American Society of Human Genetics 64th Annual Meeting
October 18–22, 2014 San Diego, CA

POSTER ABSTRACTS

Session Topic/Title	Abstract/Poster Board Numbers Start #	End #
Epigenetics	407	510
Genome Structure, Variation, and Function	511	678
Pharmacogenetics	679	741
Complex Traits and Polygenic Disorders	742	1119
Psychiatric Genetics, Neurogenetics, and Neurodegeneration	1120	1361
Bioinformatics and Genomic Technology	1362	1721
Statistical Genetics and Genetic Epidemiology	1722	1890
Evolutionary and Population Genetics	1891	2034
Cardiovascular Genetics	2035	2168
Therapy for Genetic Disorders	2169	2213
Metabolic Disorders	2214	2302
Genetics/Genomics Education	2303	2329
Ethical, Legal, Social, and Policy Issues in Genetics	2330	2383
Genetic Counseling	2384	2402
Health Services Research	2403	2419
Clinical Genetic Testing	2420	2599
Clinical Genetics and Dysmorphology	2600	2785
Prenatal, Perinatal, and Reproductive Genetics	2786	2865
Molecular Basis of Mendelian Disorders	2866	3128
Development	3129	3157
Cytogenetics	3158	3223
Cancer Genetics	3224	3509

Posters should remain on the board for all three days (Sunday through Tuesday)

POSTER AUTHOR SCHEDULE

The program and abstract/poster board number next to each listing is followed by an **S** (Sunday), **M** (Monday), or **T** (Tuesday) to indicate the day on which authors must be present at their poster boards. Refer to the schedule below for presentation times and for the poster mounting/removal schedule.

Poster Mounting and Removal

Authors must put up and take down their posters according to the schedule below. Authors must be present at their boards based on their odd or even abstract/program/board number, and must remain at their boards for the duration of their scheduled presentation times. **Posters should remain on the boards for all three days.**

Sunday, October 19

11:00 am–1:00 pm	All poster authors (Sunday, Monday, and Tuesday) place posters on boards
11:00 am–7:00 pm	Posters available for general viewing
4:00 pm–6:00 pm	**Poster Session I (Sunday Authors Present)**
	4:00 pm–5:00 pm (*odd poster board numbers; author must be present*)
	5:00 pm–6:00 pm (*even poster board numbers; author must be present*)

Monday, October 20

10:00 am–4:00 pm	Posters available for general viewing
2:00 pm–4:00 pm	**Poster Session II (Monday Authors Present)**
	2:00 pm–3:00 pm (*odd poster board numbers; author must be present*)
	3:00 pm–4:00 pm (*even poster board numbers; author must be present*)

Tuesday, October 21

10:00 am–4:00 pm	Posters available for general viewing
2:00 pm–4:00 pm	**Poster Session III (Tuesday Authors Present)**
	2:00 pm–3:00 pm (*odd poster board numbers; author must be present*)
	3:00 pm–4:00 pm (*even poster board numbers; author must be present*)
4:00 pm	Posters closed
4:00 pm–4:15 pm	All authors remove posters from boards
4:15 pm	Exhibit Hall closed

407M

5-hydroxymethylcytosine Dysregulation in Neurodegenerative Disorders. *B. Yao[1], L. Lin[1], C. Street[1], S. Yang[1], C. Stoyas[2], F. Ayhan[3], L. Duvick[4], Z. Zalewski[5], A. La Spada[2], L. Ranum[3], H. Orr[4], D. Nelson[5], P. Jin[1].* 1) Human Genetics, Emory University, Atlanta, GA; 2) Pediatrics and Cellular and Molecular Medicine, University of California San Diego, San Diego, CA; 3) Molecular Genetics & Microbiology, University of Florida, Gainesville, FL; 4) Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN; 5) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Neurodegenerative disorders are a heterogeneous group of chronic progressive diseases characterized by neuronal dysfunction, progressive degeneration, and progressive neuronal loss. It has been well established that epigenetic plasticity in DNA methylation-related regulatory processes influences activity-dependent gene regulation, learning and memory, and aging in the central nervous system. Recent studies that methylcytosine (mC) can be oxidized to 5-hydroxymethylcytosine (5hmC) present a particularly intriguing epigenetic regulatory paradigm in the mammalian brain. Our previous works suggest 5hmC modification at selective loci is altered in the cerebellum of FXTAS mouse model. In the present study, we systematically characterize the genome-wide 5hmC distributions in several neurodegenerative disorder mouse models with common features of ataxia, including Fragile X ataxia and tremor syndrome (FXTAS), spinocerebellar ataxia (SCA) type 1, 7, 8 and 17, together with their age-matched wildtype littermates. 1113 common loss-of-5hmC regions in all disease mouse models comparing to their wildtype littermates are identified, whereas 2182 common gain-of-5hmC regions are also generated. Gene Ontology analysis on these 5hmC differential regions indicates a strong correlation of neuronal pathways, such as Purkinje neuron differentiation and cerebellum development, with common loss-of-5hmC regions. Motif identification and transcription factor analysis predict several key transcription factors in neurodevelopment and neuronal functions, including RE1-Silencing Transcription factor (REST), Transcription Factor 4 (TCF4) and Foxhead box L1 (FoxL1), which could bind to DNA in a methylation-dependent manner. Altered cytosine modification at these loci thus could change their binding affinity, and affect their normal function in neurodegenerative mouse models. In addition, Approximately 15% of the 5hmC differential regions in these disease overlap with cerebellum enhancer regions, suggesting their potential epigenetic impact on distal regulatory regions. Our study provides the first systematic investigation of 5hmC profiles in various neurodegenerative disease mouse models, and highlights its key role in neuronal functions. The common differential 5hmC regions in various ataxia-related disorders may sever as “hotspots” for early diagnosis and important therapeutic targets.

408T

Modeling DNA methylation dynamics with approaches from phylogenetics. *J. Capra[1,2], D. Kostka[3].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 3) Departments of Developmental Biology and Computational & Systems Biology, University of Pittsburgh, Pittsburgh, PA.

Methylation of CpG dinucleotides is a prevalent epigenetic modification that is required for proper development in vertebrates. Genome-wide DNA methylation assays have become increasingly common, and this has enabled characterization of DNA methylation profiles from distinct stages across differentiating cellular lineages. Changes in CpG methylation are essential to cellular differentiation; however, current methods for modeling methylation dynamics do not account for the dependency structure between precursor and dependent cell types. We developed a continuous-time Markov chain approach, based on the observation that changes in methylation state over tissue differentiation can be modeled similarly to DNA nucleotide changes over evolutionary time. This model explicitly takes precursor to descendant relationships into account and enables inference of CpG methylation dynamics. To illustrate our method, we analyzed a high-resolution methylation map of the differentiation of mouse stem cells into several blood cell types. Our model can successfully infer unobserved CpG methylation states from observations at the same sites in related cell types (90% correct), and this approach more accurately reconstructs missing data than imputation based on neighboring CpGs (84% correct). Additionally, the single CpG resolution of our methylation dynamics estimates enabled us to show that DNA sequence context of CpG sites is informative about methylation dynamics across tissue differentiation. Our approach also enabled identification of genomic regions with clusters of highly dynamic CpGs. We found that enhancers, in addition to promoters and CpG islands, contain many dynamic CpG sites that are likely relevant to shifts in gene expression across the development of this lineage. Finally, we characterized genetic variation in CpG sites with a focus on mutations that disrupt the potential for CpG methylation. Our work establishes a framework for inference and modeling that is well-suited to DNA methylation data, and our success suggests that other approaches used to analyze DNA nucleotide substitutions will also translate to the modeling of epigenetic phenomena.

409M

Genome-wide Methyl-seq analysis reveals changes in hypothalamic DNA methylation patterns in response to the enhanced maternal care paradigm. *D.H. Yasui[1], T.Z. Baram[2], A. Singh[2], K.W. Dunaway[1], J.M. LaSalle[3].* 1) Medical Microbiology, UC Davis School of Medicine, Davis, CA; 2) Anatomy and Neurobiology, UC Irvine School of Medicine, Irvine, CA; 3) Medical Microbiology and Genome Center, UC Davis School of Medicine UC Davis MIND Institute, Davis, CA.

Enhanced maternal nurturing in rat models leads to a long-lasting modulation of subsequent stress responses and altered expression of key stress-controlling genes. In the “augmented maternal care” (AMC) paradigm, enhanced mothering from postnatal days 2–9 (P9) leads to reduced corticotropin releasing hormone (Crh) expression in the hypothalamic paraventricular nucleus (PVN) which persists through life. We hypothesize that AMC epigenetically sets future gene expression levels by altering DNA methylation levels. Gene levels of Crh were assayed in AMC and control PVN by bisulfite pyrosequencing analysis. As expected, Crh methylation levels were significantly higher in AMC vs control PVN, consistent with repression of Crh expression. However, Crh methylation levels were unaffected in thalamus where no differences in gene expression were observed, suggesting that elevated PVN Crh methylation levels are due to AMC and related to Crh repression. To investigate potential additional genes contributing to the stress-resilient phenotype produced by AMC, Methyl-seq analysis was performed on AMC and control PVN to identify other differentially methylated regions (DMRs). We focused on partially methylated (>70% CpG methylation) domains (PMDs), because they comprise up to 40% of the genome in early life and are enriched in repressed genes. Thus, PMDs are likely to contain DMRs responsive to AMC. We employed a hidden Markov model (HMM) trained to detect PMDs to screen for potential DMRs and identified 38 chromosomal loci, covering ~5% of the uniquely aligned rat genome. Computational analyses confirmed lower average methylation levels in 37 of the 38 identified differential PMDs, consistent with differential methylation in these regions. Interestingly, PVN DMRs included Pcdh gene clusters involved in neuronal identity. Methyl-seq analysis also identified a Crh promoter region that has potentially elevated methylation levels in AMC PVN. In summary, current studies identified both marker gene (Crh) and genome-wide methylation changes related to augmented maternal care. Current and future studies will confirm DMRs by pyrosequencing and Methyl-seq analysis and correlate these with PVN gene expression changes identified by RNA-seq. These results will provide insights into normal brain function as well as early-life influences on genome wide epigenetic reprogramming of hypothalamic genes with implications for certain human for stress-related, neuropsychiatric disorders.

410T

Whole blood DNA methylation changes are associated to malignant pleural mesothelioma. *E. Casalone[1], S. Guarrera[1,2], G. Fiorito[1,2], M. Betti[3], D. Ferrante[4], C. Di Gaetano[1,2], F. Rosa[1,2], A. Russo[1,2], S. Tunesi[4], M. Padoan[4], C. Casadio[5], F. Ardissone[6], E. Ruffini[7], P.G. Betta[8], R. Libener[8], R. Guaschino[9], E. Piccolini[10], D. Mirabelli[11,12], C. Magnani[4,12], I. Dianzani[3,12], G. Matullo[1,2].* 1) Laboratory of Genetic Pathology, Department Health Sciences, University of Piemonte Orientale, I-28100, Novara, Italy; 2) Department of Medical Sciences, University of Turin, Turin, Italy; 3) Laboratory of Genetic Pathology, Department Health Sciences, University of Piemonte Orientale, Novara, Italy; 4) 4CPO-Piemonte and Unit of Medical Statistics and Epidemiology, Department Translational Medicine, University of Piemonte Orientale, Novara, Italy; 5) 5Thoracic Surgery Unit, University of Piemonte Orientale, Novara, Italy; 6) Chest Surgery, Department of Clinical and Biological Sciences, University of Turin, Orbassano, Italy; 7) Thoracic Surgery Unit, University of Turin, Turin, Italy; 8) Pathology Unit, Azienda Ospedaliera Nazionale SS, Antonio e Biagio e Cesare Arrigo, Alessandria, Italy; 9) Transfusion Centre, Azienda Ospedaliera Nazionale SS, Antonio e Biagio e Cesare Arrigo, I-15121, Alessandria, Italy; 10) Pneumology Unit, Santo Spirito Hospital, Casale Monferrato, Italy; 11) Unit of Cancer Epidemiology, CPO-Piemonte and University of Turin, Turin, Italy; 12) Interdepartmental Center for Studies on Asbestos and other Toxic Particulates "G. Scansetti", University of Turin, Turin, Italy.

Malignant pleural mesothelioma (MPM) is a rare and aggressive tumor strongly associated with asbestos exposure. Its onset is usually 30–40 years after the first exposure, and it is characterized by a poor prognosis with a median survival of 12 months. Alterations in DNA methylation have been reported in several cancers, and are becoming an established hallmark of tumor. MPM is frequently associated with genetic mutations but also epigenetic changes leading to gene expression modifications. The identification of MPM-specific epigenetic markers in peripheral blood might be a useful methodology for defining biomarkers for potential early detection and may define methylation changes due to asbestos exposure. We conducted an epigenome-wide analysis (>450K CpG sites) on DNA from whole blood cells of 129 MPM cases and 127 controls to evaluate differences in methylation profiles. The sample population was randomly split into two sets: training and test set. In the training set 60 differentially methylated regions (DMRs) between cases and controls, adjusting for gender, age, asbestos exposure, and white blood cells percentage were found (FDR adjusted $p<0.01$). Using a cluster algorithm we validated the DMRs prediction performance in the test set (AUC=0.7625). We found significant enrichment for genes involved in leukocyte trans-endothelial migration, natural killer cell mediated cytotoxicity and cell adhesion molecules pathways. Moreover we identified several genes belonging to the inflammation pathways and related to the cancer progression. Our results suggested that methylation status in whole blood DNA might provide a useful biomarker for potential MPM early detection.

411M

UHRF1 as an epigenetic regulator of CDH1 (E-cadherin) in human breast cancer cell invasion and metastasis. *S.Y. Jang[1], H. Kim[1], C.H. Kim[1], A. Kim[1,2], J. Kim[1].* 1) korea university, seoul, South Korea; 2) Korea University Guro Hospital, Seoul, South Korea.

Human UHRF1 (ubiquitin-like PHD and RING-finger-associated domain-containing) is an epigenetic regulator that is involved in DNA methylation and histone modification. Overexpression in several types of human cancer including breast cancer has suggested the UHRF1 as a putative oncogene. However, molecular mechanisms underlying the role of the UHRF1 in the cancer are still unclear. To understand the epigenetic role of the UHRF1 in regulating the progression of human breast cancer, we analyzed changes in cell morphology, migration, and invasion as well as epithelial and mesenchymal marker proteins. We found that the UHRF1 enhanced motility and invasiveness in the human breast cancer cells. Chromatin immunoprecipitations (ChIP) and immunoprecipitation (IP) assays showed that the UHRF1 protein directly bound and repressed the CDH1 though H3K9me2/3 and H3K9Ac. Inhibition of UHRF1 expression by RNA interference restored CDH1 expression in the breast cancer cells and suppressed the cell invasiveness. Taken together, our study suggested the UHRF1 is critical to epithelial-mesenchymal transition (EMT) through epigenetic regulation of CDH1 that is a hallmark of EMT and contributes to the metastatic advantage of human breast cancer.

412T

Analyses and characterization of adjacent CpG sites of TCGA BRCA HM450K methylation data. *M.P. Lee, H. Liu, H.H. Yang.* CCR, National Cancer Institute, Rockville, MD.

DNA methylation plays important roles in maintaining genome stability and regulating gene expression. It is fundamentally important for cell differentiation and dynamically responding to the physiological and pathologic conditions. The Infinium HumanMethylation450 BeadChip has emerged as a powerful platform to investigate DNA methylation across the human genome. That is the platform used by TCGA for the DNA methylation study. The current challenge in TCGA and other genomic initiatives is in the understanding of these high throughput data. DNA methylation analysis is particularly challenging, especially regarding how to combine methylation data from multiple CpG sites per gene. To address this challenge, we investigated methylation measurements as beta-values among adjacent CpG sites. We analyzed TCGA breast cancer methylation data. The latest release has HM450k data for 812 samples. An example of our analyses of these samples is summarized below. After excluding missing beta-value, we have 302559 CpG sites within 23905 genes for analysis. 21754 genes are represented by at least 2 CpG sites, with an average of 13.8 CpG sites. The variation of beta-value among the CpG sites with a gene is substantial, with standard deviation of 0.2438. The mean difference between minimum and maximum beta-value is 0.6345. We then focused on the analysis of methylation difference between the two adjacent CpG sites. As expected, we found that the beta-value difference increases as the distance between the two CpG sites increases. When the two CpG sites are within 1 kb, the difference in beta-values is usually less than 0.1. We also analyzed the beta-value differences in terms of genomic context, including sites in the promoter region, 5'UTR, first exon, gene body, and 3'UTR, CpG island and CpG island shore, transcription levels, and clinical phenotypes. The insights gained from our analyses of relationship among multiple CpG sites of a gene can help determine how to aggregate multiple CpG sites per gene and how to combine DNA methylation data with gene expression data at gene-level and thus can provide an effective strategy to integrate methylation and gene expression and clinical phenotypes.

413M

Sports Related Concussions Induce DNA Methylation Changes in Immune Cell Trafficking and Cell Survival Pathways. *H. Kim[1], J. Gill[1], S. Yun[3], A. Cashion[1], D. Wang[1], K. Merchant-Borna[2], H. Lee[1], J. Bazarian[2].* 1) SML, NINR/NIH, Bethesda, MD; 2) University of Rochester Medical Center, Rochester, NY; 3) Yotta Bioinformatics, Bethesda, MD.

Sports related concussions (SRC) are common among athletes and place them at risk for neurological and psychological dysfunction. There is considerable variation and outcome from SRC; with most athletes recovering within one week, while others may remain symptomatic for more than one year. The purpose of this study was to determine changes in DNA methylation after a SRC in order to characterize the underlying molecular processes that accompanied recovery. Pre-season blood samples were obtained from 253 college athletes. Athletes were followed over the season and 11 sustained a SRC with subsequent blood obtained 7 days following the injury. In the SRC athletes, whole genome scale DNA methylation using Methyl Binding Domain (MBD)-seq method in peripheral blood mononuclear cells (PBMCs) was compared at baseline (pre-injury) and 7 days following the injury (sub-acute). Genomic regions that differed between baseline and the sub-acute period were determined by MACS for peak calling, PAVIS for annotation, and in-house PERL scripts provided a list of genes that significantly differed in methylation between the two time points. Ingenuity Pathway Analysis was used to determine candidate biological pathways and gene networks most related to SRC. Top canonical pathways related to SRC included: the TCA cycle II, glucocorticoid receptor signaling, as well as cellular salvage pathways, which includes xanthine and xanthosine. Cell death and survival were the primary biological function network related to SRCs. Genes with reduced DNA methylation during the sub-acute period included allograft inflammatory factor 1 (AIF1), a gene that mitigates microglial activation following TBI, and is related to inflammatory activity and regulation. The gene baculoviral IAP repeat containing 3 (BIRC3) which functions to reduce apoptosis following TBI, was also reduced in methylation. Reduced methylation was also found in regions of genes related to protein ubiquitination including the gene SNF8. Our findings illustrate that genes related to neurological disease and psychological disorders are significantly affected by SRC, resulting in reduced methylation of genes related to cellular functions that promote recovery including immune cell trafficking and cell survival. These findings provide unique insights into molecular recovery mechanism following SRC.

414T

Genome-wide DNA methylation analyses identify loci influencing recurrent stroke risk in samples from the Vitamin Intervention for Stroke Prevention clinical trial. *K.L. Keene[1,2], W.M. Chen[2], S.D. Turner[2], A.F. Koeppel[2], S.R. Williams[2], M.M. Sale[2], B.B. Worrall[2].* 1) East Carolina University, Greenville, NC; 2) University of Virginia, Charlottesville, VA.

DNA methylation is a widely accepted epigenetic factor that is important in many diseases. The folate one carbon metabolism (FOCM) pathway plays a critical role in DNA methylation through the conversion of homocysteine to methionine, wherein s-adenosyl methionine is the primary donor for de novo methyltransferase reactions. In addition to its role in DNA methylation, elevated homocyteine levels have been associated with several diseases, including cardiovascular disease and stroke. Due to the cross-talk between genes in the FOCM pathway, DNA methylation, and stroke risk, we have performed analyses of genome-wide DNA methylation status using DNA extracted from buffy coats and Illumina HumanMethylation450 BeadChip arrays, to identify regions of hypo- or hypermethylation that may influence recurrent stroke risk in 183 (105 European descent (ED) and 78 African descent (AD)) samples from the Vitamin Intervention for Stroke Prevention (VISP) clinical trial. VISP was a multi-center, double-blind, randomized, controlled clinical trial designed to determine if daily intake of a multivitamin tablet with high dose folic acid, vitamin B6 and vitamin B12 reduced recurrent cerebral infarction, nonfatal myocardial infarction (MI), or mortality. Ethnicity and gender stratified analyses identified two loci with significant ($P < 10^{-8}$) differential DNA methylation in ED males with (n= 18) and without (n= 39) a recurrent stroke, and one locus with differential DNA methylation in ED females with (n= 14) and without (n= 34) a recurrent stroke. The most significant differentially methylated locus in ED males was located near the cysteine/histidine-rich 1 (CYHR1) gene ($P= 4.85 \times 10^{-08}$; Qval= 0.005). In ED females, the most significant locus was near the centlein, centrosomal protein (CNTLN) gene ($P= 1.70 \times 10^{-08}$; Qval= 0.004). No loci showed significant differential methylation in the AD samples. The CNTLN gene is located adjacent to the 9p21 locus, a region with prior genetic associations with ischemic stroke and coronary diseases. The protein encoded by the CYHR1 gene has been identified as a biomarker of good response to erythropoietin (EPO) in hemodialysis patients. Furthermore, recent studies suggest that EPO might have neuroprotective properties in the ischemic brain. Collectively, our findings could provide insight into epigenetic marks that influence recurrent stroke risk and uncover targeted treatments and regimens to alleviate recurrent stroke.

415M

Vitamin B deficiency and gestational programming of genes related to Alzheimer's disease. *V.C. Silva[1], A.L.D.A Agamme[1], L. Fernandes[1], M.T.C Muniz[2], V. D'Almeida[1].* 1) Psychobiology, Universidade Federal de So Paulo, So Paulo, So Paulo, Brazil; 2) Department of Biological Sciences, Universidade de Pernambuco, Recife/PE.

Fetal antecedents have been associated with increased offspring disease risk through maternal environment, placental changes, and epigenetic programming. Epigenetic regulation is crucial in the development of organisms, since it is required to achieve either stable genes expression or repression at various stages of development. As methionine-homocysteine pathway supports methyl groups among other purposes, to DNA and protein methylation and this flow is sensitive to the supply of amino acids and vitamins, we investigated whether maternal Vitamin B deficiency during early development alters the gene expression of *App*, *Ps1* and *Bace*, which are related to Alzheimer's disease. To this end, we submitted female mice to experimental diet one month before and during pregnancy. After 20 days, plasma homocysteine levels from deficient gestational group (DG, n=7) were approximately 50% higher than control group (CT, n=7) (CT= 9.219 μmol/L, DG= 13.606 μmol/L; p=0.0009). The offspring's male mice (n= 6–8, per group) from control and experimental dams (CTO and DGO) were euthanized after birth, the total brain was harvested and mRNA isolated by Ilustra triplePrep® kit (GE HealthCare). Gene expression of *App*, *Bace* and *Ps1* was quantified by real time PCR using *Gapdh* as the housekeeping gene and data were analyzed by 2-ΔΔCT method. T-test was used to compare the results considering the level of significance $p<0.05$. Our results demonstrated a significant decrease in the *App* and *Ps1* gene expression and no significant difference in *Bace* expression in DGO male offspring at postnatal day 0. The decrease in gene expression showed in present study can be due to an imbalance in the one carbon metabolism, since in a previous study of our group a decrease in S-adenosylmethionine (SAM)/S-adenosylhomocysteine (SAH) ratio was observed. Considering that the manipulation occurred in crucial period of development (prenatal), a higher vulnerability to degenerative processes could be expected as a long lasting consequences of this imbalance. Sources of research support: FAPESP, CNPq and AFIP.

416T

Global Methylation of Fracture Risk in a Cohort of Young African Americans with Forearm Fractures. *C. Sprouse[1], B.T. Harmon[1], H. Gordish-Dressman[1], L.M. Ryan[3], L.L. Tosi[1,2], J.M. Devaney[1,2].* 1) Children's national medical center, Washington, DC; 2) The School of Medicine and Health Sciences, George Washington University, Washington DC; 3) Department of Pediatric Emergency Medicine, The Johns Hopkins Hospital, Baltimore MD.

Forearm fractures account for roughly 25% of all pediatric fractures. The estimated direct costs of treating childhood forearm fractures currently exceed $2 billion dollars per year in the United States. Despite an overall trend of declining injury rates in children, multiple studies have shown that the incidence of forearm fractures among children is increasing. The reasons for this increase in pediatric forearm fracture rates remains unclear. The objective of this study was to provide new insights into the influence of methylation on fracture risk in a young African American cohort. Additionally, we hoped to identify novel epigenetic markers and elucidate novel genetic pathways, which contribute to fracture risk. Our cohort included 138 African American children ages 5 to 9 who were recruited as part of a forearm fracture study. The DNA was extracted from whole blood and analyzed for methylation changes using the Illumina HumanMethylation450 BeadChip (485,577 CpG sites). Statistical analysis was completed on 395,899 CpG sites (Partek Genomics Suite) comparing participants with forearm fractures (n=69) and age-matched controls (n=69). Associations between methylation and fracture status were examined using ANCOVA models with gender and age as covariates. After scrubbing the data and adjusting for multiple testing, 490 CpG sites were significantly different between children with forearm fractures and control participants. The top ranked CpG site by p-value (p= 1.1 X 10-08; cg23369647) was located on a CPG island of the gene SCML4, a novel gene for fracture risk. This gene demonstrated a 58% higher beta value in children with forearm fracture compared with age matched control participants.

417M

Whole-blood DNA methylation patterns and glycemic traits in the KORA F4 study. *J. Kriebel[1,2,3], S. Wahl[1,2,3], S. Zeilinger[1,2], K. Schramm[4,5], W. Rathmann[6], M. Roden[7,8,9], A. Peters[1,2], T. Illig[1,10], M. Waldenberger[1,2], H. Prokisch[4,5], H. Grallert[1,2,3], C. Herder[7,9].* 1) Research Unit of Molecular Epidemiology, Helmholtz Zentrum Mnchen, German Research Center for Environmental Health, Neuherberg, Germany; 2) Institute of Epidemiology II, Helmholtz Zentrum Mnchen, German Research Center for Environmental Health, Neuherberg, Germany; 3) German Center for Diabetes Research (DZD), Neuherberg, Germany; 4) Institute of Human Genetics, Helmholtz Zentrum Mnchen, German Research Center for Environmental Health, Neuherberg, Germany; 5) Institute of Human Genetics, Technische Universitt Mnchen, Munich, Germany; 6) Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Dsseldorf, Dsseldorf, Germany; 7) Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Dsseldorf, Dsseldorf, Germany; 8) Department of Endocrinology and Diabetology, University Hospital Dsseldorf, Dsseldorf, Germany; 9) German Center for Diabetes Research (DZD), Dsseldorf, Germany; 10) Hannover Unified Biobank, Hannover Medical School, Hannover, Germany.

Recent publications indicate an involvement of DNA methylation in the development of type 2 diabetes (T2D). Thus, it is important to study the associations between genome-wide DNA methylation and blood glucose and insulin levels. The aim of the present study was to investigate the role of DNA methylation in whole blood in the development of T2D. Therefore, we performed a methylome-wide association study of glycemic traits (fasting insulin, fasting and 2h glucose) in 1448 whole blood samples of the population-based KORA F4 study using the HumanMethylation 450k BeadChip. Data were BMIQ normalized (beta mixed quantile normalized) and analyzed using linear mixed effect models. We identified four CpG sites that were significantly associated with glycemic traits after correction for multiple testing (Benjamini-Hochberg procedure). In detail, DNA methylation at cg11024682 (located in the *SREBF1* gene locus) was associated with fasting glucose (B-H-adjusted $p = 1.7\times10^{-3}$), cg09613192 and cg22065733 (both unannotated) were associated with fasting insulin (B-H-adjusted $p = 1.8\times10^{-2}$ and 4.2×10^{-2}, respectively), and cg06500161 (located in the *ABCG1* gene locus) was associated with fasting glucose, 2h glucose, and fasting insulin (B-H adjusted $p = 3.0\times10^{-4}$, 1.8×10^{-2}, and 4.3×10^{-3}, respectively) after adjustment for age and sex. After adjustment for body mass index (BMI) or estimated white blood cell counts, we did not detect any genome-wide significant signals. We additionally tested associations between DNA methylation and gene expression in a subset of 538 participants with available gene expression data using linear model and Benjamini-Hochberg correction for multiples testing including an area of ± 500kb around significant CpG sites. The CpG site cg06500161 (*ABCG1*) showed an association with *ABCG1* gene expression level (B-H adjusted $p = 3.0\times10^{-10}$). Our findings support previous evidence of a role of the ABCG1 gene in the regulation of glucose metabolism and suggest that the association might be modulated by epigenetic processes. Our study points towards an association between DNA methylation and T2D related traits. Analyses in insulin-responsive tissues such as adipose tissue, muscle or liver might help to elucidate the underlying mechanisms.

418T

The effect of fertility treatments on differential DNA methylation in autism spectrum disorder. *M.T. Siu[1], D. Grafodatskaya[1], D.T. Butcher[1], S. Choufani[1], Y. Chen[1], A. Pietrobon[1], Y. Lou[1], R. Weksberg[1,2,3].* 1) Genetics & Genome Biology, Hospital for Sick Children, Toronto, ON, Canada; 2) Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, ON, Canada; 3) Department of Paediatrics, The University of Toronto, ON, Canada.

Autism spectrum disorder (ASD) is one of the most common neurodevelopmental disorders, with a prevalence of approximately 1 in 88 in North America. Both genetic and environmental factors likely contribute to ASD risk, and the role of epigenetic dysregulation requires further clarification. One possible environmental exposure that may impact epigenetic regulation is the use of fertility treatments (FT). It has previously been shown that the frequency of DNA methylation (DNAm) alterations are increased following FT in etiologically heterogeneous disorders such as Angelman syndrome. To date, FT, including assisted reproductive technologies (ART) and hormonal stimulation of ovulation, contribute to 6% of live-births in the United States per year. FT is associated with increased risks for multiple births, prematurity and low birth weight. In addition, concerns have been raised about the potential for enhanced risks for birth defects and neurodevelopmental disorders. Whether FT contributes to the risk of ASD has yet to be established due to inconsistent results found in epidemiological studies. Our study aims to investigate the intersection of ASD and FT with respect to epigenetic alterations, as measured by changes in DNAm levels in white blood cells. Further, as the epigenetic effects of FT occur during oocyte maturation and early embryonic development, they are likely to be reflected in multiple tissues of the individual, including clinically accessible samples such as blood. Using the Illumina Infinium 450K platform, which covers over 450, 000 CpG sites, we compared genome-wide DNAm levels in five groups: 19 patients diagnosed with ASD and conceived with FT (ASD-FT), 17 patients with ASD conceived naturally (ASD-NC), their respective sex-matched, unaffected siblings, and 23 unrelated controls. Following correction for multiple comparisons in our preliminary analysis, two CpG sites were found to be significantly hypermethylated in the ASD-FT group compared to controls ($q < 0.05$), whereas no probes were found to be significantly affected in any of the other three groups. Furthermore, when global DNAm was assessed, there was a trend towards hypermethylation in the ASD-FT group only. In summary, our results demonstrate both site specific and global effects on DNAm in the ASD-FT group, suggesting that FT may predispose embryos to epigenetic changes, contributing to ASD risk.

419M

Distinct Patterns of DNA Methylation in Labial Salivary Gland Tissue Based on Sjgren's Syndrome Disease Status. *M.B. Cole[1], X. Shao[2], D. Quach[2], H. Quach[2], A. Baker[2], L.F. Barcellos[2], L.A. Criswell[3].* 1) Department of Physics, University of California, Berkeley, Berkeley, CA; 2) Division of Epidemiology, Genetic Epidemiology and Genomics Laboratory, School of Public Health, University of California, Berkeley, Berkeley, CA; 3) Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California, San Francisco, CA.

Sjgren's Syndrome (SS, OMIM #270150) is a chronic, multisystem autoimmune disease characterized by progressive destruction of the exocrine glands, with subsequent mucosal and conjunctival dryness. A growing body of evidence indicates that epigenetic changes, in particular, altered patterns of DNA methylation, contribute to the development of this complex disease, modulating risk and severity. We report on an expanded case-control study of DNA methylation differences within labial salivary gland tissues, using biopsies sampled from 12 primary SS cases and 5 controls in the Sjgren's International Collaborative Clinical Alliance (SICCA; http://sicca.ucsf.edu/; HHSN268201300057C) collection. These subjects are part of a larger, 36-subject study group for which blood, gland tissue, and cell-sorted blood samples have been methylotyped (110 samples total). We generated genome-wide DNA methylation profiles using Illumina HumanMethylation450 BeadChips and further characterized full genome SNP profiles using the Illumina HumanOmni2.5-Quad platform. All methylation results were background subtracted ('noob' method in methylumi Bioconductor package) and normalized via all sample mean normalization (ASMN) and beta-mixture quantile normalization (BMIQ). Multidimensional Scaling (MDS) applied to the 360,546 CpG sites passing strict QC criteria visibly separates cases from controls within the first 2–3 components, and this clustering changes substantially with the inclusion of a separate set of 9 symptomatic SICCA controls without true SS (based on SICCA's extensive clinical and serologic data). We demonstrate significant gene-centered mean hypomethylation across *IL10* and *IRF5* in SS cases (FDR ≤ .05). Mean methylation levels within 15 other putative SS-associated genes were similar between cases and controls. We report median methylation levels in specific *BLK* and *KLHL24* CpGs that are 15–25% hypermethylated in SS cases in addition to other sites in *IRF5* and *BLK* displaying 10–20% hypomethylation. Single-site methylation differences across 7 of the 17 genes show model-dependent significance under multiple-testing correction, with 6 of the 7 genes exhibiting hypomethylation at a majority of sites. Our results emphasize the utility of DNA methylation as a potential biomarker of disease status. Additional research, including studies of pathway-specific gene expression will be required to fully define the role of DNA methylation in SS-affected salivary glands.

420T

Age and aging dependent epigenetic drift in Danish twins. *Q. Tan, K. Christensen, L. Christiansen.* Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Odense, Denmark.

Aging is a biological process involving genetic and epigenetic regulation in the cross-talk between the organism and environment leading to the plasticity of aging. Revealing the molecular and epigenetic mechanisms underlying aging is a challenging aspect in biomedical research. We performed genome-wide DNA methylation profiling on Danish twins collected using cross-sectional and longitudinal designs to look for age and aging dependent epigenetic modification across the genome. The cross-sectional data include 150 pairs of monozygotic twins aged 30–74 who were extremely discordant for birth-weight and the longitudinal data contain 43 pairs of like-sex twins (28 monozygotic pairs and 25 dizygotic pairs) aged from 73–82 followed up 10 years until age 83–92. Data were analyzed using the mixed effect model with a kinship matrix to account for the genetic relatedness in twin data. Our epigenome-wide association study (EWAS) on cross-sectional twin data identified 1088 CpG sites differentially regulated (p<5e–05) by age with 610 hypo- and 478 hyper-methylated. The CpG showing highest significance (p = 1.08e-20, cg16867657) on chromosome 6 is located in the CpG island for gene ELOVL2. This gene has been recently reported as an epigenetic marker for age. With p<5e-05, only 21 CpGs were identified with 13 hypo- and 8 hyper-methylated. However, our analysis revealed a predominant pattern of age dependent hypomethylation on the X-chromosome in males which was replicated by a sub-sample of the cross-sectional twins aged over 56 suggesting more progressive loss of the DNA methylation status at advanced ages in males than in females. Overall, our results showed aging is accompanied by loss of control over the DNA methylome with more hypo- than hyper-methylations especially on X-chromosome in males.

421M

Differentially co-methylated genes in postmortem prefrontal cortex of individuals with alcohol use disorders. *H. Zhang[1], F. Wang[1], H. Xu[1], H. Zhao[2], J.H. Krystal[1], J. Gelernter[1].* 1) Psychiatry Dept / VAMC, Yale University School of Medicine, West Haven, CT; 2) Biostatistics Dept, Yale University School of Public Health, New Haven, CT.

Alcohol use disorders (AUDs) may be associated with regional DNA methylation changes reflecting the vulnerability to heavy drinking and the resulting impact on the brain. Genome-wide DNA methylation was examined in postmortem prefrontal cortex (PFC) of 46 European Australians (23 AUD cases: 16 men and 7 women; 23 matched controls: 16 men and 7 women) using Illumina's HumanMethylation450 BeadChip assays. Differentially methylated CpGs and CpG clusters were identified by multiple linear regression and co-methylation analyses. 58,581 CpGs showed differential methylation in AUD subjects ($P_{nominal}$=2.510^{-7}-0.05), but the significance did not survive multiple testing correction (q >0.05). Secondary analyses of data from the 16 male cases and 16 male controls identified 93,919 differentially methylated CpGs in male AUD subjects ($P_{nominal}$=3.110^{-8}-0.05), and the findings from 21,902 CpGs withstood multiple testing correction (q≤0.05). The majority (86.1%) of the 1,178 differentially methylated CpGs ($P_{nominal}$ ≤1.010^{-4}, q≤0.05) were hypermethylated in male AUD subjects. These 1,178 CpGs were clustered into two modules of inter-correlated CpGs, and both modules were significantly associated with AUDs (Module 1: 38 CpGs, $P_{nominal}$ = 6.310^{-7}; Module 2: 1,139 CpGs, $P_{nominal}$ =1.710^{-8}). 1,139 CpGs in Module 2 were mapped to 868 unique genes associated with Wnt signaling (PBH-adj=0.01) and synapes (PBH-adj =0.02) pathways. Taken together, this genome-wide analysis revealed large numbers of postmortem PFC DNA methylation alterations associated with AUDs, predominately increases in DNA methylation. These alterations were enriched for genes associated with neuronal function and were more prominent in males.

422T

Applying the Infinium HumanMethylation450 BeadChip assay to archival formalin-fixed paraffin embedded material for large epidemiological studies. *E.M. Wong[1], J.H.E Joo[1], C.A. McLean[2], L. Baglietto[3], D.R. English[4], G. Severi[3], J.L. Hopper[4], R. Milne[3], L.M. FitzGerald[3], G.G. Giles[3], M.C. Southey[1].* 1) Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Victoria 3010, Australia; 2) Anatomical Pathology, Alfred Health, The Alfred Hospital, Victoria 3181, Australia; 3) Cancer Epidemiology Centre, Cancer Council Victoria, Victoria 3053, Australia; 4) Centre for Epidemiology and Biostatistics, The University of Melbourne, Victoria 3010, Australia.

For many large epidemiological studies, archival formalin-fixed paraffin-embedded (FFPE) tumour material represents both a precious and technically challenging resource for molecular studies due to low yields of often highly-degraded DNA. Using FFPE tumour-derived DNA resources is particularly challenging for current genome-wide detection assays that typically require large amounts of high quality DNA. We have thus devised an improved and robust experimental workflow which includes two quality control (QC) checkpoints to address the issues of low yield and quality of this sample type prior to application on the Infinium HumanMethylation450 Beadchip (HM450K) assay.

Using resources from the Melbourne Collaborative Cohort Study, tumour-derived DNA was extracted from 475 macrodissected FFPE breast tumours. Four hundred and seventy FFPE tumour-derived samples with detectable amounts of DNA progressed through the experimental workflow to QC checkpoint 2, where 42 (8.93%) samples failed to pass the hurdle. Of the remaining 428 samples assayed on the HM450K platform, only 4 (1%) samples did not have average detection p-values across all probes of p≤0.01 and thus, failed our QC criterion. Moreover, any probe within a sample with a detection p-value of ≥0.01 was considered technical noise and subsequently removed. The remaining samples (n=424) had an average detection P-value across all probes of 4.54×10^{-4} with high correlations between biological replicates (correlation coefficient r>0.983) further confirming the capacity of our protocol to generate reproducible high quality data.

We have devised a protocol suitable for the analysis of macrodissected FFPE material on the Infinium HumanMethylation450K Beadchip assay and have achieved a 99% success rate on the HM450K platform when applied to FFPE tumour-derived DNA. This protocol will have utility in epidemiological studies that wish to apply this assay to large numbers of samples in a standard and controlled manner and is applicable to a wide variety of studies that wish to investigate the role of methylation in the predisposition and progression of disease.

423M

Allele-specific distribution of 5-hydroxymethylcytosine at imprinting control regions. *K. Yamazawa[1,2], A. Ferguson-Smith[2].* 1) Pediatrics, Keio University School of Medicine, Tokyo, Japan; 2) Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK.

Covalent modification of cytosine methylation at carbon five represents a major epigenetic mark of mammalian genome and plays a pivotal role in various biological processes. Recent studies have indicated that the Ten-eleven translocation (Tet) family proteins can enzymatically convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). Additional processing steps and subsequent base excision repair mechanism would then result in the removal of the methylated base and its substitution with an unmethylated cytosine. However, very little is known about the distribution and function of 5hmC at differentially methylated regions (DMRs) which regulate the expression of imprinted genes. To clarify the role of 5hmC in the epigenetic landscape of genomic imprinting, we profiled the allele-specific distribution of 5hmC at representative DMRs in brain, liver, placenta and ES cells derived from reciprocal hybrid mice. The hydroxymethylated and methylated DNA immunoprecipitations with specific antibodies to 5hmC and 5mC, respectively, were followed by quantitative PCR and pyrosequencing for SNP genotyping in hybrid crosses. We found that brain contains a substantial amount of 5hmC and that 5mC and 5hmC exist on the same allele at DMRs in principle. Our findings raise the possibility that 5hmC could be generated on conversion of 5mC as an intermediate inactive demethylation process at DMRs. It is also noteworthy that methylation analysis with conventional bisulfite conversion could overestimate methylation levels at DMRs.

424T

Obesity accelerates epigenetic aging of human liver. *J.E. Hampe[1], W. Erhart[4], M. Brosch[1], O. Ammerpohl[5], W. von Schnfels[6], M. Ahrens[6], N. Heits[6], J.T. Bell[7], T.D. Spector[7], P. Deloukas[7,8,9], R. Siebert[5], B. Sipos[10], T. Becker[6], C. Roecken[11], C. Schafmayer[6], S. Horvath[2,3].* 1) Medical Department I, TU Dresden, University Hospital Dresden, Dresden, Saxony, Germany; 2) Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA; 3) Biostatistics, School of Public Health, University of California Los Angeles, Los Angeles, California, USA 90095; 4) Department of Medicine I, University Hospital Schleswig-Holstein, 24015 Kiel, Germany; 5) Department Institute of Human Genetics, Christian-Albrechts-University Kiel & University Hospital Schleswig-Holstein, Campus Kiel, 24015 Kiel, Germany; 6) Department of Visceral and Thoracic Surgery, University Hospital Schleswig-Holstein, 24015 Kiel, Germany; 7) Department of Twin Research and Genetics Epidemiology, Kings College London, London SE1 7EH, UK; 8) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, EC1M 6BQ, UK; 9) Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah 21589, Saudi Arabia; 10) Institute of Pathology, University Hospital Tbingen, 72074 Tbingen, Germany; 11) Institute of Pathology, University Hospital Schleswig-Holstein, 24015 Kiel, Germany.

Obesity entails an increased risk of many chronic diseases that are typically associated with older age. While it is plausible that obesity increases the biological age of some cell types, it has been difficult to test this hypothesis due to the lack of appropriate biomarkers of aging due to difficulties in quantification of tissue age across tissues. Here, we utilize a novel biomarker of aging known as epigenetic clock to study the effect of high body mass index (BMI) on the DNA methylation ages of blood, liver, muscle and adipose tissue. Using a two-stage replication design, we show that liver tissue from obese subjects exhibits strong age acceleration effects which are is independent from histological nonalcoholic fatty liver disease. The accelerated aging in liver is not reversible after weight loss induced by bariatric surgery. In summary, obesity increases the epigenetic age of liver tissue which in turn may give rise to liver-related comorbidities of obesity such as insulin resistance and liver cancer.

425M

Epigenome-wide association study to identify new biomarkers for heart failure. *U. Afzal[1,2], M. Loh[1], J.S. Kooner[2,3,4], J.C. Chambers[1,2,4].* 1) Department of Epidemiology and Biostatistics, Imperial College London, London, UK; 2) Ealing Hospital NHS Trust, Middlesex, UK; 3) National Heart and Lung Institute, Imperial College London, London, UK; 4) Imperial College Healthcare NHS Trust, London, UK.

Heart failure is a leading cause of morbidity and mortality. In the USA alone, 5.1 million people live with heart failure, and heart failure contributed to 11% of deaths and 32 biliion US dollars annual costs. New tools for diagnosis, prevention and treatment of heart failure are urgently needed. We carried out an epigenome-wide association study to identify DNA methylation markers associated with heart failure. We investigated 93 patients with heart failure (defined as left ventricular ejection fraction <40%) and 186 healthy controls matched 2:1 based on gender, age and ethnicity. Average age in cases and controls were 68 (SD: 10.6) and 62 (SD: 10.6) years old respectively, with 82% male and 66% Caucasian participants. Aetiology of heart failure was ischaemic heart disease (47%), dilated cardiomyopathy (46%) and other etiology (17%). We measured methylation in genomic DNA from peripheral blood using the Illumina 450K HumanMethylation array. We applied a detection p-value threshold of 1×10^{-16}, and subsequently excluded markers with call rates <95%; this left 457,060 CpG sites for analysis. Associations of DNA methylation with heart failure were tested using logistic regression with adjustment for age, gender, ethnicity, white-blood cell estimates and technical covariates. Our study had 80% power to identify CpG sites with odds-ratio for heart failure of >3 in top vs bottom tertile of methylation at $P<10^{-4}$. We find 146 methylation markers associated with heart failure at $P<10^{-4}$. The markers identify 120 unique genes. Pathway analysis (Ingenuity) identifies enrichment ($P<0.05$) for genes known to be involved with cardiovascular disease, and heart muscle function including *CTNNA2*, *CAZS1*, *SMOC2*, *ECE1*, *KCNJ1*, *CBY1* and *CELSR1* ($P=7.2\times10^{-4}$ to 2.1×10^{-6}). There were no markers reaching epigenome-wide significance ($P<10^{-7}$). Our findings provide preliminary evidence for the potential of epigenome wide association to identify DNA methylation biomarkers of heart function in peripheral blood in biologically relevant genes. Our results provide the justification for larger scale efforts to investigate the relationships between DNA methylation and heart function.

426T

X chromosome dosage may cause epigenetic sex differences in asthma region 17q12-q21. *A. Altuwaijri[1], N. Fotouhi-Ardakani[2], AK. Naumova[1,2,3].* 1) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 2) The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 3) Department of Obstetrics and Gynecology, McGill University, Montreal, Quebec, Canada.

Males and females have different susceptibility to diseases including asthma. This is due in part to differences in gene expression levels, which in turn may result from sexual dimorphisms in transcriptional regulation. In our previous work (Naumova, 2013), we demonstrated sex-specific differences in DNA methylation levels of the zona pellucida binding protein 2 (ZPBP2) promoter that is located within chromosomal region 17q12-q21, one of the best replicated GWAS regions for asthma (Moffat 2007, 2010). The differences in methylation levels between males and females were consistent with the sex-specific bias in genetic association that we found in the asthmatic families collection from Saguenay- Lac-Saint-Jean (Naumova 2013). We next hypothesized that sex-specific differences in DNA methylation levels at the ZPBP2 promoter resulted from the dosage of the sex chromosome complement. Therefore, ZPBP2 DNA methylation was investigated in DNA samples extracted from fibroblast cell lines derived from individuals with different sex phenotypes and sex chromosome dosage: e.g. 46, XY females (sex reversal); 45, X females (Turner syndrome); 46, XY males; 46, XX females; 46, XX males, and a 48, XXXY male. No significant influence of sex phenotype or the presence of the SRY gene was detected. However, significant positive correlation was found between ZPBP2 methylation levels and the number of X-chromosomes. Individuals with one X chromosome had considerably lower methylation levels (n=9, average DNA methylation level 17.5 %) compared to individuals with more than one X chromosome (n=9, average DNA methylation level 29.3 %) ($p = 0.005$, t-test statistics). Thus, our data show that the X chromosome dosage has an effect on DNA methylation and perhaps may contribute to disease susceptibility. Our data are consistent with the effect of X-chromosome dosage on gene expression in mice (Wijchers 2010) and DNA methylation in humans (Grafodatskaya 2013). Currently, our laboratory is working towards identifying the X-linked gene(s) that are responsible for methylation differences at autosomal regions and at the asthma-associated region of chromosome 17q12-q21, in particular.

427M

Epigenome-wide DNA methylation and body mass index in monozygotic twins. *J.T. Bell[1], F. Gao[2], C.G. Bell[1], Y. Xia[2], W. Yuan[1], L. Roos[1], P.-C. Tsai[1], K. Ward[1], P. Deloukas[3,4], J. Wang[2], T.D. Spector[1].* 1) Department of Twin Research, King's College London, London, United Kingdom; 2) BGI-Shenzhen, Shenzhen, China; 3) William Harvey Research Institute, Queen Mary University of London, London, United Kingdom; 4) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom.

Obesity is a major risk factor for multiple common diseases. Genetic, environmental and recently epigenetic factors been shown to associate with obesity, offering potential to help understand and modify disease risk. Here, we focus on identifying the epigenetic signatures of environmentally mediated risk of obesity and its impacts, by studying monozygotic twins. We performed an epigenome-wide association scan of body mass index (BMI) in 1,446 UK monozygotic twins from 723 pairs (mean age = 51.8, mean BMI = 25.9 kg/m^2) from the EpiTwin project (www.epitwin.eu), part of TwinsUK. Whole blood DNA methylation was assayed using methylated DNA immuno-precipitation followed by high-throughput sequencing (MeDIP-seq), allowing for full genome-wide coverage at 10.5 million overlapping bins of size 500bp. Differentially methylated regions related to BMI were identified by comparing DNA methylation discordance to BMI discordance within twin pairs (mean absolute BMI twin-pair discordance = 4.4 kg/m^2, mean relative discordance = 0.2 units). The most associated signals included a region within an intron of the *CDKAL1* gene (P = 3.6e-7), a GWAS obesity and type 2 diabetes susceptibility locus. Functional annotation based on gene structure, Encode data, and chromHMM chromatin states showed strong enrichment of differential methylation in poised promoters. Genome-wide significant effects within poised promoters were observed in genes *SUSD4* (FDR = 0.04, P = 2.0e-6) and *SLC10A4* (FDR = 0.05, P = 4.8e-6), and the *SLC10A4* result validated at an overlapping CpG-site on the Illumina 450k array in whole blood samples from 355 individuals. Replication is currently ongoing in 424 individuals using MeDIP-seq. Our results show epigenetic studies in monozygotic twins are a powerful tool for identifying functional genomic changes in complex traits such as obesity.

428T

DNA-methylation and the Down Syndrome phenotype. *A. Bouman[1], P. Henneman[1], M.M.A.M. Mannens[1], A.N.P.M. Mul[1], E.J. Meijers - Heijboer[1], R.C.M. Hennekam[1,2,3].* 1) Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands; 2) Department of Pediatrics, Emma Children's Hospital, Amsterdam, The Netherlands; 3) Department of Translational Genetics, Emma Children's Hospital, Amsterdam, The Netherlands.

Down syndrome (DS) is the most frequent genetic cause of intellectual disability known today. DS is explained by the presence of an additional chromosome 21 which serves as the basis of DS-pathogenesis. However, underlying cellular and molecular processes causing the DS-phenotype remain not well understood. Several hypotheses exist regarding the implications and effects of the presence of trisomy 21. In this study we hypothesize that the presence of an extra chromosome 21 might influence the DNA-methylation pattern of (parts of) chromosome 21 and/or other loci throughout the genome. Since DNA-methylation plays an important role regarding gene transcription regulation, differentially methylated loci of specific genes can either cause increased or decreased transcription. This altered transcription may play a major role in causing several of the specific features of the DS-phenotype. White blood cell DNA-methylation patterns of 10 individuals with DS and 10 healthy (age- and sex-matched) controls were obtained using the Illumina 450K Human DNA-methylation Array. DNA was bisulphite converted using the gold standard kit of Zymo ®. Analysis of the data was performed using two freely available "R" packages; minfi and champ. We will discuss the results of our DS methylation-study and zoom in on specific genes which show a distinct methylation patterns in DS vs. controls. Our study serves as an addition in ongoing fundamental DS-research.

429M

Low-input Detection of Whole Genome Hydroxymethylation. *J. Burgess[1], D. Han[2], X. Lu[2], A. Khanna[1], C. He[2], H. Grunenwald[1], R. Vaidyanathan[1].* 1) Epicentre, an Illumina Company, 5602 Research Park Boulevard, Madison, WI, 53719; 2) University of Chicago, Department of Chemistry, Chicago, IL, 60637.

5-hydroxymethylcytosine (5hmC) is a recently discovered nucleotide variant with a multitude of implications in human disease and development, including cancer and neuro-disorders. In addition, 5hmC has been shown to be an intermediate step in demethylation, leading to gene expression, and is typically most concentrated in stem cells and brain tissue. A robust method of 5hmC detection includes Tet-assisted Bisulfite Sequencing (TAB-Seq). Used in conjunction with EpiGnomeTM whole genome bisulfite sequencing, which only requires as little as 50ng input DNA, it is possible to sequence the entire hydroxymethylome. In this study, we leveraged the robustness and low input requirement of EpiGnomeTM to obtain good mapping at 100ng input into TAB-Seq, with measured levels of hydroxymethylation that closely mimic previously published whole genome results obtained from microgram amounts of DNA. Investigations to reduce initial TAB-Seq input even further to 10ng (1000 cells) have retrieved promising proof of concept data. Corroboration of these low-input TAB-Seq/EpiGnome data with prior analyses of these mESCs reveals a high correlation. This study demonstrates robust and efficient combination of TAB-Seq with the WGBS EpiGnomeTM Kit to allow the previously unachievable low input detection of the entire 5-hydroxymethylome.

430T

Full resolution DNA methylome analysis in multiple tissues from twins. *S. Busche[1,2], X. Shao[1,2], M. Caron[2], T. Kwan[2], B. Ge[2], F. Allum[1,2], W. Cheung[1,2], S. Westfall[1], J. Qi[1,2], M.M. Simon[2], J.T. Bell[3], P. Deloukas[4,5], M. Blanchette[6], T.D. Spector[3], G. Bourque[1,2], M. Lathrop[1,2], T. Pastinen[1,2], E. Grundberg[1,2].* 1) Human Genetics, McGill University, Montreal, Quebec, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 3) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 4) William Harvey Research Institute, Queen Mary University of London, London, United Kingdom; 5) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 6) School of Computer Science, McGill University, Montreal, Quebec, Canada.

DNA methylation is essentially involved in human trait formation and disease susceptibility. Using whole genome bisulfite sequencing we generated full genome single-base resolution methylomes of 34 adipose and 27 whole blood samples from monozygotic (MZ) and dizygotic twins of the well characterized MuTHER cohort aiming to unravel the impact of environmental, genetic, and stochastic factors underlying methylation variation. We generated 2.3 Tbp data yielding an average coverage of 6.3x and 8.7x detecting 25M and 26M CpG sites in fat and blood, respectively (mean methylation ~80%). Static methylation was detected at 23% and 5% of CpGs in fat and blood, respectively, with hypomethylation strongly enriched and hypermethylation strongly depleted in promoter regions in both tissues. Pairwise comparisons identified a total of 21% and 16% population differentially methylated CpGs (pDMCs; $\Delta \geq 40\%$), including 2.2% and 1.7% extreme pDMCs (<20% vs. ≥80%) in fat and blood, respectively. Confounding sequence variants were identified to underlay more than half extreme pDMCs. Population differentially methylated regions (pDMRs) were defined to cover 10% of CpGs per tissue displaying high variation across individuals and high consistency across sites. pDMCs and pDMRs were depleted in promoter regions in both tissues. Fat pDMRs were subdivided into hypo- (<50%) and hypermethylated (≥50%): hypo-pDMRs were 12-fold enriched in enhancers, and transcription factor (TF) binding site motif analysis identified enrichment of TFs involved in energy metabolism and adipocyte differentiation including ERRA (p=1e-95), RXR (p=1e-65), and EBF1 (p=1e-39). Hyper-pDMRs mapped to genes involved in lipid metabolism (p=1.5e-5). Using the MZ twin structure we identified ~90 environmental DMRs (eDMRs) per tissue and pair with no overlap across tissue and found links to phenotype discordance, e.g. body mass index for fat-specific eDMRs. CpH methylation was rare in both tissues impacting 0.2% of total CpH sites (49% CpA, 33% CpT, 18% CpC methylation). Within the population CpH methylation patterns were mainly bimodal although a subset of normal distributed CpH methylation was detected. Genome feature association of CpH methylation is ongoing. Our study describes the first comprehensive and unbiased analysis of human population level DNA methylation variation to date, highlighting the importance of looking beyond promoter methylation and providing a guideline for designing future methylome studies.

431M

Monoamine oxidase A (MAOA) expression level predicts alcohol consumption in Rhesus macaques. *R.P. Cervera Juanes[1], R. Lee[2], B. Park[3], G. Wand[2], K.A. Grant[1], B. Ferguson[1].* 1) Neurosciences, ONPRC, OHSU, Beaverton, OR; 2) Johns Hopkins University, Baltimore, MD; 3) 3Department of Public Health and Preventive Medicine, Oregon Health & Science University, Portland, OR.

Monoamine oxidase A metabolizes the neurotransmitters serotonin, norepinephrine and dopamine, which are key regulators of behavior. Both a repeat polymorphism in the promoter of the MAOA gene (MAOA-LPR) and a history of childhood adversity have been linked to risk for alcohol use and behavioral disorders in humans and rhesus macaques. To investigate the potential genetic and epigenetic mechanisms underlying these risks, we collected blood from twelve unrelated, male rhesus macaques before and after they participated in a 12-month alcohol self-administration study. We determined MAOA-LPR genotypes, promoter CpG methylation levels and MAOA mRNA levels in all samples. We found that mRNA levels were independent of MAOA-LPR genotype. However, there was a robust association between site-specific CpG methylation level and MAOA expression. Of particular interest, our results show a strong association between MAOA mRNA level and the subsequent amount of alcohol consumed during the 12-month self-administration protocol. After 12 months of alcohol use, rhMAOA expression was significantly reduced, and there was a moderate increase in regional and site-specific CpG methylation levels. Thus our results indicate that blood MAOA mRNA level is more tightly associated with promoter methylation than with MAOA-LPR genotype. Moreover, since the level of MAOA expression was predictive of subsequent heavy alcohol use, CpG methylation may provide a molecular link between childhood adversity and sustained risk for alcohol use. Finally, our finding that chronic alcohol consumption further dampens gene expression, suggests that reduced MAOA availability, and its downstream effects on dopamine, serotonin and epinephrine levels, may also contribute to alcohol addiction. This work was supported by NIH grants U01AA020928 and U01AA013510.

432T

Genome-wide DNA methylation study in American Indians identifies five genes associated with intrauterine diabetes exposure. *P. Chen, S. Kobes, V. Ossowski, R. Nelson, J. Weil, W.C. Knowler, P.H. Bennett, C. Bogardus, L.J. Baier, R.L. Hanson.* Phoenix Epidemiology & Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

Exposure to a diabetic intrauterine environment is a strong risk factor for the development of type 2 diabetes in the offspring. DNA methylation has an important role in the epigenetic response to the environment. However, the genes which are subject to methylation in response to intrauterine diabetes exposure remain largely unknown. To discover differentially methylated genes and relevant pathways in response to intrauterine diabetes exposure, we selected, from a longitudinal study of Pima Indians, 189 individuals whose mothers had type 2 diabetes during their pregnancy and 201 individuals whose mothers didn't have diabetes during their pregnancy. We assayed the offspring's genome-wide DNA methylation in peripheral blood, collected when the participant was not diabetic, using the Illumina HM450 array. Quality control excluded the probes with a detection P value > 0.01 or having a SNP within 10 base pairs. Following these quality control measures, 395,899 methylation sites remained for association testing. The association between intrauterine diabetes exposure and methylation of CpG sites was tested in a linear model with adjustment for maternal age (at the time of the child's birth), as well as the proportion of American Indian ancestry, age and sex of the child. We calculated the difference in methylation score (d, in SD units) in those exposed to a diabetic intrauterine environment, compared with those unexposed. Exposure to a diabetic intrauterine environment was significantly associated with methylation of CpG islands in 5 genes: *LHX3* (d=0.67, FDR=4.6E-4) and *WNT9B* (d=0.52, FDR=4.8E-2) were hypermethylated in those exposed while *PPP1R3B* (d=-0.66, FDR=9.0E-4), *ATP8B3* (d=-0.56, FDR=6.7E-3) and *LINC00839* (d=-0.55, FDR=4.8E-2) were hypomethylated. The expression level of *LHX3*, a LIM1-homeodomain transcription factor which is under the control of the ISL1 transcriptional regulator, has previously been shown to be controlled by methylation in mice. In the pathway enrichment analysis of the 103 hyper-methylated and 1262 hypo-methylated genes (P<0.005) using DAVID, the sequence-specific DNA binding pathway (P=2.9E-5, Bonferroni=7.0E-3) and the guanyl-nucleotide exchange factor activity pathway (P=1.0E-4, Bonferroni=9.4E-2) were enriched. Our results identified five genes that are differentially methylated with exposure to a diabetic intrauterine environment. Validation of these findings in additional samples is ongoing.

433M

DNA enrichment as a cost effective tool to examine DNA methylation at single nucleotide resolution. *A. Czyz, V. Ruotti, A. Tan, D. Hill, S. Kuersten, R. Vaidyanathan.* R&D Department, Epicentre an Illumina company, Madison, WI., USA.

Epigenetic modifications are punctuation marks on DNA, responsible for modification of transcriptional activities. One of the most characterized epigenetic modifications is DNA methylation. DNA methylation plays a role in a variety of biological processes including embryonic development, cell differentiation, chromosome stability, and chromatin structure. Aberrant DNA methylation has been found to be associated with various intellectual disabilities, including obesity, anemia and numerous types of cancer. In mammalian cells, methylation occurs mostly at cytosines that are contained in the symmetrical dinucleotide CpG, or in an asymmetrical CHH or CHG context. Methylated nucleotides are generally grouped in promoter regions, CpG islands, and CpG shores. EpiGnome Methyl-Seq Kit is a simple, 1-day, post-bisulfite library construction method that requires only 50–100 ng of input gDNA. However, providing enough coverage for whole-genome bisulfite libraries typically requires 60–120 Gb of sequencing, and therefore, is relatively expensive. In contrast, several strategies exist for enrichment of target sequences that are more cost effective, including probes and use of methylation-specific antibodies. The central weakness of these last methods is lack of single nucleotide resolution. But, by combining enrichment methods and DNA bisulfite treatment with the EpiGnome protocol, we circumvent this shortcoming. To demonstrate, we took MeCP2 methyl binding domain and IgG A1 antibodies derived against 5mC, and used it for enrichment of beta-lymphocyte gDNA NA18508, prior to continuing with the EpiGnome library preparation kit. The prepared libraries were sequenced and analyzed by utilization of the commercially available BISMARK and MEDIPS bioinformatic packages. We observed significant enrichment of promoter and CpG islands, comparable to standard EpiGnome libraries without enrichment. With only 8GB (50M reads) sufficient for analysis, and thus possibility of multiplexing enriched libraries on HiSeq 2500 platform, the enrichment method provides a cost effective approach for methylation study.

434T

Genapha/dbASM: web based tools to investigate allele-specific methylation. *D. Daley, B. Chen, K. Ushey, G. Ellis.* Dept Med, Center for Heart and Lung Innovation, Univ British Columbia, Vancouver, BC, Canada.

As interest in studying allele-specific methylation (ASM) and it's association with common complex diseases grows, there is a need for a resource that stores and catalogs SNPs and regions that demonstrate allele specific methylation, analogous to NCBI's dbSNP. Additionally, as ASM is a regulatory mechanism that may be associated with hits from genome-wide association studies (GWAS), researchers need a suite of tools to help them evaluate the relationship between GWAS hits and ASM. To facilitate these investigations, we have created a new web resource called dbASM, hosted on the Genapha web server (www.genapha.ca). The aim of dbASM is twofold: 1. Curate from the literature a publicly-accessible database of known sites of ASM. 2. Provide researchers with a web-based platform of tools for exploring ASM and determining regions of interest. We will present the dbASM resource including details on the underlying database construction and datasets, in addition to the web tools and example workflows. The web tools that are currently available are: GWAS Catalog SNP Search, ASM SNP Search, SNP Counter, Methylation Plots Generation, and Sequence Viewer. GWAS Catalog SNP Search allows browsing through NHGRI's Catalog of Published Genome-Wide Association Studies by phenotype and filtering GWAS SNP's based on their relation to suspected sites of ASM. For example, rs11742570 is associated with inflammatory bowel disease (p=2.0 E-82) and demonstrates ASM. ASM SNP Search supports finding SNP's based on: ASM status or interrogability; location compared to genes, a chromosomal region, or other SNP's; and filtering by population minor allele frequencies and sample size. SNP Counter uses asynchronous JavaScript calls to the database to provide real-time counts of types of SNP's in user-selected regions of chromosome. Methylation Plots Generation is a calculates SNP correlation stratifying by genotype with CpG site methlyation patterns, similar to epigenome wide association studies (but without disease status),using CEPH HapMap samples and genotypes and methylation assays on these same samples completed on the Illumina 27K array. Sequence Viewer displays SNP's in the human reference genome (based currently on GRCh37.p10 and dbSNP build 137) with annotations showing ASM SNP's and regions of interrogability via MSRE cut sites for enzymes: HpyCH4IV, AciI, HhaI, and HpaI. These tools are all freely available for use at: http://genapha.icapture.ubc.ca/asm/.

435M

DNA Methylome Modifications Associated with HPA Axis Differences in Chronic Fatigue Syndrome. *W.C. de Vega[1,2,3], P. Manser[4,5], M. Reimers[4,6], S.D. Vernon[7], P.O. McGowan[1,2,3].* 1) Centre for Environmental Epigenetics and Development, University of Toronto Scarborough, Toronto, ON, Canada; 2) Department of Biological Sciences, University of Toronto Scarborough, Toronto, ON, Canada; 3) Department of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada; 4) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; 5) Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA; 6) Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA; 7) CFIDS Association of America, Charlotte, NC, USA.

Chronic Fatigue Syndrome (CFS) is a complex multifactorial disease that is characterized by the presence of fatigue and other particular symptoms that cannot be mitigated by rest. CFS is a multisystem and heterogeneous disease, and the biological basis of CFS is poorly understood. Some of the most consistent differences observed in previous CFS studies are modified immune gene expression and changes in the hypothalamic-pituitary-adrenal (HPA) axis that include mild hypocortisolism and enhanced negative feedback to glucocorticoids. To determine a potential role of epigenetics in CFS, we examined the DNA methylome in peripheral blood mononuclear cells (PBMCs) isolated from female CFS patients (n = 46) and healthy controls (n = 25) using the Illumina HumanMethylation450 BeadChip array. All subjects were non-obese and did not consume medications with known immunological or epigenetic effects. A two-tiered approach (a high-confidence and a medium-confidence cutoff) was applied to examine evidence for differential DNA methylation in CFS. The high-confidence cutoff identified 378 differentially methylated sites and the medium-confidence cutoff identified an additional 13,059 differentially methylated sites. RAND-36 scores, a survey assessing quality of life, revealed that CFS patients had lower scores compared to healthy controls across most categories. The first two principal components of RAND-36 explained 85% of DNA methylation variability among subjects and discriminated between CFS and controls accurately. HPA axis response in PBMCs was explored using an in vitro dexamethasone suppression assay. An increased sensitivity to dexamethasone in CFS was observed with the appearance of two subgroups (responders and non-responders) among the CFS subjects. We identified 18 differentially methylated sites between CFS and controls among the medium confidence group that were associated with the HPA axis. DNA methylation differences also existed between responders and non-responders, which underscores the potential of using DNA methylation differences in determining CFS subtypes. Our work is the first to examine DNA methylome changes in CFS and supports a role for epigenetic modifications in the increased glucocorticoid sensitivity observed in some CFS patients. These epigenomic differences could serve as potential biomarkers for future clinical research and assist in determining the biological basis of CFS.

436T

Smoking Associated DNA Methylation Changes in Peripheral Blood Mononuclear Cells from African American Women and Weighted Protein-Protein Interaction Networks. *M.V. Dogan[1,2], B. Shields[2], C. Cutrona[3], L. Gao[1], F.X. Gibbons[4], R. Simons[5], M. Monick[6], G.H. Brody[7], K. Tan[1,6], S.R.H. Beach[7], R.A. Philibert[1,2].* 1) Department of Biomedical Engineering, University of Iowa, Iowa City, IA; 2) Department of Psychiatry, University of Iowa, Iowa City, IA; 3) Department of Psychology, Iowa State University, Ames, IA; 4) Department of Psychology, University of Connecticut, Storrs, CT; 5) Department of Sociology, University of Georgia, Athens, GA; 6) Department of Internal Medicine, University of Iowa, Iowa City, IA; 7) Center for Family Research, University of Georgia, Athens, GA.

Smoking is the largest yet preventable cause of morbidity and mortality in the United States. Regular smoking has been attributed to the vulnerability and progression of numerous diseases and disabilities including cancer, diabetes and stroke. Understanding mechanism(s) through which smoking increases vulnerability to complex disorders is crucial in delineating both prevention and treatment options. Our study examined DNA methylation of peripheral blood mononuclear cells from 111 African American women. In order to determine smoking associated perturbation of DNA methylation, genome wide methylation obtained from the Illumina HumanMethylation 450k BeadChip was analyzed with respect to smoking status, controlling for slide, plate and mixed cell population. This analysis yielded 910 significant loci after genome wide Benjamini-Hochberg correction. Of these loci, two loci from the AHRR gene (cg05575921, corrected p-value = 6.17E-19 and cg23576855, corrected p-value = 3.37E-12) and one locus from the GPR15 gene (cg19859270, corrected p-value = 1.19E-19) were highly significantly differentially methylated. Due to a mere 0.1 difference in average methylation values between smokers and non-smokers at numerous loci, it was crucial to validate the Illumina array findings. Since cg05575921 is not only sensitive to long-term smoking, but also to early smoking, this locus is a consistent, sensitive smoking biomarker. Hence, a quantitative-PCR primer probe set was used to independently validate the differential methylation at cg05575921. A strong correlation of 0.94 was achieved between the primer probe set and the Illumina methylation values. Since there is growing evidence suggesting that genes function in networks rather than independently, methylation at a single locus was translated into weighted protein-protein interaction networks using network theory to obtain an integrated understanding of the proteome. Critical sub-networks were generated using the miPALM algorithm. This greedy search algorithm mapped smoking associated DNA methylation to 10 significant protein sub-networks ranging from 5 to 25 proteins. BiNGO analysis of these networks demonstrated enrichment for pathways mediating inflammation, immune function and coagulation. Hence, it can be concluded that smoking potentially increases vulnerability to diseases with inflammatory components by perturbing pathways through the alteration of DNA methylation signature of peripheral blood mononuclear cells.

437M

Epigenome-wide meta-analysis of over 10,000 individuals reveals extensive perturbations in DNA methylation associated with adiposity. *A.W. Drong[1], S. Wahl[2,3,4], B. Lehne[5], .M Loh[5], G. Fiorito[6], S. Guarrera[6], S. Kasela[7], R. Richmond[8], A. Dehghan[9], L. Franke[10], T. Esko[7,11,12,13], L. Milani[7], C.L. Relton[8,14,15], J.T. Bell[16], T.D Spector[16], O.H. Franco[9], P. van der Harst[10], C.M. Lindgren[1,17], M.I. McCarthy[1,18], G. Matullo[6], C. Gieger[2,3], J.S. Kooner[19,20,21], H. Grallert[2,3,4], J.C. Chambers[5,20,21].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Institute of Epidemiology II, Helmholtz Zentrum Mnchen - German Research Center for Environmental Health, Neuherberg, Germany; 3) Research Unit of Molecular Epidemiology, Helmholtz Zentrum Mnchen - German Research Center for Environmental Health, Neuherberg, Germany; 4) German Center for Diabetes Research (DZD), Neuherberg, Germany; 5) Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK; 6) Human Genetics Foundation, HuGeF, I-10126 Torino, Italy Department of Medical Sciences, University of Torino, I-10126, Torino, Italy; 7) : Estonian Genome Center, University of Tartu, Tartu, Estonia; 8) MRC Integrative Epidemiology Unit, Oakfield House, University of Bristol, Bristol BS8 2BN, UK; 9) Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 10) Departments of Cardiology and Genetics, University of Groningen, University Medical Center Groningen, Groningen & Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, The Netherlands; 11) Divisions of Endocrinology and Genetics and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts, USA; 12) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA; 13) Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA; 14) Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, NE1 3BZ, Tyne and Wear, UK; 15) School of Social and Community Medicine, University of Bristol, Bristol, UK; 16) Department of Twin Research & Genetic Epidemiology, King's College London, St Thomas' Hospital Campus, Westminster Bridge Road, London SE1 7EH, UK; 17) Broad Institute of Haravrd and MIT, Cambridge, Massachusetts, USA; 18) Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom; 19) National Heart and Lung Institute, Imperial College London, London W12 0NN, UK; 20) Ealing Hospital NHS Trust, Middlesex UB1 3HW, UK; 21) Imperial College Healthcare NHS Trust, London W12 0HS, UK.

Obesity is a global health problem and a major risk factor for type-2 diabetes and cardiovascular disease. The mechanisms underlying obesity and its complications are incompletely understood. We carried out an epigenome-wide association (EWA) study to investigate the relationship of DNA methylation with body mass index (BMI) to examine the epigenetic perturbations associated with obesity. The EWA study included four studies comprising 5,387 whole-blood samples from European (N=2,707) and South Asian (N=2,680) individuals. DNA methylation was determined by Illumina Infinium 450K Human Methylation array. Associations of DNA methylation with BMI were tested in each cohort using linear regression with adjustment for age, gender, physical activity, smoking, alcohol intake, white-blood cell estimates and technical covariates. Meta-analysis across the studies was done by METAL. Our primary EWA yielded CpG sites at 207 independent genetic loci associated with BMI at $P<1\times10^{-7}$, with little evidence for heterogeneity between the populations. We then carried out replication testing of the 207 sentinel CpG sites amongst 4,998 whole-blood samples from 9 independent cohorts. 187 of the 207 markers reached $P<0.05$ in replication testing and remained $P<1\times10^{-7}$ in combined analysis across all stages. A gene set enrichment analysis of the 187 identified genetic loci revealed enrichment for pathways involved in transmembrane transport and metabolic signalling (lipid metabolism, insulin signalling and PPAR signalling) at the 5% FDR level. Further, the 187 sentinel CpGs were located outside CpG Islands and show a significant enrichment for "open sea" locations ($p=4.4\times10^{-9}$), mostly located within gene bodies ($p=2.6\times10^{-4}$). We overlapped the 187 sentinel CpG sites with ENCODE ChiP data peaks and observed an enrichment for DNase hypersensitivity sites ($p=2.4\times10^{-7}$), enhancers ($p=3.5\times10^{-9}$) and the activating histone mark H3K4me1 ($p=3.85\times10^{-12}$). We identify 187 genetic loci at which DNA methylation is associated with BMI. Our results provide new insight into potential DNA regulatory mechanisms associated with obesity and its complications.

438T

Hypomethylation of Synaptic Genes revealed in Dup15q Autism. *K. Dunaway[1,2,3,4], M.S. Islam[1,2,3,4], D.I. Schroeder[1,2,3,4], I.N. Pessah[3,4,5,6], P. Lott[2], M. Meguro-Horike[7], S. Horike[7], R. Chu[1,2], I. Korf[2], J.M. LaSalle[1,2,3,4].* 1) MMI, UC Davis, Davis, CA; 2) Genome Center, UC Davis, Davis, CA; 3) MIND Institute, Sacramento, CA; 4) Center for Children's Environmental Health, UC Davis, Davis, CA; 5) Environmental Toxicology, UC Davis, Davis, CA; 6) Molecular Biosciences, UC Davis, Davis, CA; 7) 7Advanced Science Research Center, Kanazawa University, 13-1 Takaramachi, Kanazawa 920-0934, Japan.

Chromosome 15q11-13 duplication syndrome (Dup15q) is one of the most common copy number variations observed in autism-spectrum disorders. Surprisingly, a prior analysis of Dup15q human brain samples showed both DNA hypomethylation measured by LINE-1 pyrosequencing and significantly higher levels of the persistent organic pollutant PCB 95 than controls or idiopathic autism cases. In order to take an unbiased view of the whole methylome, including genic, intergenic, and repetitive regions, we performed MethylC-seq and developed novel bioinformatic analyses. Six Dup15q and matched control cortical (BA19) postmortem samples were used to determine methylation patterns in Dup15q through custom hidden Markov models, demonstrating large-scale hypomethylated domains throughout the genome. We also analyzed percent methylation in repetitive regions, which are normally neglected due to the complexity of their analyses. Using a bioinformatics toolkit we created, LINE-1 hypomethylation in Dup15 brain samples compared to controls was confirmed and analysis of additional repetitive elements is underway. To experimentally model the interaction of Dup15q and PCB 95, human SH-SY5Y neuroblastoma cells containing an additional maternal chromosome 15 (SH-15M) were assayed using MethylC-seq. SH-15M cells exhibited a 2–14% increase of hypomethylated domains spanning across every autosome compared to the parental SH-SY5Y cell line, an effect independent of PCB-95 exposure. Genes within the hypomethylated regions specific to SH-15M were enriched for functions at the postsynaptic cell membrane while those genes in hypomethylated regions specific to PCB 95 exposure were enriched for ion channels, neurotransmitter, and synaptic functions. Transcript levels of 48 autism candidate genes found in these hypomethylated regions were compared them across 48 different cell lines, single-cell clones, and postmortem brain samples by quantitative RT-PCR in the Fluidigm Biomark. Aberrant transcript levels were observed for multiple glutamate, serotonin, and GABAA receptor genes, with a significant compounding effect of SH-15M and PCB 95 for GRIA1. In summary, we found large-scale epigenomic changes to the brain methylome in Dup15q syndrome that can be modeled in cell culture and further compounded by PCB 95 exposure, altering synaptic gene transcript levels. These results have implications for understanding and treating complex gene-environment interactions in autism-spectrum disorders.

439M

Genome-wide DNA methylation study identifies genes associated with GDF-15 levels. *W.E. Ek[1], A.K. Hedman[2,4], S. Enroth[1], S. Gustafsson[2,4], E. Ingelsson[2,4], U. Gyllensten[1], L. Lind[3], A. Johansson[1,4].* 1) Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Uppsala, Sweden; 2) epartment of Medical Sciences, Molecular epidemiology, and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 3) Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University Hospital, Uppsala, Sweden; 4) Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden.

Introduction. Growth differentiation factor 15 (GDF-15) is a member of the transforming growth factor-β (TGF-β) family. Under physiologic conditions, GDF-15 is expressed in low to moderate levels in most healthy tissues and levels may increase in response to pathological stress associated with inflammation, oxidative stress or tissue damage. Overexpression of GDF-15 has been seen in various cancers and growing evidence indicate that GDF-15 levels in plasma may be a new biomarker for risk stratification and therapeutic decision making in cardiovascular disease.

Methods: We performed a genome-wide DNA methylation study to determine the association between GDF-15 and DNA methylation in 717 individuals from the Northern Sweden Population Health Study (NSHPS). DNA methylation status, in white blood cells, was determined at more than 475,000 sites distributed throughout the genome. Significant findings where replicated in 950 individuals from an independent cohort (PIVUS). We also performed gene ontology (GO) enrichment analysis to identify terms enriched for differently methylated genes in relation to GDF-15 levels.

Results: A total of 31 CpG sites, corresponding to 22 genes, were significantly associated (false discovery rate [FDR] q-values < 0.05) with GDF-15 in NSHPS. Of these, 12 sites replicated in PIVUS ($p < 0.05$). One of the replicating sites is located in the promoter region of MIR21, which encodes one of the most well studied microRNAs that has been associated with cancer and cardiovascular diseases. A total of 125 enriched biological processes were identified (FDR q-value < 0.05), of which 43 (34 %) replicated ($p < 0.05$). Among the enriched biological processes, we found terms involved in developmental, biological and metabolic processes. A total of 13 enriched molecular functions were also identified (FDR q-value < 0.05), of which 5 (38 %) replicated ($p < 0.05$). Among molecular functions, we found terms involved in protein binding and kinase activity.

Conclusion: We have shown that GDF-15 levels are correlated with DNA methylation level at numerous sites in the genome providing new leads for investigating the links between GDF-15 and cardiovascular disease. However, our results suggest that GDF-15 plays an important role in many biological processes, even in the absence of response to pathological stress.

440T

Black-white difference in regional patterns of DNA methylation: the Bogalusa Heart Study. *X. Fu[1], D. Sun[2,3], S. Li[2], C. Fernandez[2], T. Chen[4], Y. Lian[4], Q. Li[4], W. Chen[2].* 1) Department of Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA; 2) Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA; 3) Department of Epidemiology, School of Public Health, Peking University Health Science Center, Beijing, China; 4) Microarray Core Facility, University of Texas Southwestern Medical Center, Dallas, TX, USA.

Although DNA methylation profiles vary by populations, regional patterns of DNA methylation in different populations are largely unknown. The objective of this study is to examine black-white difference in regional patterns of DNA methylation in relation to pathophysiological pathways. This study included 846 adults aged 28–51 years (594 whites and 252 blacks) from the Bogalusa Heart Study. Peripheral leukocyte DNA methylation was measured with Infinium HumanMethylation450 BeadChip. After sample and probe quality analysis and cell-type count estimation, a linear regression model with age, gender, cell-type count included as covariates was fitted to find the significantly differential methylation sites between races. To avoid bias towards high probe density region, a feature-oriented dynamic window method "lasso" was used to capture significantly differential regions. Enrichment analysis in KEGG pathways was performed with selected differential methylation-related gene sets. Besides, correlation patterns of methylation were examined by race groups, followed by a comparison with genetic LD structure. A total of 1,225 regions were captured among 48,241 sites that were found to be significantly, differentially methylated between blacks and whites. Enrichment analysis using genes located in these 1,225 regions identified 15 pathways in the KEGG database which were related to human complex diseases known to have different prevalence rates among populations. In correlation analysis, significant differences in both position and size of clusters were observed in every chromosome, especially in chromosome 2, 4, 10, 13 and 22, between blacks and whites. These correlation patterns were not all caused by genetic LD structures. In conclusion, the findings of the current study on black-white difference in regional patterns of DNA methylation has implications in understanding of potential race-specific epigenetic mechanisms and pathophysiological pathways underlining chronic diseases.

441M

Identification of epigenetic signatures of childhood socioeconomic-related adversity. *S.J. Goodman[1,2], W.T. Boyce[3,4,5], S. Lam[4], J.L. MacIsaac[1], M.J. Jones[1], S.M. Mah[1], A. Zaidman-Zait[4], M.S. Kobor[1,2].* 1) Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Vancouver, British Columbia, Canada; 2) Medical Genetics, University of British Comlubia, Vancouver, British Columbia, Canada; 3) School of Population and Public Health, University of British Comlubia, Vancouver, British Columbia, Canada; 4) Human Early Learning Partnership, University of British Comlubia, Vancouver, British Columbia, Canada; 5) Pediatrics, UCSF School of Medicine, San Francisco, California, USA.

Rationale: Socioeconomic status (SES) is a potent yet enigmatic environmental determinant of human health. Childhood experiences of SES are known to alter development and well-being and cause health disparities later in life, suggesting that they are somehow 'biologically embedded'. Gene Expression Collaborative for Kids Only (GECKO) was established to study the mechanisms of biological embedding in socioeconomically diverse children. An important part of the study is investigating how DNA methylation may interact with both the environment and DNA variants, thereby stably imprinting genetic and environmental information into the epigenome. We hypothesize that socially driven developmental trajectories in children are mediated by the interactions of environment, epigenetics and genetic variants, which ultimately result in phenotypic differences and health disparities later in life. **Methods:** Using the Illumina Human Methylation 450K array platform, we have examined genome-wide DNA methylation of buccal epithelial cells (BECs) in 400 8–10 year old individuals from the GECKO cohort. For a subset of participants, whole blood and newborn dried blood spot (DBS) samples were obtained to compare BEC methylation patterns to those present in peripheral blood mononuclear cells (PBMCs) and neonatal blood from each child. Genetic variants were evaluated by running saliva DNA from each participant on the Illumina PsychArray BeadChip. **Results and implications:** We interrogated epigenomic correlates of childhood experiences and exposures to identify a set of loci showing experience-dependent epigenetic modifications. A differential analysis of methylation by SES, as measured by family income, parent occupation and education, produced a list of candidate loci which may be involved in biological embedding. Replication of this analysis on the DBSs and PBMCs reinforced the tissue-specific nature of DNA methylation. Unsupervised clustering of the paired DBS and PBMC samples showed that the methylation of two unrelated individuals in a single tissue is more alike than the methylation of two tissue types from one individual. Candidate methylation loci will be interrogated for DNA variants to identify potential risk alleles which are more prone to methylation disturbances by SES. Understanding the molecular and epigenetic mechanisms by which early adversity and SES lead to disturbances in health is a fundamental first step in elucidating the origins of health disparities.

442T

Methylation patterns are associated with chronological age in the Khomani San of South Africa. *S. Gopalan[1], O. Carja[2], E. Patin[3], M.S. Kobor[4], H. Fraser[2], M. Feldman[2], A. Froment[5], L. Quintana-Murci[3], B. Henn[1].* 1) Ecology and Evolution, Stony Brook University, Stony Brook, NY, United States; 2) Department of Biology, Stanford University, Stanford, CA, United States; 3) Department of Genomes and Genetics, Human Evolutionary Genetics, Institut Pasteur, Paris, France; 4) Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada; 5) Musum National d'Histoire Naturelle, Paris, France.

Changes in genomic methylation have been found to be associated with advancing age. Increased age is correlated with an overall genome-wide reduction in methylation, but also an increase in methylation of particular regions, such as CpG islands. So far, studies of methylation as it relates to aging in humans are primarily done in populations of European origin. It remains challenging to control for the effects of extrinsic factors, for example diet and smoking, which are known to be substantial.

We set out to identify CpG sites associated with aging in the Khomani San, a population of former hunter-gatherers living in the Kalahari Desert of South Africa. Previous studies have found that the KhoeSan populations are the most genetically diverse in the world and are among the few remaining societies that have not adopted an agriculturalist lifestyle. The Baka Pygmies are another group of African hunter-gatherers who live in the rainforests of Central Africa. Despite sharing similar subsistence strategies, and societal circumstances of poverty and malnutrition, the age structure of their populations differs greatly. It is common for Khomani San individuals to reach to age of 70 or more, while Baka Pygmies rarely reach the age of 50. We investigated the methylation profiles of individuals in both groups and compared them to European methylation profiles. Saliva samples from 56 Khomani San and 36 Baka individuals were methyltyped on the Illumina 450K array. Khomani individuals ranged from 21 to 91 years old (median 62.5) whereas Baka individuals ranged from 5 to 59 years old (median 29.5). We performed an epigenome-wide association study (EWAS) to identify sites that were significantly associated with chronological age in the Khomani San individuals. This resulted in over 100 CpG sites. We replicated the strong signal of hypermethylation in the CpG island of the *ELOVL2* gene, which was recently identified as a biomarker for aging in blood cells (Garagnani et al., 2012). These results demonstrate that some key sites of methylation are associated with aging in diverse human populations. Our results may prove useful for forensic work as well as future epidemiological studies.

Garagnani, P., et. al (2012), Methylation of ELOVL2 gene as a new epigenetic marker of age. Aging Cell, 11: 1132–1134.

443M

Characterizing a genomic map of 5-hydroxymethylcytosine in human brain at single base resolution through next-generation sequencing. *J. Gross, G.G. Chen, A. Diallo, R. Poujol, C. Ernst, G. Turecki.* Douglas Mental Health Univ Inst, Montreal, Quebec, Canada.

The recent discovery that methylated cytosines are converted to hydroxymethylated cytosines (hmC) by the family of ten-eleven translocation enzymes has sparked significant interest in the field of epigenetics. This finding, along with that of Kriaucionis and Heintz who described the presence of hmC in purkinje neurons, stimulated growing interest in research describing the genomic location, the abundance in different tissues, and the putative functions of hydroxymethylation. To date, there is a lack of reference maps of hmC in post-mortem human brain tissue, which makes challenging to understand the functional role of hmC. To characterize hmC in human brain, DNA from post-mortem brain tissue from 24 subjects who died by natural causes was extracted, glucosylated, and digested with AbaSI, an enzyme that specifically recognizes glucosylated cytosines. Subsequently, custom biotinylated adaptors were ligated to the cleaved DNA, which was then further sheared. Standard libraries were prepared using the enriched fraction and were sequenced on Illumina's HiSeq 2000 sequencers. Bowtie 2.0 was used to map raw reads to the reference genome Hg19 and a custom perl script was used to determine the locations of hmC at single base pair resolution. A combination of BEDtools, R packages, and custom scripts were used to determine enriched and depleted regions in the genome, links to regulatory elements, and correlations with available gene expression data. The results present a unique characterization of hmC in human brain and provide an important reference for future research. Understanding the differences in the genomic locations of hmC will shed light on the growing debate as to whether hmC represents a novel epigenetic mark or whether it is simply an intermediate product of active DNA demethylation.

444T

Global DNA hypermethylation is associated with increased myopia risk. *E. Hsi[1,2], K.C. Chen[3], C.W. Huang[4], M.L. Yu[5,6], C.L. Liang[7], S.S. Juo[2,8].* 1) Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan; 2) Department of Medical Research, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan; 3) Department of Biochemistry, College of Medicine, Taipei Medical University, Taipei, Taiwan; 4) Department of Biotechnology, Kaohsiung Medical University, Kaohsiung, Taiwan; 5) Hepatobiliary Division, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan; 6) Faculty of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan; 7) Bright-Eyes Clinic, Kaohsiung, Taiwan; 8) Department of Genome Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan.

Purpose: DNA methylation is the most extensively studied epigenetic process that modulates gene expression without changing DNA sequence and is known to be altered by environmental factors. We aimed to discover the association between global DNA methylation and high myopia and find the relevant consequences. Methods: The methylated levels of long interspersed nucleotide elements (LINE-1) are regarded as a surrogate marker of global DNA methylation. Three hundred high myopia and 300 sex-age matched control subjects were used to find the association between high myopia and LINE-1 methylation in leukocytes. We analyze the global DNA methylation among leukocytes, retina and sclera tissues in form deprivation myopia (FDM) mice model. Immunochemistry demonstrates the distribution of homocysteine and methionine. We use dopamine to mimic brighter environment to rescue the myopia state. Results: Subjects with high (82.1% - 91.5%) and middle tertile (79.6% - 82.1%) of LINE-1 methylation have 2.10 and 2.08 fold risk to have high myopia, significantly (p=0.002 and 0.001, respectively). Significantly elevated LINE-1 methylated level in blood and both retina and sclera tissues of both eyes in FDM mice. Immunohistochemical staining indicates elevated homocysteine, methionine levels in both FDM and fellow eyes. Finally, the global methylated level is reduced 2% significantly after 24 hours of dopamine stimulation. Conclusions: Our study demonstrates that LINE-1 DNA hypermethylation is associated with high myopia. Global hypermethylation and accumulated homocysteine indicate a constitutional change which is induced by FDM. Reduced global methylated level is observed after treatment of dopamine and is a novel mechanism in treatment of myopia.

445M

Micoarray approach reveals the differentially methylated regions as potential epigenetic markers in preeclampsia. *H.J. KIM[1], S.Y. KIM[1], S.Y. PARK[1], D.J. KIM[1], D.E. LEE[1], J.H. LIM[1], M.Y. KIM[2], J.H. CHUNG[2], H.M. RYU[1, 2].* 1) Laboratory of Medical Genetics, Cheil General Hospital and Women's Healthcare Center, Seoul, Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women's Healthcare Center, Kwandong University College of Medicine, Seoul, Korea.

Objective: The development of preeclampsia (PE) seriously affects the health of the mother and the child, but the precise pathogenesis of PE remains elusive. The placenta is considered to play a key role and DNA methylation may be associated with altered placental development and function. There have been few known fetal-specific marker using the differences between maternal and fetal epigenetic characteristics. We analyzed methylation status of maternal blood and fetal placental DNA using microarray to find a differentially methylated region as potential epigenetic markers for PE. **Methods:** We performed a high-resolution tiling array analysis using a methyl-CpG binding domain-based protein (MBD) method with blood samples from normal pregnancies, placenta samples from normal pregnancies, and blood samples from non-pregnancies. To validate the differentially methylated loci identified from a high-resolution tiling array, bisulfite direct sequencing was carried out in four groups: 1) blood from normal pregnancies, 2) placenta from normal pregnancies, 3) blood from pregnancies with PE, and 4) placenta from pregnancies with PE. **Results:** Differentially hypermethylated loci in the placenta were selected by Agilent Genomic Workbench software using an arbitrary differential ratio cut off of normalized log2 ratio > 2. We confirmed the methylation patterns of CpG sites in the 5 associated regions of selected loci using bisulfite direct sequencing. In the normal and PE group, all selected region showed significant hypermethylation in placenta compared with blood, respectively (P=0.004 for SOD1, P<0.00001 for DSCR3, P=0.0004 for C2CD2, P=0.00003 for UMODL1, P=0.00002 for ENST00000450830). Interestingly, the methylation ratio in each DSCR3 CpG site was significantly increased in PE placental samples compared with normal placental samples (P<0.05 for all CpG sites). **Conclusion:** We found 5 fetal-specific hypermethylated regions. Among these genes, DSCR3 showed significantly aberrant DNA methylation in PE, which may serves as a potential candidate to be developed into epigenetic biomarkers in PE.

446T

Quantification of the placental epigenetic signature of the maspin gene in maternal plasma of pregnancies complicated by small for gestational age. *S.Y. Kim[1], H.J. Kim[1], S.Y. Park[1], D.E. Lee[1], J.H. Lim[1], S.Y. Lee[1], K.S. Kim[1], J.Y. Han[2], M.H. Kim[2], D.W. Kwak[2], H.M. Ryu[1,2].* 1) Laboratory of Medical Genetics, Cheil General Hospital and Women's Healthcare Center, Seoul, Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women's Healthcare Center, Kwandong University College of Medicine, Seoul, Korea.

Objective: Quantification of cell-free fetal DNA (cffDNA) by methylation-based DNA discrimination has been used in non-invasive monitoring of adverse pregnancy outcomes. The maspin (serpin peptidase inhibitor, clade B (ovalbumin), member 5; SERPINB5) gene is hypomethylated in the placenta and completely methylated in maternal blood cells. The objective of this study was to examine the possible association between women with small for gestational age (SGA) neonate and cffDNA concentration in maternal plasma during first and second trimesters using tissue-specific epigenetic characteristics of the maspin gene. Methods: We performed a real-time quantitative methylation-specific PCR to quantify the concentrations of unmethylated-maspin (U-maspin) and methylated-maspin (M-maspin) in maternal plasma at 11–26 gestational weeks of women with SGA neonates (n = 55) and with appropriate for gestational age (AGA) neonates (n = 106). Results: At 11–14 gestational weeks, median U-maspin and M-maspin concentrations in women with SGA neonates did not differ significantly from those of women with AGA neonates. There were also no significant differences in the concentrations of U-maspin and M-maspin between the two groups at 15–26 weeks of gestation. The ratio {[U-maspin/(M-maspin+U-maspin)]X100} in women with SGA neonates was similar to that of controls. In addition, statistically significant correlations between U-maspin or M-maspin concentration at first and second trimesters and birth weight and gestational age at delivery were not observed. Conclusion: Concentrations of U-maspin and M-maspin in maternal plasma of the first and second trimesters were not associated with pregnant women who delivered SGA neonates. Our findings indicate that quantification of cffDNA in maternal plasma during early pregnancy using tissue-specific methylation changes of the maspin gene promoter is not a useful predictor of SGA pregnancies.

447M

Methylation levels in peripheral blood may reflect central mechanisms mediating GxE associations with cardiovascular outcomes. *L.C. Kwee[1], C.S. Haynes[1], E.A. Grass[1], I.C. Siegler[2,3], B.H. Brummett[2,3], M.A. Babyak[2,3], R. Jiang[2,3], A. Singh[2,3], R.B. Williams[2,3], E.R. Hauser[1], W.E. Kraus[1,4], S.H. Shah[1,4], S.G. Gregory[1].* 1) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC; 2) Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC; 3) Behavioral Medicine Research Center, Duke University Medical Center, Durham, NC; 4) Division of Cardiology, Department of Medicine, Duke University Medical Center, Durham, NC.

Previous studies have established associations between candidate genes, psychosocial factors, and cardiovascular (CV) outcomes. For example, a genetic variant (rs6318) in the serotonin receptor gene *HTR2C* is associated with both hypothalamic-pituitary-adrenal response to a stress recall task and increased risk of CV mortality. Also, SNPs in the *TOMM40/APOE* region are associated with triglyceride levels, and this effect is moderated by chronic stress. One model for how psychosocial factors affect CV outcomes is that epigenetic changes, such as alterations in methylation status, may mediate the relevant effects. However, it is difficult or impossible to assay epigenetic markers in the tissues that may mediate these effects, such as brain and heart. For cardiovascular disease, however, blood borne factors are in contact with vascular tissue and are of considerable interest. Here, we explore whether methylation levels in peripheral blood provide information to dissect the interaction of psychosocial factors, genetics, and CV outcomes. DNA was obtained from 206 subjects selected from the CATHGEN biorepository of patients referred for cardiac catheterization at Duke University. We selected candidate genes previously associated with coronary heart disease endophenotypes and outcomes (*HTR2C, TOMM40/APOE, ADRB2, BDNF*, and *EBF1*) via interactions with psychosocial stress, and assayed methylation at CpG sites in these genes using the Infinium HumanMethylation450 BeadChip. After QC, data from 169 CpGs in these five regions were available for analysis. We used linear mixed models to test for association between methylation at each probe and an endpoint composed of all-cause mortality or incident myocardial infarction. All models were adjusted for age, sex, race, and batch. Ten CpG sites were nominally associated ($p<.05$) with the endpoint: six in *BDNF* (brain-derived neurotrophic factor; min p=9.1×10–4), and four in *EBF1* (early B-cell factor 1; min p=8.5×10–4). (Due to the correlation in methylation across a gene, strict Bonferroni correction is overly conservative.) We observed consistent methylation changes, as all but one of the significant sites were methylated at slightly higher levels in controls vs. cases ($|\Delta\beta|<.03$ for all). These results indicate that it may be useful and relevant to assay peripheral blood biomarkers in the search for genetic mechanisms that mediate the central effect of psychosocial stress factors on cardiovascular events.

448T

Genome-wide Association Study of DNA-methylation identifies 9.8 million SNP-methylation associations and provides insight into global regulatory pathways. *B. Lehne[1], W. Zhang[1,2], M. Loh[1], J.C. Abbott[3], S. Burbidge[4], A. Drong[5], W.R. Scott[1,6], S.-T. Tan[2,6], U. Afzal[1,2], J. Scott[6], M.-R. Jarvelin[1,7,8,9,10], M.I. McCarthy[5,11], P. Elliott[1], J.C. Chambers[1,2,12], J.S. Kooner[2,6,12].* 1) Epidemiology and Biostatistics, Imperial College London, London, UK; 2) Ealing Hospital NHS Trust, Middlesex, UK; 3) Centre for Investigative Systems Biology and Bioinformatics, Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College London, London, UK; 4) Information and Communication Technologies, Imperial College London, London, UK; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 6) National Heart and Lung Institute, Imperial College London, London, UK; 7) Institute of Health Sciences, P.O.Box 5000, FI-90014 University of Oulu, Finland; 8) Biocenter Oulu, P.O.Box 5000, Aapistie 5A, FI-90014 University of Oulu, Finland; 9) Unit of Primary Care, Oulu University Hospital, Kajaanintie 50, P.O.Box 20, FI-90220 Oulu, 90029 OYS, Finland; 10) Department of Children and Young People and Families, National Institute for Health and Welfare, Aapistie 1, Box 310, FI-90101 Oulu, Finland; 11) Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, UK; 12) Imperial College Healthcare NHS Trust, London, UK.

The mechanisms linking DNA sequence variation to phenotypic patterns are incompletely understood. Methylation of genomic DNA at CpG sites regulates gene expression and cellular function. We carried out large-scale Genome-wide Association (GWA) studies to investigate whether the relationship between DNA sequence variation and phenotype may be mediated by DNA methylation. Methylation levels were measured at CpG dinucleotides in genomic DNA from peripheral blood of 1,844 individuals using the Illumina Infinium 450K methylation array. We excluded cross-hybridising markers and markers containing known SNPs within their probe-sequence, resulting in 370,394 CpG sites. Methylation levels were quantile normalised and adjusted for technical biases, gender, age and white-blood cell subpopulations. We generated high density genotype data for the same individuals, including imputation based on 1000 Genomes reference panels. We excluded SNPs with low call rate (<98%), minor-allele-frequency (<1%), Hardy-Weinberg Equilibrium $P<10^{-6}$ and imputation quality (info-score<0.5). This left 7,859,898 SNPs for analysis. Using a highly-parallelized approach we then tested the 7,859,898 SNPs for association with the 370,394 DNA methylation markers, resulting in $\sim 2.9\times10^{12}$ statistical tests. We find 9,883,723 SNP-CpG pairs with evidence for association at $P<10^{-14}$ (corresponding to $P<0.05$ after Bonferroni correction for the $\sim 2.9\times10^{12}$ tests performed). For 9,027,340 associations the SNPs are located in *cis* (within 1mb of the CpG site), while for 856,382 association the SNPs are located in *trans* (SNP-CpG distance >1mb). SNPs associated with CpGs are highly enriched for variants associated with Height, Lipid Levels, Inflammatory Bowel Disease, Multiple Sclerosis and other traits identified by GWA studies. Furthermore, CpGs influenced by genetic variants are enriched for regulatory features such as enhancers and DNase hypersensitive sites. The trans-associations identify multiple SNPs that have a widespread impact on DNA-methylation across multiple regions of the genome, including variants in well-known regulatory genes such as *CTCF*, *NFKB* and the *MHC* region. In conclusion, we report over 9.8 million associations between common genetic variants and methylation of genomic DNA, with evidence for functional enrichment. Our data provides new insight into molecular pathways by linking DNA sequence variation to human phenotype and disease.

449M

Birth weight and DNA methylation. *S. Li[1], D. Sun[1,2], X. Fu[3], C. Fernandez[1], T. Chen[4], Y. Lian[4], Q. Li[4], W. Chen[1].* 1) Epidemiology, Tulane School of Public Health and Tropical Medicine, New Orleans, LA; 2) Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China; 3) Department of Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA; 4) Microarray Core Facility, University of Texas Southwestern Medical Center, Dallas, TX.

Low birth weight for gestational age is a risk factor for adult cardiometabolic diseases. It has been hypothesized that intrauterine growth retardation may result in changes in DNA methylation, by which low birth weight leads to increased risk of cardiometabolic diseases. However, it is not known what specific DNA methylation changes occur as a result of intrauterine growth retardation as indicated by low birth weight. We aimed to examine the association between birth weight and DNA methylation profiles in adults. The study included 679 white and 289 black adult participants aged 28–51 years of the Bogalusa Heart Study who had birth weight data and DNA methylation data (the Infinium HumanMethylation450 BeadChip). After quality control, beta values from 420200 CpG sites were used as outcome variables and birth weight as an independent variable in linear regression models, adjusted for age, sex, cigarette smoking, and body mass index, separately in whites and in blacks. In whites, beta values of ten CpG sites were significantly (false discovery rate <0.05) associated with birth weight, of which three were significant after Bonferroni correction ($P<10^{-7}$). These significant sites are in or near genes EBF4, RAMP2, ARL4C, TBCD, PRRT1, and BAI1, representing cellular component organization or biogenesis, developmental process, immune system process, localization, metabolic process, multicellular organismal process, reproduction, and response to stimulus, in PANTHER analysis. In blacks, only one CpG site was significant ($P<10^{-7}$), with no clear functional implications. We are seeking replications in independent samples and will present updated results in the meeting. In conclusion, we have identified CpG sites whose methylation is associated with birth weight, indicating that prenatal growth environment affects DNA methylation profiles.

450T

Maternal age-related changes in DNA methylation in newborns and adults. *C.A. Markunas[1], A.J. Wilcox[1], Z. Xu[1], B.R. Joubert[1], S. Harlid[2], S.E. Hberg[3], W. Nystad[3], S.J. London[1], D.P. Sandler[1], P.A. Wade[2], R.T. Lie[4,5], J.A. Taylor[1,2].* 1) Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, RTP, NC, USA; 2) Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, NIH, RTP, NC, USA; 3) Norwegian Institute of Public Health, Oslo, Norway; 4) Department of Global Public Health and Primary Care, University of Bergen, Norway; 5) Medical Birth Registry of Norway, Norwegian Institute of Public Health, Bergen, Norway.

Older maternal age during pregnancy has been associated with adverse birth outcomes, childhood cancer, type 1 diabetes, and neurodevelopmental disorders in offspring. For many of these conditions the underlying biologic mechanism is unknown. One mechanism by which maternal age may affect the health of the offspring is through epigenetic modifications such as DNA methylation. Using the Norway Facial Clefts Study, a national population-based case-control study of cleft lip and cleft palate, we conducted the largest epigenome-wide association study to date investigating alterations in DNA methylation in newborns related to maternal age at delivery. The Illumina HumanMethylation450 BeadChip was used to assess whole blood DNA methylation in samples collected 2–3 days after delivery from 890 newborns. After data pre-processing, robust linear regression was used to identify CpG sites related to maternal age, adjusting for facial cleft status, birth weight, maternal alcohol use, maternal smoking, maternal education, parity, sex, and technical factors (batch, bisulfite conversion efficiency, and birth year). Additional adjustment for five blood cell subtypes following the method developed by Houseman and colleagues was performed as a sensitivity analysis. Replication of select findings is underway in an independent pregnancy cohort, the Norwegian Mother and Child Cohort Study, using Illumina HumanMethylation450K data from 1068 umbilical cord blood samples. In addition, using data from the Sister Study, a nationwide prospective US cohort of women with a sister with breast cancer, we are exploring whether maternal age-related methylation changes persist into adulthood. The adult analysis includes 1006 women with Illumina HumanMethylation27K data and 181 women with Illumina HumanMethylation450K data generated from whole blood. Identification of the same maternal-age related DNA methylation changes in adults would suggest that these altered methylation states persist for an extended period of time, which may have implications for later health outcomes.

451M

Genetic control of the human blood methylome. *J.L. McClay[1], A.A. Shabalin[1], D.E. Adkins[1], G. Kumar[1], S. Nerella[1], S.L. Clark[1], S.E. Bergen[2], C.M. Hultman[2], P.K.E. Magnusson[2], P.F. Sullivan[3], K.A. Aberg[1], E.J. van den Oord[1], Swedish Schizophrenia Consortium.* 1) Center for Biomarker Research and Personalized Medicine, Virginia Commonwealth Univ, Richmond, VA; 2) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 3) Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC.

Comprehensive understanding of genetic influence on the epigenome could shed new light on human genotype-phenotype relationships. In this study, we precision map genome-wide methylation quantitative trait loci (meQTLs) in 697 Swedish subjects. Using SOLiD next-generation sequencing (Life Technologies), we generated an average of 67.3 million methylation-enriched reads per subject to assay whole blood DNA methylation levels at ~27 million autosomal CpGs. We then correlated the resulting methylation measures with 4.5 million SNPs, genotyped on Affymetrix 5.0/6.0 or Illumina Omni quad arrays, with imputation using 1000 Genomes panels. Testing used Matrix eQTL to run all >20 trillion tests in 1.5 days on 50 nodes of a computing cluster. Using stringent (1%) false discovery rate control, 15% of methylation sites showed genetic influence, while 98% of SNPs were associated with methylation at one or more sites. The high proportion of SNP effects was due to linkage disequilibrium (LD), whereby many SNPs tagged each meQTL, whereas long-range correlation between methylation sites was much less extensive. Local meQTL effects (≤ 1Mb between SNP and methylation site) were 189,000-fold more common than distant effects. Most local effects could be explained by CpG-SNPs (SNPs altering CpG sequence), with 75% of methylation sites under local genetic influence having a CpG-SNP (MAF > 0.05) within 250bp, compared to 33% of all sites genome-wide. Local meQTL effects typically occurred outside known genes and were not enriched for several classes of genomic annotations (e.g. CpG islands, DNase clusters). However, these sites were strongly enriched (odds ratio=3) for reported disease-associated loci from the NHGRI GWAS catalog (www.genome.gov/gwastudies). Of the 393 methylation sites under local genetic control that overlapped the GWAS catalog, 366 encompassed a CpG-SNP. Our study confirms that genetic influence on DNA methylation is pervasive, but effects are highly localized. Most meQTLs appear unlikely to affect disease susceptibility, because they typically occur outside of known functional regions. However, mutation of CpGs at critical regulatory loci may be an important etiological mechanism for complex diseases. Our study provides a detailed map of genetic influence on the human blood methylome. Integration of meQTL and functional annotation data into gene mapping efforts could help pinpoint potentially causative mutations in large genomic regions.

452T

DNA methylation profiles of ten patients with Attention deficit hyperactivity disorder: a proof-of-principle study. *C. Milani[1], T.V.M.M. Costa[1], M.M. Montenegro[1,2], G.M. Novo-Filho[2], E.A. Zanardo[2], R.L. Dutra[2], A.T. Dias[1], F.B. Piazzon[1], A.M. Nascimento[2], V. Schuch[3], C.B. Mello[3], M. Muszkat[3], C.A. Kim[1,2], L.D. Kulikowski[1,2].* 1) Departamento de Patologia, Laboratrio de Citogenmica, FMUSP, Sao Paulo, Brazil; 2) Unidade de Gentica, Departamento de Pediatria, Instituto da Criana, FMUSP, Sao Paulo, Brazil; 3) Departamento de Psicobiologia, Universidade Federal de So Paulo, Sao Paulo SP, Brazil.

Introduction: Attention deficit hyperactivity disorder (ADHD) is one of the most common childhood brain disorders and affects approximately 5.3% of children worldwide. Despite the high heritability of the disorder (estimated at 76%), no genetic marker has been consistently identified and the effects on the variability of the ADHD phenotype calls for improvements in research strategies. We performed the proof of principle study in order to characterize DNA methylation profile in peripheral blood leukocytes of individuals with ADHD, providing the first epigenome view study in this disorder. Methods: Genome-wide methylation profile was assessed using DNA extracted from blood lymphocytes of 10 patients with ADHD. We used the Illumina Infinium HumanMethylation450 BeadChip array, which allows the annotation of approximately 480,000 CpG sites per sample at single-nucleotide resolution. Results: The results of all patients showed similar methylation profile and biostatistical tools identified sites with an elevated probability to be consistently hypermethylated in chromosome 13 and 16 and hemimethylated in chromosome 18, 19 and 20. However, some patients revealed slight different methylation profiles involved *BRI3BP*, *RPH3AL*, *ZNF516* and *TP53TG5* genes suggesting an involvement in process as apoptosis mediated by *TNF*, regulation of calcium ions, cell proliferation, differentiation and tumor suppression. In addition, some hypomethylated relevant genes were found in our samples (*NRD1*, *AK3L1*, *RFX5*, *LDLRAP1*, *FBLIM1*, *SYT2*, *BTBD8*, *ZBTB17*, *FITM2*, *TMEM184B*, *CPNE1*, *SPTBN4*). Others were hypermethylated (*ZBTB46*, *OPRL1*, *ZNF8*, *NDUFS7*, *CTDP1*, *TBCD*, *MED24*, *ANK53*, *SPG7*, *ITPK1*, *TPCN2*). In a early conclusion, this proof of principle study showed the DNA methylation status of 96% of CpG islands located in the genome in promoter regions and non-promoter regions of genes. Profile the tissue-specific DNA methylation patterns will provide novel insights into pathogenic mechanisms, as well as help in future epigenetic therapies.

453M

Age-related DNA methylation profiling in CD4 and CD8 T-cells reveals changes in infection-associated and lineage regulator genes. *L. Milani[1], L. Tserel[2], R. Kolde[3], M. Limbach[2], K. Tretyakov[3], S. Kasela[1,4], K. Kisand[2], M. Saare[2], J. Vilo[3], A. Metspalu[1,4], P. Peterson[2].* 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Molecular Pathology, Institute of Biomedical and Translational Medicine, University of Tartu, Tartu, Estonia; 3) Institute of Computer Science, University of Tartu, Tartu, Estonia; 4) Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia.

Ageing is associated with gradual deterioration of immune response, for example, elderly people have lower responses to vaccination and higher risks for infections, autoimmune diseases and tumorigenesis. Although many blood cell types are involved, the most affected cells in this process are T-cells. The reshaping of T-cells during ageing is considered to be most important feature of immunosenescence and causes an immune risk profile associated with higher morbidity and mortality among elderly individuals.

We analyzed the DNA methylation profiles of CD4 and CD8 T-cells and peripheral blood samples from 50 young and 50 elderly individuals to identify the epigenetic basis of these changes and to investigate their association with corresponding gene expression levels.

In DNA methylation analysis of CD4 and CD8 T-cells, we found more differentially methylated CpG sites and larger variability among CD8 T-cells, whereas a relatively low number of changes were detectable in peripheral blood samples after correction for cell subtype proportions. Age-related hypermethylation occurred primarily in the CpG islands of gene promoters with silenced transcriptional activity and was associated with the repressive histone H3K27me3 modification based on ENCODE data. More specifically, we identified a subset of functionally important genes in T-cells, with strong inverse correlation between methylation and expression levels. Among this subset, we found hypomethylation and higher expression of effector T-cell and infection-associated genes as well as hypermethylation of T-cell lineage-specific transcription factor genes in elderly individuals.

Our results argue that although many age-associated DNA methylation changes are shared among blood cells, they do not comprise all the changes that occur in specific blood-cell subsets, such as the T-cells described here. Our findings demonstrate links between age-related epigenetic changes and gene expression in CD4 and CD8 T-cells, and these changes are relevant to the response to infections and to potential harmful consequences leading to immunosenescence.

454T

Influence of maternal nutrition on ancestry-dependent gene methylation in newborns. *K. Mozhui[1], A.K. Smith[2], F.A. Tylavsky[1].* 1) Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN; 2) Department of Psychiatry and Behavioral Science, Emory University, Atlanta, GA.

The epigenome sits at the interface between the external environment and the underlying genetics. There is extensive individual variation in the epigenome that can be shaped by a combination of genetic and non-genetic influences. Here we compare DNA methylation between African and European ancestral groups in two different study cohorts: (1) the CANDLE (Conditions Affecting Neurocognitive Development and Learning in Early Childhood) study, which provides genome-wide methylation (HumanMethylation27) and expression (Illumina WG-6 BeadChip) measures in newborn umbilical cord blood from African Americans (N = 67) and European Americans (N = 45) in the Memphis area, and (2) methylation measures in lymphoblastoid cell lines from Yoruba African (YRI) and CEPH European (CEU) panels of HapMap. We also evaluate the influence of maternal nutrition—specifically, plasma levels of vitamin D and folate during pregnancy—on child methylation. We define ancestry-dependent differences in methylation at ~2,300 genes (over 11% of the CpG sites we surveyed) at an FDR of 5%. About 35% of these differentially methylated sites in CANDLE is replicable in HapMap and include genes involved in tumor suppression (e.g., *APC*, *BRCA1*, *MCC*). Overall, there is lower methylation of CpG islands in African ancestral groups. While maternal nutritional factors have only modest influence on DNA methylation in newborns, we note that the methylation of the immune antigen gene, *CD33* (cg11122968), in influenced by genetic ancestry and maternal vitamin D (unadjusted p = 0.00007). Our results show that ancestry-dependent DNA methylation patterns are remarkably stable and are likely influenced by a combination of genetic and non-genetic factors, including maternal nutrition.

455M

Self-report is a valid measure of maternal smoking during pregnancy and its epigenetic effects are modified by local (cis) genomic ancestry. *S.S. Oh[1], D. Hu[1], J.M. Galanter[1], C.R. Gignoux[2], C. Eng[1], S. Huntsman[1], E.G. Burchard[1], The GALA II Investigators.* 1) UC San Francisco, San Francisco, CA; 2) Stanford University, Palo Alto, CA.

BACKGROUND Maternal smoking during pregnancy (MSP) has been associated with methylation at 26 specific genomic regions (CpG loci) in a birth cohort of newborns. The Norwegian Mother and Child Cohort Study (MoBa) prospectively determined MSP using mid-pregnancy maternal cotinine and used newborn cord blood to assess methylation. We assessed the validity of retrospective self-reported MSP on differential methylation in our own cohort of Latino children by replicating the MoBa results. We further investigated whether MSP-related methylation marks were modified by locus-specific genetic ancestry at that locus. **METHODS** Methylation status for the same 26 MoBa CpGs was assessed with the Infinium 450K Methylation BeadChip using whole blood from 573 Latino children in the GALA II Study, a nationwide asthma case-control study of Latinos. We used robust linear regression to test the association between MSP exposure and methylation at each CpG locus, adjusting for child's sex, age, and ethnicity. Additional covariates included mother's age, parity, and education; child's asthma status, postnatal exposure to tobacco smoke, and proportions of Native American and African ancestry; and technical factors (estimated blood cell counts, batch, chip, position, and the first 10 principal components). We assessed the effect of local European ancestry on methylation by estimating the number of European alleles at each of the 26 loci. The same regression was then repeated after including local European ancestry as a covariate under an additive model. Genetic ancestry was estimated with the program ADMIXTURE using a 3-population model. Local ancestry for the 26 loci was estimated using LAMPLD, assuming 3 ancestral populations. **RESULTS** Remarkably, the relative methylation difference among GALA II subjects was in the same direction as MoBa subjects for 25 out of 26 CpGs. The probability of matching differences in methylation for 25 out of 26 loci occurring by chance is extremely low ($p = 3.9\times10^{-7}$). Thirteen of these CpG loci showed evidence of modification by local European ancestry. For the most significant CpG (cg11715943, mapping to *HLA-DPB2*), each additional European allele at that locus decreased methylation by 2.1% ($p = 9.9\times10^{-11}$). **CONCLUSIONS** We demonstrate in our study that self-report of MSP is comparable to prospectively ascertained, biomarker-validated MSP status. Furthermore, MSP-related methylation is influenced by local genetic ancestry.

456T

Schizophrenia EWAS Supports Findings of GWAS. *A.P.S. Ori[1], J. Listgarten[2], J.Y. Zou[3], L.M. Olde Loohuis[1], M.P. Boks[4], R.S. Kahn[4], J. Ernst[5], R.A. Ophoff[1,6].* 1) Center for Neurobehavioral Genetics, University of California, Los Angeles, Los Angeles, CA; 2) eScience Research Group, Microsoft Research, Los Angeles, California, USA; 3) The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 4) Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, The Netherlands; 5) Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, California, USA; 6) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California, USA.

Schizophrenia is a highly heritable and polygenic psychiatric disorder with largely unknown etiology. DNA methylation is an epigenetic mark that has been shown to be both heritable and influenced by environmental factors. Differential DNA methylation at specific promoter regions of schizophrenia candidate genes has previously been linked to the disease. In this study we aim to identify variation in the epigenome that is associated to schizophrenia by interrogating DNA methylation sites across the genome. We matched 450 individuals with schizophrenia and 450 healthy controls for age and gender from a relatively homogeneous Dutch population. Blood-derived whole genome DNA methylation was collected using Illumina's Human Methylation assay. We applied FaST-LMM-EWASher (a linear mixed model with principal components) to identify sites associated to our phenotype of interest while accounting for potential spurious association arising from technical artifacts and cell-type heterogeneity. In addition, we modeled top sites shown to be associated to schizophrenia for presence or absence of chromatin marks using ChromHMM and chromatin state calls derived from ENCODE and Epigenomics Roadmap data. We identify multiple CpG loci with DNA methylation profiles associated with schizophrenia disease status (q-value < 0.05). Among these sites are regions of genes important for neuronal functioning, such as vacuolar protein sorting 52 homolog (VPS52). This gene is involved in vesicle trafficking and fusion to target membranes and is located at the classical MHC locus, a region frequently associated to schizophrenia. In addition, regions close to DNA methylation sites most associated to clinical diagnosis are significantly enriched for common variants associated to schizophrenia based on recent GWAS findings. Finally, we show that DNA methylation at promoter regions associated to the disorder overlap with chromatin marks enriched for fetal brain and cell types involved in immune response. To summarize, we performed an epigenome-wide association study of schizophrenia in a homogeneous sample using whole blood. We observed significant association between DNA methylation and disease status that overall highlights genes implicated in neuronal functioning. These results are also in line with findings of recent genetic schizophrenia studies. Our findings support the potential of using whole blood for studying biological mechanisms underlying psychiatric disorders.

457M

DNA methylation of lipid-related genes affects blood lipid levels: a genome-wide screen. *L. Pfeiffer[1,2], S. Wahl[1,2,3], L.C. Pilling[4], E. Reischl[1,2], J.K. Sandling[5,6], S. Zeilinger[1,2], L.M. Holdt[7], A. Kretschmer[1,2,8], K. Schramm[9,10], J. Adamski[11], T. Illig[12], A.K. Hedman[13,14], M. Roden[3,15,16], D.G. Hernandez[17], A.B. Singleton[18], S. Bandinelli[18], W.E. Thasler[19], H. Grallert[1,2,3], C. Gieger[20], C. Herder[3,15], D. Teupser[7], C. Meisinger[2], T.D. Spector[21], F. Kronenberg[22], H. Prokisch[9,10], D. Melzer[4], A. Peters[1,2,23], P. Deloukas[5,24,25], L. Ferrucci[26], M. Waldenberger[1,2].* 1) Research Unit of Molecular Epidemiology, Helmholtz Zentrum Mnchen, German Research Center for Environmental Health, Neuherberg, Germany; 2) Institute of Epidemiology II, Helmholtz Zentrum Mnchen, German Research Center for Environmental Health, Neuherberg, Germany; 3) German Center for Diabetes Research (DZD), partner site Dsseldorf, Germany; 4) Epidemiology and Public Health Group, University of Exeter Medical School, Exeter, United Kingdom; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK; 6) Present address: Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 7) Institute of Laboratory Medicine, University Hospital Munich and Ludwig-Maximilians-University Munich, Munich, Germany; 8) Department of Dermatology, Venereology and Allergy, Christian Albrechts University Kiel, Kiel, Germany; 9) Institute of Human Genetics, Helmholtz Zentrum Mnchen, German Research Center for Environmental Health, Neuherberg, Germany; 10) Institute of Human Genetics, Technical University Munich, Munich, Germany; 11) Genome Analysis Center, Institute of Experimental Genetics, Helmholtz Zentrum Mnchen, German Research Center for Environmental Health, Neuherberg, Germany; 12) Hannover Unified Biobank, Hannover Medical School, Hannover, Germany; 13) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 14) Present address: Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 15) Institute for Clinical Diabetology, German Diabetes Center at Heinrich Heine University, Leibniz Center for Diabetes Research, Dsseldorf, Germany; 16) Department of Endocrinology and Diabetology, University Hospital, Dsseldorf, Germany; 17) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, United States; 18) Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy; 19) Department of Surgery, University Hospital Munich and Ludwig-Maximilians-University Munich, Munich, Germany; 20) Institute of Genetic Epidemiology, Helmholtz Zentrum Mnchen, German Research Center for Environmental Health, Neuherberg, Germany; 21) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 22) Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria; 23) German Research Center for Cardiovascular Disease (DZHK), Partner-site Munich, Germany; 24) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 25) Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, 21589 Saudi Arabia; 26) Clinical Research Branch, National Institute on Aging, Baltimore, MD, United States.

Background: Epigenetic mechanisms might be involved in the regulation of interindividual lipid level variability and thus may contribute to the cardiovascular risk profile. Aim of this study was to systematically investigate the association between genome-wide DNA methylation in whole blood and serum lipid levels of HDL-C, LDL-C, triglycerides (TG) and total cholesterol (TC).

Results: Genome-wide DNA methylation patterns were determined in blood samples of 1776 subjects of the KORA F4 cohort using the Infinium HumanMethylation450 BeadChip (Illumina). Associations between DNA methylation and lipid levels were analyzed using linear mixed-effects models with lipid levels as response. To correct for multiple comparisons, a genome-wide significance level of 1.1E-07 was used according to the Bonferroni procedure. We identified eleven lipid-related CpGs annotated to several genes including *ABCG1*. One CpG site located in *ABCG1* was associated in opposite directions with both HDL-C (β coefficient=0.049, p=8.26E-17) and TG levels (β=0.070, p=1.21E-27). Nine of the associations were confirmed by replication in KORA F3 (N=499) and InCHIANTI (N=472). Associations between TG levels and CpGs of two genes were also found in adipose but not in skin tissue of the MuTHER cohort, indicating tissue specificity. Expression analysis revealed that the association between *ABCG1* methylation and lipid levels is partly mediated by the expression of *ABCG1*. The identified association between *ABCG1* methylation and *ABCG1* mRNA levels is possibly based on methylation-dependent transcription factor binding as observed in electrophoretic mobility shift assays. DNA methylation of *ABCG1* was also associated with previous myocardial infarction in KORA F4 (odds ratio 1.15, 95%CI=1.06-1.25).

Conclusion: We found associations between DNA methylation and lipid levels for genes contributing to the modulation of cholesterol and fatty acid metabolism. The results indicate an epigenetic impact on metabolic regulation in humans and give new insights into the complex picture of lipid-related complex diseases.

458T

The Genome Wide DNA Methylation Signature of Subjects as They Enter and Exit Short Term Alcohol Treatment. *R. Philibert[1, 2], B. Penaluna[1], T. White[1], S. Shires[1], T. Gunter[3], J. Liesveld[1], C. Erwin[4], N. Hollenbeck[1], T. Osborn[2].* 1) Dept Psychiatry, Univ Iowa, Iowa City, IA; 2) Behavioral Diagnostics Iowa City, IA 52240; 3) Indiana Univ. Indianopolis, IN; 4) Texas Tech University Health Sciences Lubbock, TX.

Alcoholism has a profound socio-economic and personal impact on tens of millions of individuals throughout the world. Unfortunately, our ability to quantify the chronicity and quantity of alcohol intake is often hindered by reliance on self-report while our ability to monitor treatment response is challenged by the absence of robust and reliable biomarkers. Recently, in work replicated in over 12 independent genome wide analyses, our consortium has shown that quantitative DNA methylation assessments, particularly at AHRR locus cg05575921, can sensitively and specifically quantify tobacco smoking exposure. Now, in an extension of prior work with lymphoblasts, we examine lymphocyte DNA prepared from heavy, recently intoxicated subjects as they both begin and finish standard 28 day alcohol treatment using standard genome wide methylation techniques and the Infinium 450K HumanMethylationBeadChip. We show that as compared to abstinent controls, heavy alcohol intake is associated with profound widespread changes in DNA methylation with over 800 distinct loci reaching genome wide significance. In marked contrast to prior work with tobacco, there do not seem to be highly sensitive sentinel loci with the overall pattern of changes being consistent with the hypothesis that alcohol affects cellular functions by nonspecific general solvent effects (i.e. changes in cytoplasmic dialect constant). Significantly, there is a marked bias for alcohol induced changes to revert as a function of abstinence. Pathway analysis of the alcohol induced changes demonstrates a mapping of these changes to gene networks involved in apoptosis. Comparison of the methylation signatures for alcohol and tobacco smoke consumption does not demonstrate significant overlap. We conclude that DNA methylation may be useful clinically in the assessment of alcohol use status and monitor alcohol treatment response.

459M

genome wide analysis to find epigenetically inactivated regions associated with Thyroid hormones - A population based crosssectional study. *R. Rawal[1,2], A.M. Nazeer-batcha[3], M. Waldenberger[1], S. Zeilinger[1], C. Gieger[1,2], C. Meisinger[2].* 1) Department of Molecular Epidemiology, Institute of Epidemiology II, Munich, Neuherberg-Munich, Germany; 2) Institute of Genetic Epidemiology, Helmholtz Zentrum Mnchen, Neuherberg, Germany; 3) Department of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians- Universitt Mnchen, Germany.

Thyroid gland and thyroid hormones play a significant role in regulating several bodily functions such as cell growth, development and metabolism. Though, several genome-wide association studies (GWAS) have reported a number of genes to be associated with thyroid related phenotypes, some unexplained heritability prevail in human population. Thus, an Epi-genome wide association study (EWAS) was conducted with linear model for phenotype concentrations and logistic model for binary phenotypes in a population based cohort (n = 1,799) using the Illumina 450K Methylation BeadChip to find any association between thyroid related phenotypes and DNA methylation. We report nine distinct CpG sites: four were associated with Tri-iodothyronine (T3) in the gene regions of KLF9, DENND2D, PPP2R5C and one locus in 2p24 region one CpG site in HS3ST3B1 gene was associated with Tetra-iodothyronine (T4), four CpG sites in TSPAN14, CLU, TTLL3, VAT1 genes were associated with T3 to T4 ratio, All reported CpGs showed negative association except two. Smoking adjusted analyses did not provide any information about its influence on the association. These results show the effect of DNA methylation on thyroid hormones and provide additional knowledge on the possibility of methylation, associated with thyroid phenotypes, for future research.

460T

DNA methylation profiles that distinguish rheumatoid arthritis (RA) from osteoarthritis in fibroblast-like synoviocytes can be detected in immune cells from peripheral blood. *B. Rhead[1], C. Holingue[1], M. Cole[1], X. Shao[1], H. Quach[1], D. Quach[1], L.F. Barcellos[1], L. Criswell[2].* 1) Division of Epidemiology, Genetic Epidemiology and Genomics Laboratory, School of Public Health, University of California, Berkeley, Berkeley, CA; 2) Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California, San Francisco, San Francisco, CA.

RA [MIM 180300] is a chronic inflammatory disease with potential to cause substantial disability, primarily due to the erosive and deforming process in joints. RA etiology is complex, with contributions from genetic and non-genetic factors. Epigenetic changes, such as altered patterns of DNA methylation, are also present in RA. Recent work by Firestein and colleagues (Nakano, 2013) compared genome-wide DNA methylation profiles in synovium-derived fibroblast-like synoviocytes (FLS) from RA patients and osteoarthritis controls and identified a set of differentially methylated genes that appear to distinguish these two forms of arthritis. Given the greater accessibility of peripheral blood compared to synovium-derived FLS, we set out to determine whether similar methylome signatures were present in immune cell subsets in peripheral blood. We generated over 400 genome-wide DNA methylation profiles for 101 women (70 cases from the UCSF RA cohort and 31 controls) using Illumina HumanMethylation450 BeadChips. Four FACS-sorted immune cell types were assayed for each individual: CD14+ monocytes, CD19+ B cells, CD4+ memory T cells, and CD4+ nave T cells. All samples were background subtracted using the lumi Bioconductor package and normalized using all sample mean normalization (ASMN) followed by beta-mixture quantile normalization (BMIQ). All study individuals were fully characterized for whole genome SNP profiles using the Illumina OmniExpress BeadChip. We excluded CpG sites near SNPs known to be common (>1%) from analysis, as well as sites with a low detection p-value, leaving 442,797 sites for analysis. We examined the top differentially hypo- and hypermethylated (n = 17 and 18, respectively) loci from the Firestein study in each immune cell type and found strong evidence for overlapping RA methylome signatures between FLS and immune cells. The most significantly differentially methylated genes in cases vs. controls include *MGMT, ADAMTS2, CABLES1, COL4A1, COL4A2, CYFIP1, EGF, FOXO1, MAP3K1, PHLPP1, PTPN14, RXRA,* and *TGFBR2. MGMT* encodes a DNA repair enzyme and showed the strongest evidence of hypomethylation. The strongest hypermethylation was seen in *MAP3K1*, a serine/threonine kinase involved in some signal transduction cascades. Our investigation of DNA methylation in RA cases and controls underscores the importance of epigenetic mechanisms in RA pathogenesis and represents the largest study, to date, of methylome profiles derived from immune cells.

461M

Epigenome-wide analysis identified highly significant age-related DNA methylation changes. *A. Russo[1,2], S. Guarrera[1,2], G. Fiorito[1,2], C. Di Gaetano[1,2], F. Rosa[1,2], A. Allione[1,2], F. Modica[1], L. Iacoviello[3], M.C. Giurdanella[4], R. Tumino[4], S. Grioni[5], V. Krogh[5], A. Mattiello[6], S. Panico[6], P. Vineis[1,7], C. Sacerdote[1,8], G. Matullo[1,2].* 1) Human Genetics Foundation, Turin, Turin, Italy; 2) Medical Sciences Department, University of Torino, Torino, Italy; 3) Fondazione di Ricerca e Cura “Giovanni Paolo II”, Catholic University, Campobasso, Italy; 4) Cancer Registry and Histopathology Unit, “Civile-M.P. Arezzo” Hospital, ASP7, Ragusa, Italy; 5) Department of Preventive and Predictive Medicine, Epidemiology and Prevention Unit , Fondazione IRCSS Istituto Nazionale dei Tumori, Milano, Italy; 6) Department of Clinical and Experimental Medicine, Federico II University, Napoli, Italy; 7) Epidemiology and Public Health, Imperial College London, UK; 8) Cancer Epidemiology, CPO-Piemonte, Torino, Italy.

Aging is associated with an increased risk for many complex diseases. DNA methylation represents one of the most promising biomarkers of aging. To evaluate the effect of age on DNA methylation, we examined the methylation levels of more than 450K CpG sites in 206 cases and 206 matched controls belonging to the Italian section of the EPIC cohort. EPIC healthy volunteers were recruited between 1994–98 and followed up for myocardial infarction and other diseases. For the CpG methylation level assessment on blood DNA we used the Illumina HumanMethylation450 BeadChip. Data were analyzed according to standard procedures (MethyLumi, Bioconductor). All analyses were corrected for sex, BMI, season, center of recruitment, cellular subtypes and batch effect. Linear regression analysis showed that 10,000 CpG sites were significantly associated with age after Bonferroni correction ($p<10^{-7}$). Positive correlation with age was found for 5,687 CpG sites (57%), whereas 4,313 CpG sites (43%) were negatively correlated with age. The top four significant loci ($p<10^{-28}$) were located in the CpG islands of *ELOVL2* and *FHL2* genes. Methylation levels of the 4 CpG sites were positively correlated with increasing age, according to previously published results. Gene enrichment analysis highlighted that significantly age-associated genes were involved in transcriptional regulation and alternative splicing. These results confirm that DNA methylation alterations occur during aging and that *ELOVL2* and *FHL2* genes could be used as biomarker of aging. Moreover, the methylation differences associated with age could help in understanding molecular mechanisms that underlie the development of age-related diseases.

462T

Characterizing Functional Methylomes in Human Populations Using Novel Next-Generation Capture Sequencing Approach. *X. Shao[1], F. Allum[1], F. Gunard[2], M-M. Simon[1], S. Busche[1], M. Caron[1], T. Kwan[1], S. Marceau[3], M. Lathrop[1], A. Tchernof[3,4], M-C. Vohl[2,4], T. Pastinen[1], E. Grundberg[1].* 1) Department of Human Genetics, McGill University and Genome Quebec Innovation Centre, Montreal, QC H3A 1A5, Canada; 2) Institute of Nutrition and Functional Foods (INAF), Laval University, Quebec, QC G1V 0A6, Canada; 3) Endocrinology and Nephrology, CHU de Quebec Research Center, Quebec, QC, G1V 4G2, Canada; 4) Quebec Heart and Lung Institute, Quebec, QC, G1V 4G5, Canada.

Whole genome bisulfite sequencing (WGBS) has been widely used to detect DNA methylation at single CpG resolution. However, as only ~20% of CpG sites are variable across individuals or across human tissues, WGBS is inefficient and costly and thus not optimal for population studies of methylation variation and the impact on disease susceptibility. Meanwhile, targeted approaches such as Illumina Human Methylation 450K BeadArray or the reduced representation bisulfite sequencing are biased to gene promoter and cover only a minor proportion of variable CpGs. Here we introduce MethylC-Capture Sequencing (MCC-seq), a novel cost-efficient sequencing approach targeting the most informative fraction of CpGs within disease-relevant tissues. Using adipose tissue and whole blood we have designed capture panels focusing on metabolic and autoimmune diseases that target ~160Mb or ~5M dynamic CpGs/tissue mapping to regulatory elements, hypomethylated footprints or an Illumina450K site. The panels also include all SNPs and their proxies ($r^2 > 0.8$) linked to metabolic or autoimmune diseases (GWAS catalogue 9/2013). The approach is based on bisulfite-converted DNA libraries that are selectively enriched using the custom capture panels (Roche/Nimblegen) and sequenced on the IlluminaHiSeq2000 systems. With multiplexing up to 6-fold/lane we note on average ~62% on-target rate/sample corresponding to ~26X mean genome coverage/sample. We also find a high concordance of within samples CpG methylation on different runs (ρ=0.9). Methylation calls between MCC-seq and WGBS are found to be highly concordant (ρ=0.85), whereas lower concordance of MCC-seq and 450K (ρ=0.80) is observed. Moreover, our results indicate that MCC-seq is capable of simultaneous genotyping where 97% of on-target SNPs could be detected with an average concordance of 99% with genotypes from the Illumina 2.5M array. Finally, we integrated phased genotypes and the MCC-seq data to detect allele-specific methylation (ASM). Preliminary data show ~2.4% of all heterozygous SNP associated CpGs are demonstrating ASM where there is an ~2-fold enrichment in enhancer regions, most likely representing sequence-dependent effects. In addition, while imprinted loci are enriched among ASM they only account for 0.84% of detected sites. Overall, the MCC-seq is cost-efficient and provides single-base resolution of functional DNA methylome and genotype profiles suitable for large-scale disease-association studies.

463M

Facioscapulohumeral muscular dystrophy 2 (FSHD2) testing - a UK pilot study and a clinical diagnostic service. *D.J. Smith[1], JA. Whitfield[1], S. O'Shea[1], R. Whittington[1], L. Yarram-Smith[1], PW. Lunt[2], M. Williams[1].* 1) Bristol Genetics Laboratory, North Bristol NHS Trust, Southmead Hospital, Bristol, UK; 2) National Genetics & Genomics Education Centre, National School for Healthcare Science, Birmingham, UK.

FSHD (affecting ~1 in 20,000 individuals) is an autosomal dominant but epigenetically regulated disorder, with a characteristic pattern of progressive muscle involvement commencing in the face and shoulder-girdle. Two clinically indistinguishable forms (FSHD1 & FSHD2) differ by the molecular basis of epigenetic regulation, but are underpinned by hypomethylation of 4q35, leading to aberrant expression of DUX4 (a toxic transcription factor). FSHD1 (OMIM 158900) (95% cases) is caused by contraction of D4Z4 repeats, resulting in allele specific hypomethylation. FSHD2 (OMIM 158901) (~3% of cases) is contraction independent, caused by mutations in the SMCHD1 gene encoding a chromatin modulating enzyme, resulting in definitive hypomethylation and chromatin relaxation of D4Z4 array on both 4q35 alleles. A permissive haplotype at 4q35 is required for clinical expression of both FSHD1 and 2, necessitating digenic inheritance for FSHD2.

Bristol Genetics Laboratory has provided a FSHD specialist diagnostic service since 1992, and processes over 400 referrals annually. Patients negative for standard and extended deletion analysis but clinically typical for FSHD are candidates for FSHD2 testing (4.8% diagnostic referrals.)

Lemmers et al (2012) identified mutations in SMCHD1 (18p21.31) causing global hypomethylation of D4Z4. Hartweck et al (2013) identified a region showing intense hypomethylation in FSHD2 (DR1). BGL has developed a novel pyrosequencing assay for quantification of methylation using 10 statistically evaluated sites within DR1. A pilot UK cohort of 15 clinically typical FSHD1 negative patients, (assessed by clinical proforma) were assessed for hypomethylation, followed by SMCHD1 sequencing. Our results indicate that 9/15 patients showed hypomethylation (DR1) below a threshold level, with novel candidate SMCHD1 mutations found in all 9; (3 missense, 1 non-sense, 1 duplication, 2 deletion, 2 potential splice site). FSHD2 is an example of digenic inheritance, and this increases the complexity of genetic counselling, as the risk to offspring is between 25 and 50%, depending on the haplotypes of the wider family. We present a service overview, the results of this study and interesting cases highlighting the clinical utility of genetic testing, and family risks associated with this digenic disease.

464T

Sex specific epigenetic and transcriptional responses of peripheral blood leukocytes (PBLs) to lipopolysaccharide (LPS). *M. Stein[1], C. Hrusch[2], J. Nicodemus-Johnson[1], A. Sperling[2], C. Ober[1].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL.

Many immune mediated diseases show striking sex differences in prevalence, yet the genetic and epigenetic factors contributing to these differences are poorly characterized. We undertook this study to identify sex-specific differences in methylation and transcriptional responses to LPS, an important activator of the innate immune response, in PBLs collected from 153 members (55% female; range 6–76 years) of a large Hutterite pedigree who previously participated in our studies of asthma. Cells were incubated with LPS+media and with media alone (control) in a closed system (TruCulture®, Myriad RBM, Austin). After 30 hours, DNA and RNA were extracted. In a pilot study of 28 school children (12 girls, 16 boys; 6–14 years old), we used the Illumina Methylation450 Beadchip for methylation and the HumanHt-12v4 for expression studies. Studies in the larger sample are underway. We defined sex-specific differentially methylated regions (DMRs) in response to LPS as CpG sites with methylation response (q-value <0.05) in one sex, but not in the other (q-value >0.2). We observed 39 sex-specific DMRs in response to LPS in girls only and 85 in boys only. 48% of sex-specific DMRs were near a gene that was detected as expressed in the arrays. Expression levels of five (8.5% of expressed genes near a DMR) were correlated with methylation levels. These genes include intriguing candidates that have previously been associated with asthma. For example, methylation levels at a CpG site in the *ANXA2* gene were lower in LPS-treated samples compared to control samples in boys ($P=2.15\times10^{-5}$) but remained unchanged in girls (P=0.48); *ANXA2* expression was higher in LPS-treated samples compared to control samples in boys ($P=1.97\times10^{-4}$) but not in girls (P=0.16). Increased expression of *ANXA2* was previously associated with asthma exacerbations (PMID 10430734). Methylation levels at a CpG site in *ATP2C1* were significantly higher in LPS-treated samples in boys ($P=1.97\times10^{-4}$) but remained unchanged in girls (P=0.95); gene expression was also higher in LPS-treated samples in boys ($P=4.91\times10^{-4}$) but not in girls (P=0.92). Methylation levels in this gene were previously associated with a parent-of-origin effect in children with asthma (PMID 24166889). Overall, these data show that methylation and transcriptional responses to an important environment stimulus (LPS) can be sex-specific, potentially providing a mechanism for the sex disparities in occurrence and course of immune mediated diseases.

465M

CpG Methylation and G-quadruplex structure influence genotyping of the imprinted MEST gene promoter. *A.J. Stevens[1], S. Stuffrein-Roberts[1], A. Gibb[1], K. Doudney[1], A.L Miller[1], A. Aitchison[1], S. Macmill[1], M.R Eccels[2], V.V Filichev[3], P.R Joyce[4], M.A Kennedy[1].* 1) Department of Pathology, University of Otago, Christchurch, New Zealand; 2) Department of Pathology, University of Otago, Dunedin School of Medicine, Dunedin, New Zealand; 3) Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand; 4) Department of Psychological Medicine, University of Otago, Christchurch, New Zealand.

While exploring genetic variability in the maternally imprinted human gene MEST, which plays a role in development and maternal behaviour(1–3), we discovered three SNPs in a CpG island located at the 5' end of the gene(4). These SNPs are in total linkage disequilibrium, such that there are two haplotypes in the human population. When these SNPs were typed in many subjects we observed a non-Mendelian pattern whereby only one haplotype was detectable in each subject, never both, despite the use of multiple genotyping methods. Experiments in which genomic DNA from different subjects was mixed prove that the assays were capable of detecting both haplotypes simultaneously(4). The absence of observed heterozygotes was most likely due to allelic dropout of one allele. Competitive PCR performed on differentially methylated templates of opposite haplotypes showed that the methylated allele consistently fails to genotype. Using circular dichroism spectroscopy (CD), polyacrylamide gel electrophoresis, and fluorescent dimethyl sulfate footprinting we investigated the propensity for sequence motifs from the human MEST promoter region to adopt G4 structure, in vitro. Use of fluorescently labelled primers during PCR showed that these MEST G4 motifs inhibit the amplification of methylated templates for five cycles longer than unmethylated templates. We demonstrated that the combination of G-quadruplex formation and DNA methylation are both required for allelic dropout and neither factor in isolation can explain the observed phenomenon. By investigating the annealing rates of G4 motifs we found methylated oligonucleotides showed faster re-annealing into G-quadruplex structures after heating and it is possible this phenomenon could impact the efficiency of amplification during PCR and account for the observed dropout of the methylated (imprinted) maternal MEST allele. This parent-of-origin specific allelic drop-out of an imprinted locus has important implications for the diagnostic clinical laboratory, and the interaction of G-quadruplexes and methylation in the promoter of an imprinted gene raises the broader question of a potential role for these structures in genomic imprinting or regulation. 1.Lefebvre et al. (1998) Nat Genet 20, 163–169 2.Kobayashi et al. (1997) Human molecular genetics 6, 781–786 3.Riesewijk et al. (1997) Genomics 42, 236–244 4.Stuffrein-Roberts, S. (2008) Allelic expression patterns in psychiatric candidate genes. PhD Thesis in Pathology.

466T

Dose-dependent changes in DNA methylation identified in Williams syndrome and the reciprocal 7q11.23 duplication syndrome. *E. Strong[1], D. Butcher[2], C.B. Mervis[3], C.A. Morris[4], R. Weksberg[2], L.R. Osborne[1,5].* 1) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 2) Program in Genetics and Genome Biology, SickKids Research Institute, Toronto, Ontario, Canada; 3) Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY, US; 4) Department of Pediatrics, University of Nevada School of Medicine, Las Vegas, NV, US; 5) Department of Medicine, University of Toronto, Toronto, Ontario, Canada.

Williams syndrome (WS) is a multisystem developmental disorder caused by hemizygous deletion of 1.5 Mb spanning 26 genes on chromosome 7q11.23. The reciprocal duplication leads to a distinct disorder, 7q11.23 duplication syndrome (Dup7). Both syndromes are associated with developmental delay or intellectual disability, attention deficit hyperactivity disorder, and contrasting cognitive and behavioral phenotypes including different patterns of anxiety disorder. Little is known about the underlying molecular mechanisms. Five genes at 7q11.23 have been implicated in epigenetic regulation, therefore we hypothesized that deletion or duplication of this region would disrupt the epigenetic profile of individuals with WS or Dup7, possibly impacting their phenotypic presentation.

To determine if epigenetic regulation was perturbed in the 7q11.23 disorders, we measured DNA methylation in whole blood from 20 children with WS, 10 with Dup7 and 15 age-matched typically developing controls using the Illumina Infinium HumanMethylation450 array. Differential methylation was assessed using a cutoff of 17% gain or loss of methylation (FDR $p \leq .05$). We identified >1,000 differentially methylated (DM) probes across the WS and Dup7 cohorts that correspond to ~500 unique genes. Hierarchical clustering of the top DM probes resulted in three distinct clusters, corresponding to WS, control and Dup7. WS and Dup7 showed both distinct and dose-dependent changes in DNA methylation, with the WS group showing predominantly hypermethylation and the Dup7 group predominantly hypomethylation. Many of the DM probes spanned genes involved in brain development and function. For the few DM genes where we could assess expression in blood (*RGS2*, *HDAC4*, *DNMT3A*), we show corresponding changes in expression level. We validated the methylation changes in probes from three genes using pyrosequencing and found good correlation with the microarray data ($r^2 > .9$).

The aberrant DNA methylation profiles that we identified in WS and Dup7 suggest methylation differences may contribute to the phenotypic expression of these disorders. Our results suggest that one or more genes within the commonly deleted/duplicated region is important for proper DNA methylation at specific sites within the genome and acts in a dose-dependent manner.

467M

Race-specific association between DNA methylation and body mass index: the Bogalusa Heart Study. *D. Sun[1,2], X. Fu[3], S. Li[1], Q. Li[4], T. Chen[4], Y. Lian[4], C. Fernandez[1], W. Chen[1].* 1) Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA; 2) Department of Epidemiology, School of Public Health, Peking University Health Science Center, Beijing, China; 3) Department of Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA; 4) Microarray Core Facility, University of Texas Southwestern Medical Center, Dallas, TX, USA.

This study test the hypothesis that genome-wide DNA methylation is associated with adiposity measured as body mass index (BMI). As part of the Bogalusa Heart Study, 968 adults aged 28–51 years were recruited between 2006 and 2010. Of these individuals, 830 participants (584 whites, 246 blacks) were examined only once and used as the discovery cohort; 138 participants (95 whites, 43 blacks) were examined twice 3.2 years apart and used as the replication cohort. The discovery findings were replicated by the baseline and follow-up sample. Genome-wide methylation of peripheral leukocyte DNA samples was measured by HumanMethylation450 BeadChip and sample and probe quality analysis was performed. In the association analyses by race, adjusting for age, gender and smoking, DNA methylation levels at 17 sites in whites and at 4 sites in blacks were found to be significantly associated with BMI in discovery cohort. As shown in the table below, 2 (cg06500161 and cg00574958) of the 17 sites in whites and 1 (cg26403843) of 4 sites in blacks were significant in both baseline and follow-up replication cohorts. In conclusion, DNA methylation changes in lipid metabolism-related genes are associated with BMI values, which provides potential race-specific epigenetic mechanisms underlying the development of obesity in whites and blacks.

Effect sizes and Bonferroni-adjusted P values

CpG	Gene	Discovery cohort		Baseline replication		Follow-up replication	
β		P		β	P		
cg06500161	ABCG1	1.4×10^{-3}	2.2×10^{-9}	1.4×10^{-3}	2.0×10^{-3}	1.6×10^{-3}	5.2×10^{-3}
cg00574958	CPT1A	-9.8×10^{-4}	1.1×10^{-9}	-8.8×10^{-4}	1.8×10^{-2}	-7.9×10^{-4}	3.3×10^{-2}
cg26403843	RNF145	2.6×10^{-3}	4.0×10^{-3}	3.0×10^{-3}	4.6×10^{-2}	3.3×10^{-3}	4.1×10^{-2}

468T

Genome-wide association scans identify differentially methylated and expressed regions related to smoking in adipose tissue. *P-C. Tsai[1], C. Glastonbury[1], A. Viuela[1], I. Erte[1], W. Yuan[1], E. Dermitzakis[2], P. Deloukas[3], K. Small[1], T.D. Spector[1], J.T. Bell[1].* 1) Department of Twin Research & Genetic Epidemiology, King's College London, London, United Kingdom; 2) Department of Genetic Medicine and Development, Universit de Genve, Genve, Switzerland; 3) William Harvey Research Institute, Queen Mary University of London, London, United Kingdom.

Smoking is an important environmental risk factor in the development of chronic disease and cancer. Multiple epigenome-wide association scans (EWASs) for smoking have identified and replicated differentially methylated regions in blood. To explore these effects across tissues, we analysed DNA methylation profiles from the Illumina 450k array in adipose tissue samples from 349 healthy female twins, including 35 current smokers, 128 ex-smokers, and 186 non-smokers. We identified 39 CpG sites in or near 25 genes that were significantly associated with smoking at a false discovery rate of 5%. Several of the smoking-associated methylated sites, such as those in the *AHRR*, *F2RL3*, and *CYP1A1* gene, were previously identified to be associated with smoking in blood and lung tissues. Furthermore, smoking-related regions were predominately hypo-methylated in adipose tissue, consistent with previous results. We next performed a genome-wide scan of RNA-sequencing gene expression profiles in the same 349 adipose tissue samples to identify differentially expressed exons associated with smoking. In total, 48 exons in 35 genes were associated with smoking status at a false discovery rate of 5%. Of the 35 genes, 4 genes (*CYTL1*, *AHRR*, *F2RL3*, and *CYP1B1*) overlapped with the top methylation results. Inter-individual correlation showed strong associations between DNA methylation and exon expression levels at a number of genes, including several of the smoking-associated loci. To our knowledge, this is the first smoking-EWAS in adipose tissue. We identified both adipose-specific and tissue-shared DNA methylation changes related to smoking, with corresponding gene expression associations with smoking status. We conclude that smoking exerts a strong effect on DNA methylation and gene expression levels across multiple tissues, and suggest that smoking should be incorporated as a covariate in future EWAS.

469M

Downstream analyses and Mendelian randomization study on methylome-wide associations with BMI reveal biological pathways underlying obesity and related metabolic consequences. *S. Wahl[1,2,3], B. Lehne[4], A.W. Drong[5], M. Loh[4], S. Zeilinger[1,2], G. Fiorito[6], S. Kasela[7], R. Richmond[8], A. Dehghan[9], L. Franke[10], T. Esko[7,11,12,13], L. Milani[7], C.L. Relton[8,14,15], J. Kriebel[1,2,3], H. Prokisch[16,17], C. Herder[18], A. Peters[1], T. Illig[1,2,19], M. Waldenberger[1,2], J.T. Bell[20], O.H. Franco[9], P. van der Harst[10], C.M. Lindgren[5,21], M.I. McCarthy[5,22], G. Matullo[6], C. Gieger[1,2], J.S. Kooner[23,24,25], H. Grallert[1,2,3], J.C. Chambers[5,18].* 1) Institute of Epidemiology II, Helmholtz Zentrum Mnchen - German Research Center for Environmental Health, Neuherberg, Germany; 2) Research Unit of Molecular Epidemiology, Helmholtz Zentrum Mnchen - German Research Center for Environmental Health, Neuherberg, Germany; 3) German Center for Diabetes Research (DZD), Neuherberg, Germany; 4) Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 6) Human Genetics Foundation, HuGeF, I-10126 Torino, Italy Department of Medical Sciences, University of Torino, I-10126, Torino, Italy; 7) Estonian Genome Center, University of Tartu, Tartu, Estonia; 8) MRC Integrative Epidemiology Unit, Oakfield House, University of Bristol, Bristol BS8 2BN, UK; 9) Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 10) Departments of Cardiology and Genetics, University of Groningen, University Medical Center Groningen, Groningen & Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, The Netherlands; 11) Divisions of Endocrinology and Genetics and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts, USA; 12) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA; 13) Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA; 14) Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, NE1 3BZ, Tyne and Wear, UK; 15) School of Social and Community Medicine, University of Bristol, Bristol, UK; 16) Institute of Human Genetics, Helmholtz Zentrum Mnchen - German Research Center for Environmental Health, Neuherberg, Germany; 17) Institute of Human Genetics, Technische Universitt Mnchen, Munich, Germany; 18) Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Dsseldorf, Dsseldorf, Germany; 19) Hannover Unified Biobank, Medical School Hannover, Hanover, Germany; 20) Department of Twin Research & Genetic Epidemiology, King's College London, St Thomas' Hospital Campus, Westminster Bridge Road, London SE1 7EH, UK; 21) Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 22) Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom; 23) National Heart and Lung Institute, Imperial College London, London W12 0NN, UK; 24) Ealing Hospital NHS Trust, Middlesex UB1 3HW, UK; 25) Imperial College Healthcare NHS Trust, London W12 0HS, UK.

Studying body mass-related DNA methylation signatures is promising from two perspectives. First, epigenetic mechanisms might be involved in the development of obesity, and second, epigenetic regulation may partially be underlying obesity-related pathogenic processes including disturbed lipid and glucose metabolism. In an epigenome-wide approach based on the Illumina Infinium 450k assay, we have identified 187 methylation sites stably associated with body mass index (BMI) in more than 10,000 individuals of European and South Asian ancestry. In subsamples of 3961 and 703 subjects, the identified methylation sites were integrated with single nucleotide polymorphism and gene expression data. In addition, directionality of the observed BMI-methylation associations was explored by means of Mendelian randomization experiments using SNPs associated with BMI in published genome-wide association studies, and SNPs associated with methylation in our study as instrumental variables. Furthermore, we tested association of the identified methylation sites with incident type 2 diabetes (T2D) during a follow-up of 7.6 years on average (n = 3064), as well as with clinical traits reflecting glucose and lipid metabolism and inflammation (n = 4176). Methylation at 125 CpG sites showed strong genetic regulation ($p = 4.610^{-8}$ to 1.010^{-350}). 13 CpG sites were significantly associated with expression of genes located in *cis* ($p = 5.510^{-6}$ to 6.110^{-35}). Mendelian randomization experiments suggest that methylation at the majority of CpG sites might be consequential to changes in BMI. 92 methylation sites showed significant association with future risk of T2D ($p = 7.110^{-12}$ with *ABCG1*) or clinical traits independent of BMI, most prominently with triglycerides ($p = 5.610^{-65}$ with *ABCG1*), HDL cholesterol ($p = 3.810^{-45}$ with *ABCG1*), C-reactive protein ($p = 6.510^{-36}$ with *CRELD2*) and HbA1c ($p = 1.610^{-12}$ with *ABCG1*). Adjustment for methylation levels at the identified CpG sites resulted in a reduction of up to 68% in the association of these traits with BMI ($p = 1.210^{-4}$ to 3.810^{-19}). Ongoing analyses will determine the extent to which some of this reduction is due to mediation through methylation. The presented work provides a close examination of the CpG sites discovered in the first large meta-analysis of genome-wide DNA methylation in relation to BMI. Our findings provide new evidence for biological pathways underlying obesity and related metabolic disturbances.

470T

Epigenome-wide association study identifies epigenetic markers for asthma and allergic disease in the MeDALL study. *C.J. Xu[1,2,3], M. Kerkhof[2,4], S.A. Jankipersadsing[1,2,3], C. Sderhll[5], M. Wickman[6], E. Meln[6], M. Bustamante[7,8], E. Morales[7], J. Sunyer[7], N. Baz[9], I. Annesi-Maesano[9], U. Gehring[10], J.M. Ant[7], J. Bousquet[11], P. van der Vlies[3], C.C van Diemen[3], C. Wijmenga[3], D.S. Postma[1,2,3], G.H. Koppelman[2,12].* 1) University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, The Netherlands; 2) University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen, The Netherlands; 3) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands; 4) University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands; 5) Department of Biosciences and Nutrition, Karolinska Institutet, Sweden; 6) Sachs' Children's Hospital and Institute of Environmental Medicine, Karolinska Institutet, Sweden; 7) Center for Research on Environmental Epidemiology (CREAL), Barcelona, Spain; 8) Center for Genomic Regulation (CRG), Barcelona, Spain; 9) EPAR Institute Pierre Louis of Epidemiology and Public Health, INSERM and UPMC Sorbonne Universits, Paris, France; 10) Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands; 11) University Hospital, Hospital Arnaud de Villeneuve, Department of Respiratory Diseases, Montpellier, France; 12) University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital Groningen, The Netherlands.

Asthma and allergic diseases are caused by a combination of genetic and environmental factors. Although a large number of genetic variants are associated with asthma and allergic risk, many of the risk factors remain unexplained. We hypothesize that epigenetic mechanisms are partly responsible for the missing heritability and tested our hypothesis using an epigenome-wide association approach. We examined DNA methylation in asthma and allergy in 1748 blood samples collected at birth, 4 and 8 years old using a case-cohort design. Cases of asthma and allergic disease (asthma, eczema or allergic rhinitis) were selected from 4 European birth cohorts (PIAMA, BAMSE, INMA and EDEN), participating in the Mechanism of the Development of Allergy (MeDALL) consortium. DNA methylation profiling was performed using the Illumina Infinium Human Methylation 450k chip that measures 485,512 different CpG sites covering 96% of RefSeq genes. CpG sites related to asthma and allergic diseases were identified using robust linear regression. DNA methylation profiles differ strongly with childhood age, but not so much with the country of origin. We identified 7 CpG sites that are differentially methylated between individuals with and without asthma/allergy at genome wide significant level ($<1.0310^{-7}$), annotated to genes *ETV6*, *LMAN2*, *TNK2*, *RPUSD2*, and *THRA*. These top target CpG sites are currently being validated and replicated in around 6000 DNA samples from 10 European cohorts. To our best knowledge, this is the first study of differential methylation across the genome in relation to asthma and allergic diseases in childhood. Our findings will help to explore the mechanisms of initiation of asthma and allergy.

471M

Epigenome-wide DNA methylation changes in blood from infants with facial clefts in the Norway Facial Clefts Study. *Z. Xu[1], A.J. Wilcox[1], R.T. Lie[3,4,5], J.A. Taylor[1,2].* 1) Epidemiology Branch, National Institute of Environmental Health Sciences of NIH, RTP, NC, USA; 2) Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences of NIH, RTP, NC, USA; 3) Norwegian Institute of Public Health, Oslo, Norway; 4) Department of Global Public Health and Primary Care, University of Bergen, Norway; 5) Medical Birth Registry of Norway, Norwegian Institute of Public Health, Bergen, Norway.

Isolated orofacial clefts are among the most common congenital birth defects. Recent studies have shown that both genetic and environmental factors can increase risks for these defects, however the underlying biological mechanisms are unknown. Using the Illumina HumanMethylation450 BeadChips we conducted an epigenome-wide DNA methylation study in blood from 2–3 day old infants collected in the nationwide population-based Norway Facial Clefts Study. Our study included 453 healthy controls and 312 infants with isolated clefts (94 cleft lip only; 86 cleft palate only; 132 cleft lip & palate). We did not find significant evidence of association between methylation and cleft lip only; and for cleft lip & palate found only 1 significant CpG at a false discovery rate threshold (FDR) q<0.01. However, for cleft palate only cases we found evidence of association with methylation for 556 differentially methylated CpGs (dmCpG) at FDR q < 0.01. There were similar numbers of increasingly methylated dmCpGs (incCpG, 277) and decreasingly methylated CpGs (decCpG, 279). The incCpGs were enriched in genomic regions with lower CpG density and depleted for active histone marks while decCpGs were enriched in higher CpG density regions and enriched for active histone marks. Although the differences in methylation for the other two case groups were not significantly different from controls, > 90% of the dmCpGs identified for cleft palate only had concordant direction of methylation change relative to controls in both the cleft lip only and cleft lip & palate cases. We suggest that this surprising concordance among the three case groups for these epigenetic changes may reflect their common etiologic pathways.

472T

Multiplexed and Quantitative DNA Methylation Analysis Using Long-Read Single-Molecule Real-Time (SMRT) Bisulfite Sequencing. *Y. Yang[1], R. Sebra[1], I. Peter[1], R.J. Desnick[1], C.R. Geyer[2], J.F. DeCoteau[2], S.A. Scott[1].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 2) Cancer Stem Cell Research Group, University of Saskatchewan, Saskatoon, SK, Canada, S7N 4H4.

DNA methylation has important roles in transcriptional regulation, imprinting, X chromosome inactivation and other cellular processes, and aberrant CpG methylation is directly involved in the pathogenesis of human imprinting disorders and many cancers. Targeted analysis of CpG methylation typically is accomplished by pyrosequencing or laborious bisulfite amplicon cloning and sequencing. Although high-throughput targeted methods have been developed to enhance sequencing depth and CpG methylation quantitation, these techniques are limited by amplicon size and short read length. To enable targeted CpG methylation analyses across entire CpG islands and other genomic regions, we developed a quantitative method that is capable of a high degree of multiplexing using long-read single-molecule real-time (SMRT) bisulfite sequencing (Pacific Biosciences). This method was tested using three healthy peripheral blood DNA samples and analyzing the methylation levels of 13 amplicons (609 bp to 1491 bp) encompassing 10 CpG islands. Sequencing read depth ranged from 175X to 1263X per amplicon and the average depth at each CpG site (n=618) was 559X, which resulted in highly quantitative measurements of amplicon CpG methylation ranging from 0.032 to 0.558. In addition, six amplicons (702 bp to 969 bp) encompassing five CpG islands and 473 CpG sites in 30 hematological malignancy cell lines were subjected to SMRT bisulfite sequencing, achieving up to 494X coverage and identifying amplicon CpG methylation ranging from 0.017 to 0.988. Notably, statistical modeling indicated that minimum sequencing depths of 19X and 97X are necessary to accurately quantify methylation levels of 0.050/0.950 and 0.500, respectively, with a margin of error ≤ 0.1 and 95% confidence level. Reproducibility was assessed by SMRT bisulfite sequencing four amplicons of the *TUBGCP3* CpG island, which resulted in highly consistent methylation quantitation (s.d.=0.023). SMRT bisulfite sequencing was validated by comparing CpG methylation levels to both 450K-array (Illumina; 68 CpG sites, r=0.903±0.009) and SureSelect Methyl-Seq (Agilent; 226±79 CpG sites, r=0.927±0.016) data. Coupled with an optimized long-range bisulfite amplification protocol, multiplexed SMRT bisulfite sequencing is an accurate and cost-effective method for targeted CpG methylation analysis that is capable of interrogating ~1.5 kb regions in a single sequencing read, which theoretically covers ~90.5% of CpG islands in the human genome.

473M

DNA methylation in preeclamptic versus normotensive placentas. *K.R. Yeung[1], C.L. Chiu[1], R. Pidsley[2], A. Makris[1,3], A. Hennessy[1], J.M. Lind[1].* 1) University of Western Sydney, Campbelltown, Australia; 2) Garvan Institute of Medical Research, Sydney, Australia; 3) Liverpool Hospital, Liverpool, Australia.

Preeclampsia (PE) is a hypertensive disorder that affects approximately 5% of pregnancies and is a leading cause of maternal and fetal morbidity and mortality worldwide. While the underlying causes of PE are heterogeneous, it is widely accepted that clinical symptoms originate from aberrant placentation early in pregnancy leading to gene expression and physiological changes. There is evidence to suggest that epigenetic mechanisms, such as DNA methylation play a role in placental morphology, development and function. While multiple studies have investigated gene expression changes in preeclampsia, few have looked at epigenetic changes. The aim of this study was to compare DNA methylation profiles in preeclamptic and normotensive placentas. Genomic DNA was extracted from placental samples of PE (n=8), and normotensive pregnancies (n=16). The Illumina Infinium HumanMethylation450K BeadChip was used to assess DNA methylation at > 480 000 CpG sites. All analyses were conducted using publically available Bioconductor packages in R. Following standard quality control and processing steps, linear regression models were used to assess the association between methylation and case-control status at each CpG site. The DMRcate package was used to identify regions of differential methylation. A total of 63 CpG sites across 47 genomic regions were identified as having significantly different DNA methylation in PE placentas compared with controls (absolute beta difference > 5% and FDR value < 0.05). Of these, a number have been identified as having potential involvement in the placental dysregulation observed in PE. Selected gene regions with significantly different methylation profiles in cases versus controls include *SPESP1, NOX5, MIR548H4* (B= 0.28, P = 0.03), *AJAP1* (B= −0.20, P = 0.04), *PARD3* (B= −0.13, P =0.03), *ERRFI1* (B= −0.13, P = 0.03), *PDLIM2* (B= −0.07, P = 0.04), *CSNK1D* (B= −0.08, P = 0.04), and *ITLN1* (B= −0.12, P = 0.04). This study identified genome-wide changes to DNA methylation profiles in placentas from women with PE that may be associated with changes in placental development and function. Functional characterisation of genes identified to be differentially methylated in diseased placentas is required. This research may lead to improved understanding of the pathophysiological disease process.

474T

Analysis of DNA methylation by 450K BeadChip to characterize effects of early life exposures in children. *P. Yousefi, K. Huen, G. Motwani, H. Quach, V. Dav, L. Barcellos, B. Eskenazi, N. Holland.* CERCH, SPH, University of California, Berkeley, Berkeley, CA, United States.

Epigenetic mechanisms, particularly DNA methylation, are a possible link between genetic and environmental determinants of health. As the methylome undergoes rearrangement *in utero* and is susceptible to environmental insults, it may be a mechanism explaining the developmental origins of human disease. Polybrominated diphenyl ethers (PBDEs) are flame retardant chemicals that have been widely used in consumer products over the last 40 years. In the CHAMACOS longitudinal birth cohort, which follows participants from low-income Mexican-American farmworker families in California, we have previously found that prenatal and early-life exposure to PBDEs is associated with a variety of adverse health effects, including decreased fertility, abnormal mental development, and altered thyroid function in children. Further, we have demonstrated that exposure to PBDEs is also associated with altered DNA methylation in LINE-1 repetitive elements in newborn children. Here, we report results of the analysis of site-specific DNA methylation by Illumina Infinium HumanMethylation450K BeadChips assessed in cord blood from CHAMACOS newborns and from the same children at 9 years of age. The Illumina assay simultaneously interrogates methylation at 485,512 CpG sites, in 99%; of all RefSeq Genes and 96%; of CpG Islands. Multivariate linear regression and generalized estimating equation analysis identified 8 differentially methylated positions (DMPs) associated with prenatal PBDE exposure (FDR < 0.05). These DMPs were located in 4 different genes: *AMN, EPHA6, FAM150A,* and *SMAD5.* We additionally identified clusters of CpG sites associated with prenatal PBDE exposure by the bumphunter and ChAMP lasso algorithms. Differential cell counts were used to evaluate how white blood cell type composition modified the relationship between exposures and DNA methylation profiles at birth. New data will be presented on the potential role of altered DNA methylation in mediating the effects of environmental exposures on children's health.

475M

Comparing Statistical Methods for Differential Methylation Identification Using Bisulfite Sequencing Data. *X. Yu[1], S. Sun[2].* 1) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Department of Mathematics, Texas State University, San Marcos, TX.

DNA methylation is an epigenetic modification involved in organism development and cellular differentiation. Identifying differential methylations can help to study genomic regions associated with diseases. Differential methylation studies on single-base-resolution has become possible with bisulfite sequencing (BS) technologies. Several methods have been developed to identify differentially methylated regions (DMRs) in BS data. In this poster, we provide a comprehensive comparison analysis of five DMR identification methods: methylKit, BSmooth, BiSeq, HMM-DM, and HMM-Fisher. We summarize the features of these methods from several analytical aspects, and compare their performances using both simulated and real BS datasets. Based on the simulated analysis results, our discoveries are summarized below. First, parameter settings may largely affect the accuracy of DMR identification. Different from the default settings, the modified parameter settings yield higher sensitivities and/or lower false positive rates. Second, all five methods show higher accuracy in the identification of simulated DMRs that are long and have small within group variations. Third, HMM-DM and HMM-Fisher yield relatively higher sensitivities and lower false positive rates than others, especially in DMRs with large variations. Moreover, in the real data analysis, five methods show low concordance, probably due to the different approaches they have used when addressing the issues in DMR identifications. In addition, we find that among the three methods (methylKit, BSmooth, and BiSeq) that involve methylation estimation, BiSeq can best present the raw methylation signals. Therefore, based on these results, we suggest that users select DMR identification methods based on the characteristics of their data and the advantages of each method. To guarantee a higher accuracy in validation and further analysis, users may choose the identified DMRs that are relatively long and have small within group variations as a priority.

476T

Potential susceptibility factors of congenital heart disease identified by epigenome-wide association study of placenta. *C. Zeng[1], W. Wei[1], J. Zhu[2], Z. Liu[2].* 1) Laboratory of Genome Variations and Precision Biomedicine, Beijing Institute of Genomics, Chinese Academy of Sciences; 2) National office for Maternal and Child Health Surveillance, West China Second University Hospital, Sichuan University.

Congenital heart disease (CHD) is the most common birth defect and genetic, epigenetic, and environmental factors all have contributions to its occurrence. As a nutrition-exchanging organ in embryo development, placenta provides a connection between fetal and environmental factors. Epidemiology studies show that the placentas of CHD fetuses are smaller than that of the healthy ones. To find epigenetic changes in the placenta which may bear the sign of the CHD statues or link the disease with the causative environmental exposure, we performed epigenome-wide association study (EWAS) in 26 placentas of CHD fetuses and 15 placentas of the healthy ones. Considering the anatomy of placenta, each sample was divided into the maternal side and the fetal side for analysis. DNA methylation profiling was performed using Illumina Methylation 450K chip, and loci around SNP sites were ignored. To remove possible contaminations from the blood, we did principle component analysis (PCA) in placenta and blood samples. Two CHD and one control were then excluded due to their closeness to blood samples in PC2.

The DNA methylation of fetal side tissues differed significantly from the maternal side in PC4, which was set to be the covariant in EWAS. In association analysis using linear regression model, 44 loci of significantly differential methylation were observed ($p<10^{-4}$) between cases and controls in maternal side, and their corresponding genes were enriched in alcohol metabolism process, including ST3GAL6, ALDH2, PGD, and SQLE genes, as revealed by ingenuity pathway analysis. Furthermore, 12 of these 44 genes were conserved between placenta and fetal heart, suggesting these genes may harbor important methylation changes in CHD. In fetal side, however, there were only 9 loci, including CASQ2 gene relating to ventricular tachycardia. For the regulation elements, 65 genes showed significant difference ($p<10^{-3}$) on their promoters, TSSs or gene bodies in maternal side. These genes were enriched in small molecular metabolism including nicotine and pathways involved in neuron system development. Our results suggest that environmental exposure such as nicotine and alcohol are related to CHD possibly affecting the methylation of placenta. Methylation changes in certain genes may provide clues for the occurrence of CHD.

477M

An epigenetic map of age-associated autosomal loci in Northern European families at high risk for the Metabolic Syndrome. *Y. Zhang[1], D. Cerjak[1], J. Kent Jr.[2], R. James[1], J. Blangero[2], M. Carless[2], O. Ali[3].* 1) Medicine, Medical College of Wisconsin, Milwaukee, WI; 2) Genetics, Texas Biomedical Research Institute, San Antonio, TX; 3) Pediatrics, Medical College of Wisconsin, Milwaukee, WI.

Purpose: Prevalence of chronic diseases such as cancer, type 2 diabetes, metabolic syndrome (MetS) and cardiovascular disease increases with age in all populations. Epigenetic features, including DNA methylation, change with age and are hypothesized to play important roles in the pathophysiology of age-associated diseases. We searched for age-associated methylation signatures in individuals at high risk for MetS. Methods: Genome-wide DNA methylation was profiled in 192 individuals of Northern European ancestry using the Illumina HumanMethylation450 array. Subjects (aged 6–85 yrs) belonged to 7 extended families and 73% of adults and 32% of children were overweight or obese. A strict marker filtering system was applied and CpG sites that passed quality control were subjected to quantitative assessment against age using the computer package Sequential Oligogenic Linkage Analysis Routines(SOLAR). We then applied a positional density-based clustering algorithm to generate a map of epigenetic "hot-spots" of age-associated genomic segments. Gene/pathway enrichment analyses were performed on these clusters using the KEGG and GO databases. Results: We identified 21,432 age-associated autosomal CpG sites that surpassed the genome-wide significance threshold ($p<3.65\times10^{-7}$ after correction for multiple testing) of which 12,511 are positively associated with age and 8,922 are negatively associated. We identified 246 age-associated differentially methylated CpG clusters (aDMCs); in 185 of these clusters, all the CpG sites within are positively associated with age, in 7 all are negatively associated, and in 54, sites with both directions were observed. Within the 185 positive aDMCs reside 199 genes that are enriched in specific KEGG pathways and GO terms. The three most significantly enriched pathways are the hedgehog signaling pathway (adjusted $p=3.87\times10^{-05}$), maturity onset diabetes of the young (MODY) (adjusted $p=2.58\times10^{-03}$) and basal cell carcinoma (adjusted $p=3.30\times10^{-03}$). Significant GO terms include those related to embryonic development, stem cell development, regulation of cell differentiation and proliferation, insulin secretion, metabolism homeostasis and aging. Conclusion: In a family cohort at high risk for MetS, we identified 185 positively associated aDMCs enriched in several important biological pathways. These findings will guide further studies to determine the mechanisms whereby genomic CpG methylation is involved in aging and aging-related diseases.

478T

A microRNA self-regulatory network in testicular germ cell tumor. *W. Chan[1], Y. Suen[1], S. Gu[1], L. Li[1], B. Chen[1,2].* 1) CUHK-Shandong University Joint Laboratory on Reproductive Genetics, School of Biomedical Sciences, Chinese University of Hong Kong, Shatin, Hong Kong SAR; 2) Department of Biological Science and Biotechnology, School of Science, Wuhan University of Technology, Wuhan, Hubei, China.

It was previously demonstrated that microRNA-199a (miR-199a) was down-regulated in testicular germ cell tumor (TGCT) partially caused by hypermethylation of its promoter. miR-199a is encoded by two loci in the human genome, namely, miR-199a-1 on chromosome (Chr) 19 and miR-199a-2 on Chr 1. Both loci encode the same miR-199a. Another microRNA, miR-214, also locates on Chr 1 and in close vicinity of miR-199a. Previous study revealed that it is co-transcribed with miR-199a-2. However, the biological significance of the co-transcription of miR-199a and miR-214 remains largely unknown. In this study, we showed that miR-199a and miR-214 were concordantly expressed in a TGCT cell line NT2 and in TGCT patient tissues. After 5-aza treatment, miR-199a and miR-214 expression was significantly increased in NT2. Silencing of DNMT1 [DNA (cytosine-5)-methyltransferases 1] with siRNA restored the expression of miR-199a and miR-214, accompanied by de-methylation of the promoters of miR-199a-1/2. Tumor protein p53 (TP53) down-regulated the expression of DNMT1 in NT2 cells and overexpression of TP53 restored the expression of miR-199a and miR-214. In addition, silencing of PSMD10 up-regulated the expression of TP53, while miR-214 over-expression resulted in PSMD10 down-regulation and TP53 up-regulation. Collectively, our findings highlighted a miR-199a/miR-214/TP53/DNMT1 self-regulatory network, which caused the down-regulation of miR-199a, miR-214, and TP53, respectively, as well as the up-regulation of DNMT1 in TGCT. These observations partially explain the mechanism of promoter DNA hypermethylation of miR-199a in TGCT. They also suggest a potential therapeutic approach by targeting the miR-199a/miR-214/TP53/DNMT1 regulatory network in the treatment of TGCT.

479M

Genetic and epigenetic control of regulation of miR-9 expression by alcohol. *A. Pietrzykowski[1, 2], Y. Wang[1], E. Mead[1], A. Thekkumthala[1], L. Tejeda[1], D. Pham[1].* 1) Department of Animal Sciences, Rutgers University, New Brunswick, NJ; 2) Department of Genetics, Rutgers University, New Brunswick, NJ.

Alcohol Dependence (AD) is a complex, chronic disease with high morbidity in humans. Some individuals are genetically predisposed to the development of AD. AD progresses due to frequent alcohol exposures and withdrawals, which cause permanent changes in complex gene expression networks in the brain. However, the molecular mechanisms of these changes and their connection to predisposition are largely unknown. microRNAs are master regulators of gene expression networks. To better understand the role of microRNAs in AD we are using a multiplex, genetic and epigenetic approach focused on a specific microRNA called miR-9. miR-9 is one of the most important brain microRNAs and a regulator of neuronal function. Our and others work showed previously that alcohol affects miR-9, and that it plays a key role in the development of molecular tolerance to the drug. Using mouse striatal cultures we observed that the expression levels of miR-9 depend highly on the interplay between alcohol exposure and withdrawal. Additionally, using human DNA samples (COGA - the Collaborative Studies on Genetics of Alcoholism) we determined that there are specific SNPs in miR-9 regulatory regions (promoters), which are highly associated with AD. Interestingly, presence of minor alleles could change binding of transcription factors as well as affect DNA methylation, which could lead to altered expression of miR-9. Indeed, the levels of miR-9 in the postmortem prefrontal cortex of individuals who suffered from AD are decreased. These results can enhance our insight into the molecular differences between binge drinking and chronic consumption associated with AD and shed some light on genetic predisposition to AD. Together, we observed that microRNAs are intermediaries of alcohol actions in the brain, thus could be considered as attractive, novel therapeutic targets of AD.

480T

DNA methylation patterns of specific L1 loci on the short arm of chromosome 21. *S. Tincher, C.A. Ruggeri, A. Khan, J.L. Doering.* Biology, Loyola University Chicago, Chicago, IL.

While 10–15% of the human genome is composed of heterochromatic DNA, these regions are not included in the completed genome sequence. We are using the short arm of chromosome 21 (HC21p) as a model for understanding the structure and function of heterochromatin. While L1 retrotransposons are underrepresented in heterochromatin, including HC21p, there is a disproportionate number of full length L1s on HC21p. L1 elements are also overrepresented on the X chromosome, and it has been proposed that these L1s are involved in the assembly of heterochromatic nuclear compartments during X chromosome inactivation. The L1 internal promoter is highly conserved and includes a CpG island involved in the DNA methylation silencing mechanism. The activity of full-length L1s is usually suppressed in normal tissues by DNA methylation, but L1 promoters on the inactive X are hypomethylated. The L1s on HC21p may be similarly involved in organizing the heterochromatin in that genomic region, possibly through epigenetic modifications. L1 structure and function, including promoter methylation levels, have not previously been studied in heterochromatic regions of the genome like HC21p. Our goal is to examine the promoter methylation status of specific L1s on HC21p to determine if there is hypomethylation like that on the inactive X. Using bisulfite sequencing PCR on both chromosome-specific hybrid cell lines and leukocytes, we determined the DNA methylation patterns in the promoter regions of six specific full length L1s on HC21p, three of which are located within alphoid sequences. We also included in our analysis two control L1s from the euchromatic HC21q, as well as LRE3, a highly active L1 on HC2. The promoters in these control loci have 93–99% overall methylation levels, while the promoters of evolutionarily older L1 elements on HC21p have lower methylation levels, ranging from 35–57% in both cell types. Interestingly, the one HC21p L1 locus that is of the same recently evolved L1PA2 family as the control loci is less methylated than the controls in both the hybrid cell line and leukocytes, with values of 80% and 73%, respectively. Four CpG sites in the L1 promoter have been specifically implicated in the suppression of L1s. In most of the HC21p L1 promoters these sites are either absent or hypomethylated compared to the same sites in the control loci. Thus, the promoters of the L1s on HC21p are hypomethylated and may be contributing to this region's heterochromatic state.

481M

Inter-individual Variation in Epigenetic Susceptibility to D4Z4 Chromatin Relaxation Explains Clinical Variability in FSHD1 and FSHD2. *R.J.L.F. Lemmers[1], J.J. Goeman[2], P.J. van der Vliet[1], M.P. van Nieuwenhuizen[3], J. Balog[1], G.W. Padberg[4], S. Sacconi[5], R. Tawil[6], S. Tapscott[7], E. Bakker[3], S.M. van der Maarel[1].* 1) Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 2) Department of Biostatistics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 3) Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; 4) Department of Neurology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 5) Centre de rfrence des Maladies neuromusculaires and CNRS UMR6543, Nice University Hospital, Nice, France; 6) Department of Neurology, University of Rochester Medical Center, Rochester, NY, USA; 7) Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

Facioscapulohumeral muscular dystrophy (FSHD) is a common myopathy that progressively affects the facial and upper extremity muscles with marked clinical variability, within and between families. FSHD is caused by contractions of the D4Z4 repeat array on chromosome 4 to 1–10 units (FSHD1), or by mutations in the chromatin modifier *SMCHD1* on chromosome 18 (FSHD2). In somatic cells, SMCHD1 binds to the D4Z4 repeat array to maintain a repressed D4Z4 chromatin structure. Both FSHD situations lead to a partial failure in repeat mediated epigenetic repression of the D4Z4 repeat array and transcriptional derepression of D4Z4-encoded DUX4 that is toxic to muscle cells. We demonstrate that the level of CpG methylation at D4Z4, as a measure of epigenetic repression, linearly correlates with the size of the D4Z4 repeat array. More importantly, we show that the variability in clinical severity in FSHD1 and FSHD2 individuals is caused by individual differences in the degree of CpG hypomethylation at D4Z4. In FSHD2 the epigenetic susceptibility depends on both the D4Z4 repeat array size and the nature of the *SMCHD1* mutation: SMCHD1 open reading frame-preserving mutations being more deleterious for the maintenance of a repressive D4Z4 chromatin state than open reading frame-disrupting mutations. In FSHD1 families, the clinical severity not only depends on the size of the D4Z4 repeat array, but also on the degree of hypomethylation. We show that affected FSHD1 individuals have a greater degree of D4Z4 chromatin relaxation than expected based on their pathogenic D4Z4 repeat array size. In contrast, non-penetrant gene carriers of the same pathogenic array are protected from disease having normal levels of D4Z4 methylation. This study suggests that natural variation in chromatin repressor activity affects D4Z4 methylation providing an epigenetic basis for the striking variability in onset and disease progression, a mechanism that may well translate to common diseases.

482T

Identification of tumor supressor genes modulated by histone acetylation in gastric cancer. *F. Wisnieski[1], D.Q. Calcagno[2], M.F. Leal[1,3], L.C. Santos[1], C.O. Gigek[1], E.S. Chen[1], T.B. Pontes[1], S. Demachki[2], R. Artigiani[4], P.P. Assumpo[2], L.G. Loureno[5], R.R. Burbano[6], M.C. Smith[1]*. 1) Genetics Division, Department of Morphology and Genetic, Federal University of So Paulo, So Paulo, Brazil; 2) Nucleu of Research in Oncology, Joo de Barros Barreto University Hospital, Federal University of Par, Belm, Brazil; 3) Department of Orthopaedics and Traumatology, Federal University of So Paulo, So Paulo, Brazil; 4) Department of Pathology, Federal University of So Paulo, So Paulo, Brazil; 5) Department of Surgical Gastroenterology, Federal University of So Paulo, So Paulo, Brazil; 6) Human Cytogenetics Laboratory, Institute of Biological Sciences, Federal University of Par, Belm, Brazil.

Despite the fact that overall rates of gastric cancer (GC) continue to decline worldwide, the majority of patients are still diagnosed with advanced disease in Western countries. In these cases, surgical resection of the primary tumor offers limited value for a cure and has high morbidity rates. New strategies for early diagnosis and new therapeutic methods in GC continue to be explored. Epigenetic control using histone deacetylase inhibitors, such as trichostatin A (TSA), is a promising cancer therapy. This study aimed to identify genes modulated by TSA in gastric cell lines, and to evaluate the expression of selected genes in gastric resection specimens. ACP02 and ACP03 GC cell lines were treated in triplicate with 250 nM TSA for 24 hours. Differentially expressed genes (DEGs) in treated cell lines compared to controls were identified using microarray assay. The validation of selected DEGs and the assessment of their expression in 46 pairs of primary gastric adenocarcinoma and adjacent non-tumor tissues were performed using qRT-PCR. Possible associations between the expression level of these genes and clinicopathological features were also evaluated. Microarray analysis revealed 42 DEGs (20 upregulated and 22 downregulated genes). Ten relevant functions were observed in which 2 were enriched with a significant number of genes: a) connective tissue, immune, and inflammatory diseases; b) cell cycle, drug, and lipid metabolism. The upregulated DEGs *BMP8B*, *BAMBI*, and *LRRC37A2* were validated and initially selected for gastric tissue analysis. Reduced transcript levels of *BMP8B* were found in diffuse-type and poorly differentiated gastric tumors compared with adjacent non-tumor tissues (p=0.0001 and p=0.006, respectively). In addition, reduced expression of *LRRC37A2* was observed in early staging gastric tumors compared with adjacent non-tumor tissues (p=0.001). On the other hand, the level of expression of *BAMBI* did not differ between gastric tumor and non-tumor samples. Our results demonstrated, for the first time, *BMP8B* and *LRRC37A2* are possible tumor suppressor genes in gastric cancinogenesis and constitute potential targets for epigenetic therapy through histone deacetylase inhibitors.

483M

***KDM5C* SINEUP activates an epigenetic path damaged in XLID/Epilepsy diseases.** *L. Poeta[1], S. Zucchelli[2], A. Padula[1], A. Ranieri[1], S. Filosa[3, 4], P. Collombat[5], M.B. Lioi[6], M.V. Ursini[1], S. Gustincich[2], M.G. Miano[1]*. 1) Institute of Genetics and Biophysics "Adriano Buzzati Traverso", CNR, Naples, Italy; 2) SISSA, Area of Neuroscience, Trieste, Italy; 3) Institute of Biosciences and Bioresources, CNR, Naples, Italy; 4) Istituto Di Ricovero e Cura a Carattere Scientifico, Neuromed, Pozzilli, Italy; 5) Inserm U1091 Diabetes Genetics Team, Nice, France; 6) University of Basilicata, Potenza, Italy.

Mis-steps in histone methylation-demethylation rounds have been directly involved in several forms of Intellectual Disability (ID) with Epilepsy. Lysine-specific demethylase 5C (*KDM5C*) is an X-linked gene which encodes a chromatin JmjC eraser with H3K4me2/3 demethylase activity. *KDM5C* is frequently mutated in a spectrum of XLID and/or malignant Epilepsy. It functions as a transcriptional repressor and interacts with REST/NSRF, a master epigenetic hub that is critical for transition of neural progenitors to neurons. Noteworthy, a defective KDM5C-H3K4me3 path has been found in association with a mis-regulation of XLID/Epilepsy effector genes. We report here the application of SINEB2 long non coding RNAs (SINEUPs) technology to target KDM5C-H3K4me3 axis. SINEUPs are a new functional class of antisense lncRNAs with embedded repetitive elements that activate the production of target proteins to physiological levels. We generated synthetic lncRNAs antisense to *KDM5C*. In transfected cell lines, HEK293T and P19, we obtained increased levels of endogenous KDM5C protein by using both human and murine SINEUPs. As expected, increasing doses of transfected antisense *KDM5C/Kdm5C* SINEUPs titrated KDM5C protein, in both cell lines. When increasing amounts of the SINEUP were co-transfected with human KDM5C into HEK293T cells, dose-dependent KDM5C protein upregulation was observed in the absence of change in the quantity of exogenous *KDM5C* mRNA. Given the H3K4 demethylase activity of KDM5C, we therefore tested the level of H3K4me3 and verified that the increase amount of KDM5C inversely correlates with a decrease in H3K4me3 signal. However, the snapshot of this captivating stratagem may become more well-defined once we will decipher the effect of KDM5C knock-up in *Arx* KO/*Kdm5C*-depleted ES cell line. Noteworthy, SINEUPs are a new class of RNA therapeutics that allows to design gene-specific approaches to treat ID, malignant epilepsy diseases and many other neuropathologies with seizure linked to insufficient activity of chromatin and/or transcriptional regulators. *Italic Text*.

484T

The impact of antisense transcription on epigenetic signals. *C. Wadelius[1], S. Bornelöv[2], J. Komorowski[2,3]*. 1) Dpt of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 2) Dpt of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden; 3) Polish Academy of Sciences, Warszawa, Poland.

Several histone post-translational modifications are preferentially located in the promoter region of genes and are associated to the promoter activity. It is debated whether the histone modifications regulate the gene expression i.e., the histone code hypothesis or if they are put there as a consequence of transcription. In an earlier study we observed that the histone 3 acetylation signal upstream of the transcriptional start site (TSS) was lower in unidirectional compared to bidirectional genes. Following this observation, we hypothesized that the transcribed region is modified during transcription and that the observed upstream signal is caused by transcription in the opposite direction from the TSS. Promoters directing transcription in both directions are frequent in the genome, which would explain the common occurrence of upstream signals. Here we identified bidirectional or unidirectional genes based on TSS identified from cap analysis of gene expression and database searches. We compared histone modification signals between these two classes of genes across several cell types. We have found significant differences for well-known histone modifications, e.g. H3K4me3, H3K9ac and H3K27ac for which the upstream signal is higher in the bidirectional genes. Furthermore, we have compared transcription factor bindings between bidirectional and unidirectional genes and found examples of differences in their prevalence and position relative to the TSS. In conclusion, our results support the model of histone modifications occurring in transcribed regions and thus being a consequence of transcription. In addition, we have identified transcription factors which may be involved in the direction of transcription initiation.

485M

Methylation analysis and diagnostics of Beckwith-Wiedemann syndrome in 1000 subjects. *A. Ibrahim[1,2], G. Kirby[3], C. Hardy[4], R. Dias[3], L. Tee[3], D. Lim[3,4], J.N. Berg[2], F. MacDonald[4], P. Nightingale[5], E.R. Maher[1,3].* 1) Department of Medical Genetics, University of Cambridge and NIHR Cambridge Biomedical Research Centre, Cambridge, United Kingdom; 2) Department of Clinical Genetics, University of Dundee, Dundee, United Kingdom; 3) Centre for Rare Diseases and Personalised Medicine, School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Birmingham, United Kingdom; 4) West Midlands Regional Genetics Service, Birmingham Women's Hospital, Birmingham, United Kingdom; 5) Wellcome Trust Clinical Research Facility, University Hospitals Birmingham NHS Foundation Trust, Queen Elizabeth Hospital, Birmingham, United Kingdom.

Background - Beckwith-Wiedemann syndrome (BWS), a congenital overgrowth disorder with variable expressivity and a predisposition to tumorigenesis, results from disordered expression and/or function of imprinted genes at chromosome 11p15.5. There are no generally agreed clinical diagnostic criteria, with molecular studies commonly performed to confirm diagnosis. In particular, methylation status analysis at two 11p15.5 imprinting control centres (IC1 and IC2) detects up to 80% of BWS cases (though low-level mosaicism may not be detected). In order to evaluate the relationship between the clinical presentation of suspected BWS and IC1/2 methylation abnormalities we reviewed the results of >1000 referrals for molecular diagnostic testing.

Methods and Results - 507/1091 (46.5%) referrals had a positive diagnostic test for BWS. The frequency of tumours was 3.4% in those with a molecular diagnosis of BWS. Previously reported genotype-phenotype associations with paternal uniparental disomy, IC1, and IC2 epimutation groups were confirmed and potential novel associations detected. Predictive values of previously described clinical diagnostic criteria were compared, and although there were differences in their sensitivity and specificity, receiver operating characteristic (ROC) analysis demonstrated that these were not optimal in predicting 11p15.5 methylation abnormalities. Using logistic regression, we identified clinical features with the best predictive value for a positive methylation abnormality. Furthermore, we developed a weighted scoring system (sensitivity - 75.9% and specificity - 81.8%) to prioritise patients presenting with the most common features of BWS, and ROC analysis demonstrated superior performance (area under the curve - 0.85; 95% CI: 0.83–0.87) compared to previous criteria.

Conclusion - We suggest that this novel tool will facilitate selection of patients with suspected BWS for routine diagnostic testing and so improve the diagnosis of the disorder.

486T

Identification of imprinted 4q35 variant associated with the combined asthma-plus-rhinitis phenotype using both genetic and epigenetic data. *C. Sarnowski[1,2], C. Laprise[3], M. Moffatt[4], G. Malerba[5], A. Morin[3,9], Q. Vincent[6], K. Rohde[7], M-H. Dizier[1,2], J. Esparza-Gordillo[7], P. Margaritte-Jeannin[1,2], L. Liang[8], Y-A. Lee[7], P-F. Pignatti[5], W.O.C. Cookson[4], T. Pastinen[9], M. Lathrop[9], F. Demenais[1,2], E. Bouzigon[1,2].* 1) U946, INSERM, PARIS, France; 2) Universit Paris Diderot, Sorbonne Paris Cit, Institut Universitaire d'Hmatologie, France; 3) Universit du Qubec Chicoutimi, Canada; 4) National Heart Lung Institute, Imperial College, UK; 5) Section of Biology and Genetics, Department of Life and Reproduction Sciences, University of Verona, Italy; 6) U1163, INSERM, PARIS, France; 7) Max-Delbrck-Center for Molecular Medicine (MDC), Berlin, Germany; 8) Department of Epidemiology, Harvard School of Public Health, Boston, USA; 9) McGill University, Montreal, Canada.

Several lines of evidence support the role for epigenetic mechanisms such as imprinting in asthma and allergic diseases. We conducted a genome-wide linkage scan for the combined asthma-plus-rhinitis phenotype (AST+AR) in 615 families from European ancestry (French (Epidemiological study on the Genetics & Environment of Asthma, EGEA), British (Medical Research Council Asthma, MRCA) and Italian (Verona)) and detected a linkage signal in the 4q35 region with increased evidence when accounting for maternal imprinting ($p=7\times10^{-5}$). We then investigated further this region using a panel of 1,300 SNPs (spanning 6 Mb) genotyped in 162 families from the EGEA study (207 offspring). We tested the association between these SNPs and AST+AR using the Parent-of-Origin-Likelihood Ratio Test (PO-LRT) which allows detecting parent-of-origin and/or maternal genotype effects. We identified 18 SNPs associated with AST+AR at $p<0.005$ that were investigated for replication in 152 asthmatic French Canadian families from the Saguenay-Lac-Saint-Jean study (SLSJ). Combination of EGEA and SLSJ results under a maternal imprinting best-fitting model showed evidence for association for one SNP ($p_{meta}=4\times10^{-5}$), lying at 1.6 Mb from the linkage peak, that accounted for most of the 4q35 linkage signal. Many cis-regulatory elements (enhancers, silencers, insulators) are described in a 50 kb surrounding region of the replicated SNP. Using the Quantitative Transmission Disequilibrium Test (QTDT), we tested for association between the replicated SNP and 26 DNA methylation probes of that region, measured in white blood cells of 159 individuals (40 SLSJ families), while accounting for parent-of-origin effect and adjusting for AST+AR. Maternally inherited risk allele of the replicated SNP was associated with increased methylation of the top-ranked probe ($p<10^{-5}$ after permutations). This probe was located at 529 bp from the SNP and lies within regulatory elements that include a predicted active promoter in lung fibroblasts, DNase I hypersensitive clusters, and binding sites of two transcription factors involved in inflammatory response initiation (RelA and NF-κB). This study identified a maternally imprinted SNP that affects AST+AR through an epigenetic mechanism. Funded: Conseil Rgional Ile de France, ANR-GWIS-AM, EC-FP6, Canada Research Chair, Canadian Institutes for Health Research.

487M

Prediction of imprinted genes based on the genome-wide methylation analysis. *N. Ternikova[1,2,4], N. Tnisson[1], K. Lokk[1], A. Salumets[3,5,6], A. Metspalu[1,2], R. Mgi[2].* 1) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 2) Estonian Genome Center, University of Tartu, Tartu, Estonia; 3) Competence Centre on Reproductive Medicine and Biology, Tartu, Estonia; 4) Quantitative Genetics Group, Public Health Genomics Unit, Institute for Molecular Medicine, Helsinki, Finland;; 5) Institute of Biomedicine, University of Tartu, Tartu, Estonia; 6) Department of Obstetrics and Gynecology, University of Tartu, Tartu, Estonia.

Genomic imprinting is an epigenetic gene-marking phenomenon that is established in germline. The importance of genomic imprinting is to regulate gene expression in parental-specific manner. One of the new findings has shown that imprinting does not always occur similarly across all tissues. Recent studies suggest a more complicated scheme of tissue-specific promoters and epigenetic marks. We test the hypothesis that imprinted genes can be predicted by the methylation level. In case of genomic imprinting one of the parental copies is usually silenced through DNA methylation. Based on this knowledge we expect semi-methylation in imprinted genes. In order to prove this hypothesis we analysed the DNA methylation in well-known imprinted genes across the tissue panel from the same individuals. All 17 tissues from every 4 individuals were collected during the autopsy. DNA methylation analysis of the total 72 tissue samples and controls was performed with the Illumina Infinium HumanMethylation450 BeadChip in Estonian Genome Centre. We used Levene's test for comparison of known imprinted genes with the rest of the genes captured by 450K methylation array. As a result, all imprinted genes (n=92) demonstrated less variability in the methylation level ($p < 0.01$) across all 17 tissues when compared to non-imprinted genes. We also visualized the CpG patterns of known imprinted genes across all tissues. Each CpG was annotated to its exact location in the genome in exon, gene body or UTR region. Visualized CpG patterns also confirmed tissue-specific nature of imprinted genes. For example, gallbladder shows medium methylation of KCNQ1DN gene as CpG sites are only partially methylated, while in isciatic nerve the CpG sites are not methylated. Using this mapping method, we have narrowed down the list of potential candidate genes to 3 000. We found that some genes meet the criteria for candidate imprinted genes in all somatic tissues, while other genes meet those criteria only in some of the tissues. As the next step we are using the RNAseq data to further narrow down the list of candidate genes. Our method can be regarded as a tool to identify the tissue specificity of the already established imprinted genes as well as to discover new imprinted genes across the whole human genome. Our method will give a better understanding of the nature of the imprinted genes.

488T

Characterizing the processing, localization, and function of *Snord116* noncoding RNAs at the Prader-Willi locus. *R. Coulson, W. Powell, D. Yasui, J. Aflatooni, S. Wong, J. LaSalle.* Medical Microbiology and Immunology, Genome Center, M.I.N.D Institute, University of California Davis, Davis, CA.

Prader-Willi syndrome (PWS) is a neurodevelopmental disorder caused by paternal deficiency of the imprinted 15q11-13 noncoding RNA cluster SNORD116. This repeat cluster is GC-skewed, resulting in R-loop formation, histone displacement, and chromatin decondensation specifically of the paternal allele upon neuronal transcription. SNORD116 produces two noncoding RNAs: SNORD116 snoRNAs, which localize to the nucleolus of maturing neurons, and 116 host gene (116HG), which forms an RNA cloud at its paternally decondensed site of transcription. To understand the processing, localization, and functional relevance of each component of SNORD116, we engineered novel transgenic mice to test complementation of the PWS mouse model, Snord116del. Complete transgene wild-type (Ctg/WT) mice containing 27 copies of the transgenic Snord116 repeat were generated and bred to Snord116del males to produce offspring lacking paternal Snord116 but expressing the transgene (Ctg/Snord116del). RNA fluorescence in situ hybridization (FISH) analysis of brain showed that spliced 116HG localized in a distinct cloud at its decondensed transcription site and mature snoRNAs localized to the nucleolus in wild-type (WT) but not Snord116del neurons, as previously described. In Ctg/WT neurons, a single 116HG RNA cloud was detected that was significantly larger than that observed in WT neurons, suggesting that the Ctg-derived 116HG colocalizes with the endogenous 116HG RNA cloud at the paternal Snord116 locus. In addition, snoRNAs were detected in the nucleolus at a significantly higher level in Ctg/WT than in WT neurons. In non-neuronal Ctg/WT tissues and all Ctg/Snord116del tissues, no RNA cloud or snoRNAs were detected. qRT-PCR analysis demonstrated that splicing of Snord116 was largely restricted to neuronal tissues of both Ctg/WT and Ctg/Snord116del mice, despite expression of the transgene in many tissues. These combined results suggest that processing of the Snord116 transcript is dependent on neuronal factors and/or chromatin states, including the decondensed Snord116 paternal allele. Analyses of a spliced 116HG/Snord116del transgenic mouse is underway to further test the hypothesis that localization of the 116HG RNA cloud requires the decondensation of paternal Snord116 in neurons. Understanding how DNA-RNA interactions mediate the processing and localization required for Snord116 function and phenotypic rescue is critical for the development of effective PWS therapies in the future.

489M

Evolutionarily-conserved Imprinting Between Mouse and Human Orthologs Identified 17 Novel Candidate Genes for Human Imprinting Disorders. *E.J. Bhoj[1], F. Rajabi[2], E.H. Zackai[1], W.H. Tan[2].* 1) Department of Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Boston Children's Hospital, Boston, MA.

There are about a dozen human imprinted disorders that cause a wide range of dysmorphic and neurocognitive phenotypes. Some are only recently being appreciated as they present with a subtle phenotype, such as maternal uniparental disomy (UPD) of chromosome 20, which results in failure to thrive in infancy. We hypothesize that there are additional human imprinted disorders with subtle or late-onset effects that have not been previously recognized. To identify these novel imprinting disorders we have searched the literature for every human imprinted gene, identified each mouse ortholog, determined its imprinting status in the mouse, and collected information about the function of each mouse gene. We anticipate that by focusing on those genes that through evolutionary conservation are imprinted in both the mouse and human, we can derive information about their potential pathology in humans. Using this method we were able to discover 17 human genes that are imprinted in both humans and mice, and have functional data in mice to suggest that disturbed expression would lead to an abnormal phenotype in humans. These genes are: DIRAS3, TP73, SLC22A3, DDC, GRB10, MAGI2, PEG10, PPP1R9A, CALCR, DLGAP2, GLIS3, INPP5F, ANO1, SLC38A4, GATM, PEG3, and NLRP2. Many of these genes have mouse models that demonstrate abnormal phenotypes as varied as profound tracheomalacia, hydrocephalus, dysregulated fetal growth, neurologic disorders, impaired lactation, chronic inflammation and infection, abnormal social functioning, calcium metabolism derangement, and a range of congenital dysmorphisms. If there are additional patients with deleterious mutations in these genes they would likely not be identified through whole exome sequencing as the mutation could be inherited from a normal parent, and therefore discarded in the variant analysis. In addition, patients with UPD may lead to an abnormal phenotype. If there is a large region of isodisomy it may be identified by single nucleotide polymorphism (SNP) array, but smaller regions of isodisomy or any heterodisomy will be missed by conventional SNP array. Therefore, evolutionary conservation is a powerful tool to identify novel genes involved in human imprinting disorders with subtle or late-onset effects.

490T

Random monoallelic expression in neural stem cells, induced pluripotent stem cells and neural committed cells. *A.R. Jeffries[1,2], D.A. Uwanogho[2], G. Cocks[2], L.W. Perfect[2], E.L. Dempster[1], J. Mill[1], J. Price[2].* 1) University of Exeter Medical School, Medical Research, RILD level 4, Royal Devon & Exeter Hospital, Barrack Road, Exeter. EX2 5DW. United Kingdom; 2) King's College London, Centre for the Cellular Basis of Behaviour, The James Black Centre, 125 Coldharbour Lane, London. SE5 9NU. United Kingdom.

Random monoallelic expression is where the choice of which alleles to express is made at random early in development and maintained in subsequent clonal progeny. This results in cellular level heterogeneity of allele specific gene expression together with any resulting phenotypic effects. Up to 10% of autosomal genes can show random monoallelic expression yet little is known on the precise mechanisms behind the allelic choice made early in development.

We use epigenetic reprogramming to investigate the effects on previously characterized random monoallelic expressed genes. We also neuralize the resulting induced pluripotent stem cells (iPSCs) to a more committed state to further examine allelic choice.

Results: Genes which previously showed random monoallelic expression in neural stem cells reverted to biallelic expression after epigenetic reprogramming into iPSCs. Lineage commitment into neural stem cells followed by clonal isolation revealed a number of genes showing random monoallelic expression choice, many of which represented new gene loci. Global DNA methylation analysis was also performed using the Illumina 450k methylation beadchip. Clones undergoing monoallelic expression showed increased DNA methylation at gene promoters compared to sister clones which showed biallelic expression.

In summary, random monoallelic expression is lost during epigenetic reprogramming but re-established in a stochastic fashion when cells are committed towards a neural lineage. Clonal differences in allelic expression status show associated promoter level DNA methylation differences.

491M

Somatic and genetic variations in regulatory regions revealing discordance between monozygotic twins. *K. Kim[1], HJ. Ban[2,3], J. Seo[4], K. Lee[1], M. Yavartanoo[1], SC. Kim[5], K. Park[2], SB. Cho[2], JK. Choi[2].* 1) Department of Bio and Brain Engineering, KAIST, Daejeon 305–701, Republic of Korea; 2) Department of Biomedical Informatics, Center for Genome Science, National Institute of Health, KCDC, Choongchung-Buk-do 363–951, Republic of Korea; 3) Division of Molecular and Life Sciences, Hanyang University, Ansan, Gyeonggi-do 425–791, Republic of Korea; 4) Research Institute of Bioinformatics, Omicsis, Inc. Daejeon 305–333, Republic of Korea; 5) Samsung Genome Institute, Samsung Medical Center, Seoul 135–710, Republic of Korea.

Open chromatin has important role as regulatory region in transcriptional process, and genetic variations in open chromatin have been known to contribute to discordance of gene expression and phenotype. However, the effect of somatic variation on regulatory region still remains to be elucidated. In this study, we performed FAIRE sequencing (~72X) to analyze open chromatin, array-based genotyping across 72 monozygotic (MZ) twins to identify genetic and somatic variations that can explain twin discordance in chromatin accessibility. First, we focused on the spectrum of somatic and genetic sequence variations underlying discordant open chromatin mainly via the disruption of transcription factor binding sites (TFBSs). The change of chromatin accessibility due to C:G > A:T transversions seemed to be subject to purifying selection, because of their strong impact on chemical bonding, as indicated by a lower frequency of polymorphisms than somatic mutations. Of note, CpG located in TFBSs were found to be mutated or polymorphic only 1/5 as often as expected. In particular, those CpGs whose methylation is specifically regulated during cellular differentiation appeared to be protected from the high mutation rates associated with 5'-methylcytosines, thereby implying that the spectrum of CpG variations may be shaped at developmental level but not through natural selection. Next, we used genetic association mapping of within-pair chromatin differences to search for cases in which twin siblings with a particular genotype associated with chromatin discordance at the relevant locus. At an FDR of 0.01 for local associations, ~1000 chromatin sites were found to be differentially accessible depending on the genotype of a nearby locus. This demonstrates that epigenetic differences can bring about regulatory variations through interactions with genetic factors. Remarkably, poised promoters present high levels of chromatin discordance in association with either somatic mutations or genetic interactions and therefore most likely induce differential regulatory responses to environmental stimuli. Our observations illustrate how somatic mutations and genetic polymorphisms may contribute to regulatory discordance.

492T

Deciphering the role of DNA methylation in osteoporosis using MeDIP-seq: A twin-based study design. *J.A. Morris[1,2], F. Gao[3], V. Forgetta[2], Y. Xia[3], P.C. Tsai[4], W. Yuan[4], C.M. Greenwood[1,2,5,6], E. Grundberg[1], T.D. Spector[4], J. Wang[3], J.T. Bell[4], J.B. Richards[1,2,6].* 1) Department of Human Genetics, McGill University, Montreal, Canada; 2) Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Canada; 3) BGI-Shenzhen, Shenzhen, China; 4) Department of Twin Research & Genetic Epidemiology, King's College London, London, United Kingdom; 5) Department of Oncology, McGill University, Montreal, Canada; 6) Department of Epidemiology, Biostatistics, & Occupational Health, McGill University, Montreal, Canada.

Osteoporosis is a common disease that leads to an increased risk of fracture through decreased bone mineral density (BMD) and defects in bone. Osteoporotic fractures are a socio-economic burden and are increasing in incidence with the ageing of industrial societies. Since 2007, numerous genome-wide association studies (GWAS) for osteoporosis and related traits have identified multiple common genetic variants associated with BMD, however little is known concerning the role of epigenetic changes influencing BMD. The aim of this study is to investigate the association between epigenetic changes and BMD phenotypes by undertaking an epigenome-wide association study of monozygotic (MZ) and dizygotic (DZ) individual twins from EpiTwin (http://www.epitwin.eu/), a TwinsUK project (http://www.twinsuk.ac.uk/).

We have completed whole genome methylated DNA immunoprecipitation sequencing (MeDIP-seq) for 1,248 MZ twins and 320 DZ twins to quantify genome-wide methylation levels. Methylation levels for each 500 bp region (250 bp overlap) of the genome will be compared with BMD for all twins using a linear mixed effects model to account for the twin relationships. Additional effects will include age, gender, and batch. Significant regions will be differentially methylated regions (DMRs), identified to be associated with BMD through variation in methylation levels. Analysis is currently ongoing. Each twin has three measurements for BMD at total forearm, total hip, and total spine.

Identified DMRs suggest that region-specific variation of methylation levels is associated with osteoporotic phenotypes. Significant findings will be qualitatively compared to regions already identified by GWAS to influence BMD, where any overlaps would suggest an additional mechanism of influence for the region on BMD. DMRs could also identify novel drug targets to treat osteoporosis, as the majority of osteoporosis drug targets have been identified through studying the determinants of BMD.

493M

Epigenome-wide association study of sexual orientation in monozygotic twins. *T.C. Ngun[1], W. Guo[2], N.M. Ghahramani[3], K. Purkayastha[1], F.J. Sanchez[4], S. Bocklandt[1], M. Zhang[2, 5], M. Pellegrini[6], E. Vilain[1].* 1) Department of Human Genetics, David Geffen School of Medicine at University of California Los Angeles (UCLA), Los Angeles, CA, USALos Angeles, CA; 2) Bioinformatics Division and Center for Synthetic & Systems Biology, TNLIST, Tsinghua University, Beijing 100084, China; 3) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 4) Department of Counseling Psychology, The University of Wisconsin-Madison, WI, USA; 5) Department of Molecular and Cell Biology, Center for Systems Biology, The University of Texas at Dallas, Richardson, TX 75080, USA; 6) Department of Molecular, Cellular, and Developmental Biology, UCLA, Los Angeles, CA, USA.

Sexual orientation is one of the most pronounced sex differences in the animal kingdom. Although upwards of 95% of the general population is heterosexual, a small but significant proportion of individuals (3–5%) is homosexual. Previous work has consistently shown that sexual orientation has a strong genetic component. Male sexual orientation has been linked to several genomic loci including Xq28, 8p12, and 10q26, which is an imprinted region. As with other complex traits, environmental factors also play an important role. For instance, monozygotic twins show substantial levels of discordance for this trait. Additionally, each biological older brother a man has increases the probability of him being gay by 33%. Taken together, the evidence strongly suggests a role of epigenetic mechanisms in the regulation of sexual orientation. We aim to elucidate the molecular mechanisms underlying sexual orientation by investigating DNA methylation patterns on a genome-wide basis in 34 monozygotic male twin pairs that are discordant for sexual orientation using reduced representation bisulfite sequencing (RRBS). As a control population, we also performed RRBS on 10 monozygotic twin pairs concordant for homosexuality. Subjects were classified as homosexual or heterosexual based on their sexual feelings (as measured by the Kinsey scale). We first consolidated nearby CpG sites into short regions (100–500 bp) to increase the signal to noise ratio in our data. A linear mixed modeling approach was used to identify regions that were significantly associated with sexual orientation (sexual orientation DMRs). At a false discovery rate of 10%, we identified 136 sexual orientation DMRs. One of these regions was in 10q26 but there were no other overlaps between our data and previous linkage findings. This discrepancy was not surprising as prior studies primarily used families with evidence of maternal loading, which was not a criteria for inclusion in our study. Several genes associated with sexual orientation DMRs were involved in the androgen signaling pathway: (1) *AR*, which encodes the androgen receptor; (2) *ZDHHC7*, which regulates the localization of sex steroid receptors; and (3) *SULT1A3*, which is part of the alternative androgen synthesis pathway. Our findings demonstrate that numerous epigenetic changes are associated with sexual orientation in humans and that our approach has the potential to identify novel genes that influence this trait.

494T

Genes that escape from X inactivation vary in mouse tissues. *C. Disteche[1,4], J. Berletch[1], W. Ma[2], F. Yang[3], J. Shendure[2], W. Noble[2], X. Deng[1].* 1) Dept Pathology, Univ Washington, Seattle, WA; 2) Dept Genome Sciences, Univ Washington, Seattle, WA; 3) Fred Hutchinson Cancer Research Center, Seattle, WA; 4) Dept of Medicine, Univ Washington, Seattle, WA.

X chromosome inactivation (XCI) silences most genes on one X chromosome in female mammals, but some genes escape XCI. Surveys in cultured human/mouse hybrid cells and in cell lines from individuals with skewed XCI have shown that about 8–15% of human genes consistently escape XCI, 10–13% display variable levels of escape, and 10–20% vary between cell lines and individuals. Escape from XCI results in significant sexual dimorphisms in levels of gene expression, and bi-allelic expression of at least some escape genes is important for a normal phenotype in human females. Indeed, the presence of a single X chromosome (45,X) results in Turner syndrome. To identify escape gene in vivo and to explore molecular mechanisms that regulate this process we analyzed the allele-specific expression and chromatin structure of X-linked genes in mouse tissues and cells with skewed XCI and distinguishable alleles based on single nucleotide polymorphisms. Using a new method to estimate allelic expression, we demonstrate a continuum between complete silencing and significant expression from the inactive X (Xi). Few genes (2–3%) escape XCI to a significant level and only a minority differs between mouse tissues, suggesting stringent silencing and escape controls. Allelic profiles of DNase I hypersensitivity and RNA polymerase II occupancy of genes on the Xi correlate with escape from XCI. Allelic binding profiles of the DNA binding protein CCCTC-binding factor (CTCF) in different cell types indicate that CTCF binding at the promoter correlates with escape. Importantly, CTCF binding at the boundary between escape and silenced domains may prevent the spreading of active escape chromatin into silenced domains.

495M

Male Rett syndrome: Clinical profiling and insights into epigenomics. *J. Duis[1], S. Gupta[2], S. Naidu[2].* 1) The McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins, Baltimore, MD; 2) Department of Neurogenetics, Kennedy Krieger Institute, Baltimore, MD.

Rett syndrome (RTT [MIM 312750]) is an X-linked disorder caused by mutations in the gene encoding methyl CpG binding protein 2 (MECP2), which is known to be essential in neuronal function. Once thought to be lethal, male Rett is now extensively described. We report 7 cases of male Rett with special attention to the clinical profile including phenotype and genotype correlations. We review what is known in the literature, report novel mutations with insights into inheritance of these mutations, the roles of maternal inheritance gleaned from these cases, and how this facilitates global understanding of the roles epigenetics may play in this disorder.

496T

Examining Escape from X Chromosome Inactivation and Sex-Differential Gene Expression. *P.G. Bronson, T.R. Bhangale, K.T. Cuenco, W. Ortmann, R.R. Graham, T.W. Behrens.* Genentech, Inc., South San Francisco, CA.

Autoimmune disease displays a predilection for females, with a sex ratio of up to 12:1 for systemic lupus erythematosus in women. The cause of this sex imbalance is unknown, but hormones are suspected to play a role. A gene dosage effect for the X chromosome may also be involved. Female embryonic cells undergo X chromosome inactivation (XCI), through CpG hypermethylation and chromatin remodeling initiated by the noncoding RNA XIST, to randomly turn all but one X into a compact, inactive state. Previous studies have established that ~15% of the 1,031 genes on X escape XCI. We analyzed 195 lymphoblastoid cell lines (LCLs) from females with RNA-Seq expression data (Geuvadis RNA Sequencing Project of 1KGP) and DNA sequence genotypes available (1KGP) to identify a subset of "monoclonal" LCLs. We defined a "monoclonal" LCL as one in which the same X was inactivated in ≥95% of the LCLs, as determined by informative (heterozygous variants) in non-escaping genes on X. RNA-Seq reads were mapped to the reference genome using GSNAP, which performed splice-junction-aware split-read alignment. Using known variants in the samples, GSNAP performed SNP-tolerant alignment and removed the reference-bias in read mapping. Gene expression levels were quantified in terms of RPKMs using DEseq. We identified 63 "monoclonal" LCLs of the 195 examined, and observed escape from XCI in 80 genes using informative SNPs; escape was defined as having >=1 heterozygote cell line with skewed XCI (≥90% of RNA reads coming from only one of the two alleles). From the literature we identified an additional 41 escaping X genes for a total of 121 genes that escape XCI. The Kolmogorov-Smirnov test was used to test 13,615 genes for expression differences between females (n=241) and males (n=214). Significant gene expression differences, defined as >1.2 fold difference and $p \leq 1.5 \times 10^{-3}$, were observed in 13.3% of escaped genes. Expression differences were observed for an additional 7 X genes and 16 autosomal genes, including the T-cell activation RhoGTPase activating protein (an autoimmunity locus) and adrenomedullin (previously reported to have higher expression in females). Further work is underway to annotate the genes that escape XCI, and those that show differential expression between females and males. Additional datasets are also being examined. Understanding the mechanisms that control expression of genes on the X chromosome may provide important new insights into autoimmunity.

497M

Placental microRNAs as potential biomarkers for noninvasive detection of trisomy 21. *H.M. Ryu[1,2], J.H. Lim[1], H.J. Kim[1], A.R. Oh[1], S.Y. Kim[1], D.E. Lee[1], S.Y. Park[1], Y.J. Han[2], J.S. Choi[2], K.H. Choi[2].* 1) Laboratory of Medical Genetics, Medical Research Institute, Cheil General Hospital and Women's Healthcare Center, Seoul, Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women's Healthcare Center, Kwandong University College of Medicine, Seoul, Korea.

Objectives: The discovery of fetal nucleic acids in the plasma of pregnant women has led to the development of noninvasive prenatal tests (NIPT). The occurrence of placental microRNAs (miRNAs) in maternal blood has been demonstrated and then proposed as a particular class of molecular biomarkers for the NIPT. However, miRNAs for NIPT of fetal trisomy 21 (T21) have yet to be reported. The objective of this study was to discover a panel of placental miRNAs as potential novel biomarkers for NIPT of T21 fetus and to predict biological functions of new biomarkers using bioinformatics tools. Method: Using microarray-based expression profiling, we compared the expression levels of miRNAs in whole blood samples from non-pregnant women, whole blood samples from pregnant women, euploid fetal placenta samples from pregnant women, and T21 fetal placenta samples from pregnant women. We analyzed the differentially expressed miRNAs according to the presence or absence of disease and tissue type (p value < 0.05 and two-fold expression change). Potential target genes of miRNAs were predicted using the miRBase program. To predict their functions, the functional annotation tools provided by the WebGestalt database were used. Results: We identified 299 miRNAs which reasonably separate the whole blood from the placenta. Among identified miRNAs, 150 miRNAs up-regulated in the placenta and 149 miRNAs down-regulated. Most of up-regulated miRNAs in the placenta were members of mir-498, mir-134, and mir-127 clusters as placenta specific miRNA located on chromosome 19 and 14. Among up-regulated miRNAs in the placenta, mir-1973 and mir-3196 were highly expressed in the T21 placenta than in the euploid placenta. The two miRNAs potentially regulated 203 target genes. Interestingly, the genes were significantly associated with T21-related disorder such as congenital abnormalities and mental disorders. The genes were involved in various pathways such as ErbB signaling, calcium signaling, chemokine signaling, focal adhesion, MAPK signaling, endocytosis, and pathways in cancer. Conclusions: In conclusion, our study indicated miRNAs that may be potential biomarkers for NIPT of fetal T21 as well as provided new insights by regulation of miRNAs into the molecular mechanisms in the placenta of T21 fetus.

498T

Detection of *in vivo* G-quadruplex structure of the *ANXA5* promoter that contributes to the recurrent pregnancy loss. *H. Inagaki[1], S. Ota[1], H. Miyamura[1,2], M. Tsutsumi[1], T. Kato[1], H. Nisizawa[2], I. Yanagihara[3], H. Kurahashi[1].* 1) Molecular Genetics, ICMS, Fujita Health University, Toyoake, Aichi, Japan; 2) Obstetrics and Gynecology, Fujita Health University School of Medicine, Toyoake Aichi, Japan; 3) Developmental Medicine, Osaka Medical center and Research Institute for Maternal and Heath, Izumi, Osaka, Japan.

Recent findings have highlighted the possibility that polymorphisms within the annexin A5 gene (*ANXA5*) promoter contribute to the etiology of recurrent pregnancy loss. However, the underlying mechanisms are unknown. An M2 haplotype of the polymorphism confers a high risk of onset for recurrent pregnancy loss. The M2 haplotype shows lower promoter activity and expresses lower level of *ANXA5* mRNA. This promoter region could form guanine-quadruplex (G4) structure, and the M2 allele possesses less potential than the major allele in vitro. In addition, the promoter activity correlated with the *in vitro* G4 propensity estimated by circular dichroism. In our present study, we assessed *in vivo* G4 structure of the *ANXA5* promoter region using bisulfite method. The genomic DNA from the placenta was treated with sodium bisulfite without dissociation of the DNA. Under this condition, the nucleotide Cs could be changed to U only at the single-stranded regions, but not at the double-stranded regions. After the conversion, we amplified this region by PCR, cloned the corresponding region and determined the conversion rate of the Cs. The promoter region showed about 24% conversion of Cs, while other regions resulted in 8%. When the conversion rate was calculated in each strand, the complementary strand of the promoter region showed the higher conversion rate. This phenomenon could be explained that the strand forming G4 structure is resistant about bisulfite attack, while the complementary strand is left to be single-stranded that resulted in higher conversion rate. The lower conversion rate was observed in the DNA with M2 haplotype than that of major haplotype. These results suggest that the G4 structure at the promoter drives expression of *ANXA5*, and the M2 haplotype reduces the G4 propensity and the *ANXA5* expression resulting in hypercoagulation within the placental villi leading to the recurrent pregnancy loss.

499M

An epigenetically regulated expression-variable class of genes depleted in neurodevelopmental CNVs. *A. Gimelbrant[1,2], L.A. Weiss[3,4].* 1) Cancer Biology, Dana-Farber Cancer Institute, Boston, MA., United States; 2) Genetics, Harvard Medical School, Boston, MA, United States; 3) Psychiatry, University of California San Francisco, San Francisco, CA, United States; 4) Institute for Human Genetics, University of California San Francisco, San Francisco, CA, United States.

We recently discovered that >3,000 human autosomal genes can be expressed from a single allele in one cell, and from the other allele -or both- in a neighboring cell. This epigenetic phenomenon is known as variable monoallelic expression (MAE). MAE is a mitotically stable process generating a somatic mosaic of MAE and BAE cells in a given tissue, with mRNA expression levels higher in BAE cells compared with MAE cells for the same gene. Little is known about functional consequences of MAE. We hypothesize that MAE increases expression level variability in the population, with potential impact on phenotypic variation.

In this study, we used a chromatin signature to identify MAE genes in lymphoblastoid cell lines and in human fetal brain tissue. We assessed available data to compare expression variation in MAE genes to BAE genes at three distinct scales. At the trans-species level, we evaluated gene expression differences between humans and non-human primates. In the human population, we used genetic mapping of expression variability data to identify genes with polymorphisms influencing expression variance. Finally, we provide an analysis of pathogenic and non-pathogenic neurodevelopmental copy number variant (CNV) data in humans.

We find that MAE genes show higher interspecies variance in expression level at the mRNA and protein level and are more likely to have acquired differences in expression during evolution compared with constitutive BAE genes ($P_{RNA}<3\times10^{-8}$; $P_{PROTEIN}<4\times10^{-5}$). We show that MAE genes are more likely to have evQTLs (expression variability quantitative trait loci) than BAE genes ($P<10^{-23}$). Finally, we observe that fetal brain MAE genes are specifically depleted in pathogenic neurodevelopmental CNVs but not polymorphic CNVs ($P<2\times10^{-4}$).

Our results suggest that genes undergoing MAE appear to be a special set of epigenetically-regulated genes subject to both more extreme species differences and more extreme individual differences in expression level. Depletion of these genes in pathogenic neurodevelopmental CNVs suggests that neurodevelopmental processes subject to genomic haploinsufficiency are also less tolerant to epigenetically-mediated expression variation by MAE.

500T

Defining the role of CGGBP1 protein in FMR1 gene expression. *M. Goracci, S. Lanni, F. Palumbo, G. Mancano, P. Chiurazzi, E. Tabolacci, G. Neri.* Institute of Medical Genetics, Catholic University, Rome, Italy.

Fragile X syndrome (FXS) is the most common heritable form of cognitive impairment and is caused by the expansion over 200 repeats and subsequent methylation of the CGG triplets at the 5' UTR of the FMR1 gene, leading to gene silencing. The epigenetic and molecular mechanisms responsible for FMR1 gene silencing are still unclear. To outline structure-specific proteins that could recruit components of the silencing machinery we investigated the role of CGGBP1 in FMR1 gene transcription. CGGBP1 is a highly conserved protein which binds specifically unmethylated CGG tracts. The role of CGGBP1 on FMR1 transcription is yet to be defined. Sequencing analysis and expression studies through quantitative PCR of CGGBP1 were performed in cell lines with different allele expansions (wild-type WT, premutation, methylated full mutations FXS and unmethylated full mutation UFM), demonstrating no differences between them. ChIP assays showed that CGGBP1 binds unmethylated CGG triplets of the FMR1 gene proportionally to the length of the repeats. We also observed that CGGBP1 binding to the FMR1 locus was restored after pharmacological demethylation with 5-azadC of FXS alleles, suggesting a possible role for CGGBP1 in FMR1 expression. CGGBP1 silencing with siRNA (reaching ~ 85% of CGGBP1-mRNA depletion) did not affect FMR1 transcription in WT and UFM fibroblasts. Although the strong binding to the CGG tract could suggest a relevant role of CGGBP1 on FMR1 gene expression, our results demonstrate that CGGBP1 is not a direct regulator of FMR1 transcription. Supported by Telethon Onlus, FRAXA Foundation and Italian Association for fragile X syndrome.

501M

Epigenomic Analysis for Single Cells and Small Quantities of Cells. *X. Pan[1], J. Yang[1], L. Han[2], N. Liu[1], H. Zhu[1], Y. Tanaka[1], Y. Li[1], X. Zi[1,2], H. Wu[3], M. Riester[3], J. Xiong[1], J. Wu[1], G. Euskirchen[4], W. Li[1], F. Rong[2], F. Michor[3], M. Snyder[4], S. Weissman[1].* 1) Genetics, Yale School of Medicine, New Haven, CT; 2) Yale School of Engineering & Applied Science, New Haven, CT; 3) Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard University, Boston, MA; 4) Department of Genetics, Stanford University, Stanford, CA.

Functional genomics analysis in single cells provides unique power for the analysis of heterogeneous biological systems and is applicable for the study of mechanism of differentiation, reprogramming and development as well as clinics. In recent years single cell analysis (SCA) has achieved great progresses in the analysis of genomic sequences and transcriptomes, but study of epigenetic marks has lagged. Previously we have established methods for whole DNA pool amplification (WPA) and sequencing(1), full length mRNA-seq (SMA and PMA)(2), robust measurement of the integrate length of the telomeres in single cells (SCT-pqPCR)(3), and analysis of DNA and RNA from the same single cells. Representative results from our lab include unexpected levels of heterogeneity in hematopoietic early precursor cells and complexity in the early response of normal hematopoietic cells to exposure to erythropoietin, the demonstration of increasing telomere heterogeneity in aging cell cultures, and telomere expansion in the first few divisions of fertilized eggs. Recently we have worked out several new and sensitive methods for CpG methylation profiling of a low number of cells or single cells. An ultrasensitive RRBS at consistently reached >50% coverage for single cells comparing to conventional RRBS; two methods for CpG methylation pattern profiling (CpGMp) for single cells gives a much higher coverage of DMRs (differentiatially methylated regions) and CpG islands without bisulfite conversion. In addition, we have also developed a new concept and method for genome-wide extraction of large blocks of highly compacted chromatins (HCC) by sequencing, from as few as 100-cells. This compacted DNA shows positive or negative correlations with histone modifications H3K27me3, H3K27ac, H3K4me2, and H3K9ac, CTCF, and enhancers. Interestingly, HCC are inversely correlated to DHS in general, but a large set of blocks of DNA are neither HCC nor DHS (DNase I hypersensitive sites). Thus, HCC provides a new parameter to sensitively measure chromatin structure. Ref: (1). Pan X, et al. (2008) A procedure for highly specific, hypersensitive and unbiased whole genome amplification. PNAS, 105(40):15499-504. (2). Pan X, et al. (2013) Two methods for full-length RNA sequencing for low quantities of cells and single cells. PNAS. 110(2):594-9. (3). Wang F, Pan X, et al. (2013) Robust measurement of telomere length in single cells. PNAS. 110 (21): E1906-E1912. xinghua.pan@yale.edu.

502T

RNA-Seq revealed canonical RNA editing in mouse retina with laser-induced choroidal neovascularization. *J. Chen[1], S. Huang[1], J. Liang[1], Z. Lu[1], C. Pang[2], M. Zhang[1], H. Chen[1].* 1) Joint Shantou International Eye Center, Shantou University/The Chinese University of Hong Kong, Shantou, Guangdong, China; 2) Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong Special Administration Region, China.

Purpose: Laser-induced choroidal neovascularization (CNV) in mice is an important animal model to study exudative age-related macular degeneration, but molecular changes in the retina in CNV mice remain largely unclear. In this study RNA-Seq was used to study RNA-editing in the retina of mouse CNV model.

Methods: Laser photocoagulation of Bruch's membrane was used to induce CNV in the right eyes (OD) of three adult C57B/L6 mice. Total RNA was extracted from the whole retina of CNV (OD) and untreated contralateral (OS) eyes on Day 3. RNA-Seq was performed on the Illumina HiSeq 2000 platform with configuration of 20 million 100 bp pair-end reads per sample. Reads were aligned to the mouse full genome (UCSC version mm10) using Bowtie and splice junctions were identified by using TopHat2. Quantification of expression level and analysis of differential expression of RNA-Seq data were done using R language package EdgeR. An FPKM (fragments per kilobase of exon per million fragments mapped) filtering cutoff of 1.0 in at least three of the six samples was used to filter genes with low expression. Canonical RNA editing sites with editing level ≥ 1% were identified using Samtools and VarScan. Analysis of gene ontology (GO) enrichment and KEGG pathway was performed using DAVID bioinformatics resources. Gene network was constructed using GeneMANIA.

Results: In total 33 potential canonical RNA-editing sites were detected only in the retina with CNV but not in the contralateral untreated retina. Among these sites, the highest level of RNA editing was found at the 3'-UTR of Tapbp (19.60 ± 11.43%). Six protein-coding genes with RNA editing including Tapbp were also found to be differentially expressed in the retina with CNV compared to the contralateral retina. GO analysis showed that gene network of the protein coding genes with RNA editing was mainly involved in immune response.

Conclusions: In the current study, RNA-Seq analysis revealed remarkable RNA-editing in the retina of CNV mouse model. RNA editing could play a substantial role in retina inflammation during CNV.

503M

The influence of Microvesicles carrying microRNA on cross talk between FOXP3 and EZH2 in multiple sclerosis. *A. Hossein-nezhad[1,2,3], Z. Maghbooli[3], S.M. Eshaghi[3], S. Emamgholipour[3], MA. Sahraian[3].* 1) University of Miami, Miami, FL; 2) Boston University, Boston, MA; 3) Tehran University of Medical Sciences, Tehran.

Transcription factor FoxP3 is found initially as a key regulator in regulatory T cells. We have previously reported an inverse coloration between FoxP3 and inflammatory cytokine in different subtype of multiple sclerosis (MS) that were involved in the MS severity. Also we showed the effect of microvesicle on changing the specific gene expression profiles. Relatinship between FoxP3 and Ezh2, an enzyme for histone 3 lysine 27 methyl transferase, reported in cancer. In this report, we present evidence showing that circulating microvesicle of healthy subject regulates cross talk between Ezh2 and Foxp3 in peripheral blood mononuclear cells (PBMC) of multiple sclerosis patients, which is critical for MS monitoring and treatment. Fifteen primary progressive subtype of MS patients and 15 healthy age and sex-matched controls were included in the study. All patients gave their informed consent and the study was approved by the local ethics committee. PBMCs were isolated by Ficoll-Isopaque density gradient centrifugation. The gene expression profiling was determined with Real time PCR. The serum levels of TNF-α, IL-1β, and RANKL were measured by ELISA and hs-CRP levels was determined by imunoturbidimetric method. Microvesicles isolated from healthy controls sera underwent an ultracentrifuge. Gene's expression was evaluated before and after microvesicle induction of PBMC. We have also induced PBMC with RNase-treated microvesicles. MS patients had higher levels of TNF-α, IL-1β, CRP, RANKL and EZH2 and lower levels of foxp3 than healthy controls. RANKL and TNF-α showed a reverse significant correlation with FoxP3 relative expression levels. We have also observed the inverse correlation between FoxP3 and Ezh2 in MS patients. The expression of Ezh2 in PBMC of MS patients decreased after induction by microvesicles, which extracted from healthy controls and also expression of FoxP3 increased. RNase-treated microvesicles had no significant effect on gene expression profiles. This study shows an important part for epigenetic changes during cross talk between EZH2 and FoxP3 in MS. It highlights the role of RNA content of microvesicles from healthy controls to suppress Ezh2 in primary progressive multiple sclerosis. This study represents an opportunity for novel therapeutic agents in multiple sclerosis based on microRNA content of microvesicles. In addition, cross talk between Ezh2 and FoxP3 may important in monitoring and targeting therapy in multiple sclerosis.

504T

Analysis of miRNAs in MELAS mutant cells. *R. Li[1,2], D. Wang[3], N. Dasgupta[1].* 1) Human Genetics, Cincinnati Children's Hosp, Cincinnati, OH. Division of Human Genetics, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229; 2) Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45229; 3) Maxillofacial Pathology and Radiology Department, Ohio State University College of dentistry. 304 W. 12th Avenue, Columbus, OHIO 43210.

MELAS (mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes) is caused by A3243G mutation in mtDNA. The pathogenic mechanism of MELAS is not completely understood. microRNAs (miRNAs) regulate gene expression via the targeting of mRNA for degradation or by suppression of mRNA translation through complementarity to target mRNAs. The A3243G mtDNA mutation increases mitochondrial membrane permeability, may lead to the increased import of miRNAs into mitochondria. We hypothesize that MELAS-associated A3243G mtDNA mutation results in the accumulation of specific miRNAs in mitochondria, targeting on mitochondrial RNAs, causing mitochondrial dysfunction and contributing to the pathogenesis of MELAS. To study whether miRNAs is relevant to MELAS, cell pellets and pure mitochondria were obtained from A3243G homoplasmic mutant 43B cybrids and wild type HSI cybrids as well as normal human fibroblast (HF), mutant fibroblasts hf16 and hf68 with 90% and 30% of A3243G mutation respectively. Pure mitochondria were then treated by RNase A to eliminate cytoplasmic RNAs. Total RNA preparations were obtained from cell pellets of 43B, HSI, HF, hf16 and hf68 and their RNase A treated mitochondrial pellets. Next-gen sequencing of small RNA (15–49 bp) including miRNAs (18–23 bp) were carried out using cellular RNA and RNase A treated mitochondrial RNA samples. The small RNA sequence results demonstrated that some miRNAs were enriched in both 43B and HSI mitochondrial RNA. 40 miRNAs in HSI mitochondria and 7 miRNAs in 43B mitochondria were richer than in their cellular miRNAs. Most important, several miRNAs were enriched in MELAS-associated mitochondria compared with wild type mitochondria. RT-qPCR results confirmed these miRNAs were richer in 43B mitochondria than in HSI mitochondria. These miRNAs were also increased in hf16 fibroblast mitochondria. However, these miRNAs were unchanged in their cellular. These findings suggested the MELAS mutation, mtDNA A3243G, resulted in an increased import miR-181c and miR-423-5p from cytoplasm into mitochondria. We identify 7 miRNAs, particularly miR-181c and miR-423-5p, are enriched in mitochondria of MELAS cybirds and patient fibroblasts with high percentage of mtDNA A3243G mutation. These mitochondrial miRNAs provide attractive candidates for study of the pathophysiology of mitochondrial diseases and potentially establishing a miRNA-based therapeutic strategy for individuals with incurable mitochondrial disease.

505M

Genome-wide microRNA expression profiling in fetal trisomy 21 placenta. *J.H. Lim[1], S.Y. Park[1], J.W. Kim[1], D.E. Lee[1], S.Y. Kim[1], H.J. Kim[1], K.S. Kim[1], H.K. Ahn[2], S.W. Lee[2], H.M. Ryu[1,2].* 1) Laboratory of Medical Genetics, Medical Research Institute, Cheil General Hospital and Women's Healthcare Center, Seoul, Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women's Healthcare Center, Kwandong University College of Medicine, Seoul, Korea.

Trisomy 21 (T21) is the most common known aneuploidy, caused by an extra copy of all or part of human chromosome 21 (hsa21). Differential expression of microRNAs (miRNAs) has been identified in many diseases, including T21. However, the changes of the genome-wide microRNA expression in the T21 placenta have yet to be determined. Moreover, the biological functions of the differentially expressed miRNAs in the T21 placenta are still unclear. Thus, we investigated the genome-wide miRNA expression profiling of the T21 placenta using microarray and analyzed the biological functions of the differentially expressed miRNAs in the T21 placenta using bioinformatics tools. Total RNA extracted from placenta samples of T21 and euploid fetus was used for miRNA microarray. To investigate the functions of the differentially expressed miRNAs (fold change >2.0 and P<0.05), the miRWalk database was used to predict their potential targets, which were afterwards submitted to the functional annotation tool provided by the WebGestalt database. The interaction map of potential target genes was generated using STRING 9.1. In total, 34 miRNAs were significantly differentially expressed in the T21 placenta, compared with the euploid placenta (16 up-regulated and 18 down-regulated). Target genes of up-regulated and down-regulated miRNAs were identified 7,434 and 6,071, respectively. Out of total target genes, 76 genes were identified within hsa21 (10 and 34 genes being controlled by the down- and up-regulated miRNAs, respectively, and 32 by both). Target genes of abnormally expressed miRNAs were significantly associated with in T21 as well as T21-related disorder such as mental retardation, chromosome disorders, and so on. Moreover, the genes were involved in various mechanisms such as metabolic pathways, natural killer cell mediated cytotoxicity, toll-like receptor signaling pathway, and so on. Interaction network of target genes showed the regulatory mechanism of hsa21 genes targeted by miRNAs in the T21 placenta. To the best of our knowledge, this is the first genome-wide study to comprehensively survey placental miRNAs of T21 fetuses. Our results provide a new insight into understanding the expression characteristic of microRNAs in the T21 fetal placenta. Additionally, findings of our study indicate that the differentially expressed miRNAs in the T21 placenta may potentially affect various pathways related with pathogenesis of the T21.

506T

Regulation of *COL1A1* by miRNAs in scleral fibroblasts: Implications for myopia control. *R. Metlapally, K.K. Wang, C.F. Wildsoet.* School of Optometry, UC Berkeley, Berkeley, CA.

Purpose: Previous work has implicated scleral (outer white coat of the eye) miRNAs in ocular growth regulation during normal and myopia (near-sightedness) development. Scleral thinning and extracellular matrix (ECM) loss are classic features of high/pathological myopia caused by excessive axial elongation of the eye. miRNAs represent potential therapeutic targets for myopia control by promoting/enhancing ECM deposition in the sclera. Since *COL1A1*, the most prevalent protein in the scleral ECM, undergoes degradation and loss in myopia and miRNAs are known to regulate *COL1A1*, we investigated the potential of altering *COL1A1* gene expression levels via selective manipulation of miRNA activity in scleral fibroblast cultures. **Methods:** Human scleral fibroblasts from donor eyes were grown in culture using standard methods. miRNA inhibitors directed at let-7c, let-7e, miR-98 or miR-103 were introduced into the cells via electroporation. Cells were lysed after 48 hours, followed by RNA extraction and reverse transcription using the Taqman® Cells to CT kits for studying miRNA (let-7c, let-7e, miR-98 and miR-103) and mRNA (*COL1A1*) expression. Four biological replicates were assessed for each target and data normalized to the negative scramble controls. QPCRs were performed using Taqman® assays. *GAPDH*, *POLR2A*, *RNU44* and *RNU6B* served as housekeeping genes, and the delta CT method was used for data analyses. A live/dead viability/cytotoxicity assay (Life Technologies) was used for assessing toxicity of the treatments. **Results:** All four miRNAs tested were inhibited as expected with their respective inhibitors (at least 40-fold, min p=0.001). The expression of *COL1A1* was down-regulated after inhibition of either let-7c, let-7e or miR-98 (at least 2-fold, min p=5.9×10^-5), and up-regulated after inhibition of miR-103 (1.5 fold, p=0.007). The cytotoxicity assay findings did not reveal any adverse effects for any of the miRNA inhibitor treatments. **Conclusions:** Our findings show that scleral miRNAs can be manipulated in culture to regulate gene expression. The finding that scleral *COL1A1* mRNA expression is modulated by miRNAs holds promise for future *in vivo* therapeutic intervention studies. Our long-term goal is to control axial ocular elongation in myopia by inhibiting the underlying scleral changes.

507M

Human neural stem dell in the form of neurosphere: microRNA expression profile. *M. Palacios-Reyes[1,2], A.M. Espinosa-García[3], A. Contreras[4], R.M. Ordoñez[2], G. Benitez-King[5], A. Hidalgo[6], I. Rubio[1], I. Palma[1,7].* 1) Laboratorio de Morfología Celular y Molecular, Sección de Estudios de Posgrado e Investigación, Escuela Superior de Medicina, Instituto Politécnico Nacional, Plan de San Luis y Díaz Miron, SN, Col. Casco de Santo Tomás, México, 11340, D.F., México; 2) Unidad de Investigación Médica en Genética Humana, Hospital de Pediatría, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Av. Cuauhtémoc # 330 Col. Doctores. 06720 México, D.F., México; 3) Unidad de Medicina Genómica, Hospital General de México, Eje 2A Sur (Dr. Balmis) 148 Doctores, Cuauhtémoc, 06726 México, D.F., México; 4) Departamento de Biología del Desarrollo y Teratogénesis. Hospital Infantil de México Federico Gómez. Doctor Márquez 162 Doctores, Cuauhtémoc, 06720 México, D.F., México; 5) Departamento de Neurofarmacologa, Instituto Nacional de Psiquiatría 'Ramón de la Fuente Muñiz', Calzada México Xochimilco 101 San Lorenzo Huipulco, Tlalpan, 14370 México, D.F., México; 6) Laboratorio de Genómica del Cáncer, Instituto Nacional de Medicina Genómica, Periférico Sur 4809 Arenal Tepepan, Tlalpan, 14610, México, D.F., México; 7) Departamento de Morfología, Facultad de Medicina Veterinaria y Zootecnia (FMVZ); Universidad Nacional Autónoma de México (UNAM), Ciudad Universitaria No. 3000, 04510, México D.F., México.

Olfactory neuroepithelium (ONE) represents a site where neural stem cells can be recovered (NSC), besides being a readily accessible site. They can be cultured by neurosphere assay, forming spherical cluster with multipotent and progenitor cells in suspension, with ability to differentiate. These features make it a model for studying molecular and cellular neural processes. Maintaining these cells in an undifferentiated state and their ability to differentiate, requires the availability to changes cell expression, therefore to have epigenetic mechanisms related to modulate molecules that promote the multipotentiality and that suppress cells differentiation. One such mechanism depends on microRNAs, small molecule non-coding RNAs with roles in regulation of expression. miRNAs roles are relevant in brain diferentiation and function. miRNA dysregulation is present in brain disorders. Human NSCs are a powerful tool to investigate neural process, to evaluate miRNA expression in this cell model, could show new information about miRNAs roles. To this end, a nasal swab from ONE was taken from 4 healthy volunteers, Cells were cultured like neurospheres and their neural phenotype was determined by by immunofluorescence. Subsequently, miRNA expression assay was performed on plates TLDA A y B of applied V 2.0. Cq values equal or greater than 36 were removed. We assessed the expression level according to DCt method, normalizing with the global mean and by groups according to the mean plus / minus standard deviations. Finally, we performed an in silico analysis looking for possible nervous system pathways in which some microRNAs are involved. We analyzed a total of 667 microRNAs. 246 microRNAs were expressed in at least 3 of the 4 samples. According to the expression level, 56 miRNAs have very high expression (23%), 23 high (9%), 66 medium (27%), 34 low (14%) and 67 very low (28%). The 10 microRNAs with higher expression were miR-222, miR-200c, miR-191, miR-30a *, miR-30e *, miR-484, miR-146a, miR-378, miR-24 miR-574-3p. Some miRNAs are already described with roles in neurogenesis as miR-184, miR-132 family members miR-200c, miR-30, among others. As for his possible role, predictions point to possible involvement in neurodevelopmental pathways as well as in neural pathologies.

508T

Epigenetic factors regulating DUX4 expression in muscle cells. *J. Balog[1], P.E. Thijssen[1], K.R. Strassheijm[1], Y.D. Krom[1], A. de Jong[1], R.J.L. Lemmers[1], P.J. van der Vliet[1], R. Tawil[2], S.J. Tapscott[3], S.M. van der Maarel[1].* 1) Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Neuromuscular Disease Unit, Department of Neurology, University of Rochester Medical Center, Rochester, New York, United States of America; 3) Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America.

Facioscapulohumeral dystrophy (FSHD) is a progressive, dominantly inherited myopathy affecting 1:14.000–20.000 individuals. A key factor in the pathomechanism of FSHD is DUX4, a double-homeobox germline transcription factor which misexpression causes muscle cell death. Sporadic expression of DUX4 in myonuclei is preceded by chromatin relaxation of the D4Z4 macrosatellite repeat array that contains a copy of the DUX4 gene in every single D4Z4 unit. These epigenetic changes in D4Z4 are a consequence of either a contraction of the D4Z4 repeat array in >95% of cases (FSHD1), or of mutations in SMCHD1 (Structural Maintenance of Chromosomes Flexible Hinge Domain Containing) in most other cases (FSHD2). We aimed to determine the epigenetic requirement of DUX4 transcription in primary human myoblasts and myotubes. First we established the dynamics of DUX4 expression during myogenesis in a series of control (n=6), FSHD1 (n=4) and FSHD2 (n=6) human primary myoblast and myotube samples. DUX4 transcripts were detected in FSHD1 and FSHD2 myoblasts and a robust increase of DUX4 transcript levels was observed in FSHD myotubes. To understand the epigenetic changes underlying transcriptional upregulation of DUX4 we quantitatively measured DNA methylation and activating histone modifications (H3K4me2, H3K4me3 and H3K36me3) or repressive histone modifications (H3K27me3 and H3K9me3) at D4Z4 in the same samples DNA methylation levels were lower in patient myoblasts and myotubes, but no significant change was found during myogenesis suggesting that increased levels of DUX4 in myotubes do not correlate with changes in DNA methylation. Activating histone modifications did correlate with transcription, with higher levels in patient samples compared to controls. We could not confirm the reported loss of H3K9me3 at D4Z4 in patient myotube samples. Furthermore downregulation of known chromatin modifiers involved in H3K9me3 mediated repression at D4Z4 such as SUV39H1 or cohesin subunits SMC3 and RAD21 did not induce DUX4 transcription in control myoblasts and myotubes while depletion of SMCHD1 resulted in transcriptional derepression of DUX4. Interestingly the level of H3K27me3 and PRC2 subunits EZH2 and SUZ12 were increased in FSHD myoblasts but subsequently decreased during myogenic differentiation possibly contributing to the leaky expression of DUX4. We conclude that SMCHD1 is a central chromatin regulator at D4Z4 controlling DUX4 expression in FSHD1 and FSHD2.

509M

Integrative analysis reveals enhanced regulatory effects of human large intergenic noncoding RNAs in lung adenocarcinoma. *X. Kong[1], Y. Zhou[1], T. Huang[2], J. Jiang[3], J. Huang[1], P. Zhang[1], Y. Zhu[1], Y. Shi[1], L. Hu[1].* 1) Molec Gen, Inst Hlth Sci, Shanghai, China; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America; 3) State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China.

Although there is an accumulating appreciation of the key roles that large intergenic noncoding RNAs (lincRNAs) play in diverse cellular processes, our knowledge of how lincRNAs function in cancer remains sparse. Here, we present a comprehensive landscape of RNA-seq transcriptome profiles of lung adenocarcinomas and their paired normal counterparts from 72 South Korean patients to unravel the mysteries of lincRNA functions and mechanisms. We found asymmetric expression patterns of lincRNAs between cancer and adjacent paired normal tissues, with an increased breadth and quantity of lincRNAs during cancer progression. Consistent with previous findings, lincRNAs were typically coexpressed with their neighboring genes, albeit to an extent similar to that of pairs of neighboring protein-coding genes. By building a mathematical model based on correlated gene expression, we distinguished an additional subset of lincRNAs termed "regulatory lincRNAs," representing their dominant roles in gene regulation. The number of regulatory lincRNAs was significantly higher in cancerous compared to normal tissues, with greater enrichment of target protein-coding genes associated with cell growth and differentiation processes observed in cancer tissues. Moreover, dozens of lincRNAs switched from non-regulators in normal tissues to regulators in tumor tissues. Our integrated analysis reveals enhanced regulatory effects of lincRNAs in lung adenocarcinoma and provides a resource for the study of regulatory lincRNAs that facilitate tumorigenesis and may serve as new targets for clinical therapy.

510T

A Neuroepigenomic Model of the Fetal Alcohol Exposure Spectrum.

B.I. Laufer, E.J. Diehl, M.L. Kleiber, A. Chokroborty-Hoque, B. Alberry, K. Mantha, S.M. Singh. Molecular Genetics Unit, Department of Biology, Western University, London, Ontario, Canada.

Maternal alcohol consumption during pregnancy causes a continuum of heterogeneous disorders termed Fetal Alcohol Spectrum Disorders (FASD). Patients affected with FASD show life long defects, particularly affecting the central nervous system and its complex traits. Our group has developed an animal model using C57BL/6J (B6) mice and genome-wide molecular technologies. The results show that exposure of alcohol during neurodevelopment in B6 mice causes behavioral disabilities matching FASD patients in resulting offspring. Further, the resulting mice show changes in brain gene expression as well as epigenetic marks, including DNA methylation, multiple histone modifications, and ncRNA expression. Interestingly, the genes affected by the epigenetic marks last a lifetime. Also, the genes affected participate in the critical neural processes of apoptosis, neurodevelopment, cellular identity, cell-cell interaction, and signalling. Using genome-wide arrays to interrogate the cytosine methylation and ncRNA expression within the whole brain and hippocampus of adult mice prenatally exposed to alcohol, we have identified an epigenomic footprint that is characterized by alterations to imprinted regions of the genomes, controlled by DNA methylation, that encode multiple developmentally important non-coding RNAs (ncRNAs) and are regulated by CCCTC-binding factor (CTCF), a zinc finger protein. These processes are developmentally integral to the outcome of larger scale cortical brain structure formation, since the events of both pre- and post-natal neurodevelopment are highly dependent on the (epi)genotype as well as the experience and environment of the differentiating cells at the molecular level. Taken together, our models suggest that ethanol can create significant long-term changes to the molecular mechanisms that create and maintain an epigenetic landscape that is essential to normal brain function and future neurodevelopment. The proposed model suggests a potential for effective attenuation of disease endophenotypes, given the plastic nature of the epigenome in response to enriched postnatal environments.

511S

Allele-aware ChIP-seq alignments identify allelic differences in transcription factor binding at disease-associated loci. *M.L. Buchkovich[1], K.E. Eklund[1], Q. Duan[1], L. Y[1,2,3], K.L. M[1], T.S. Furey[1,4].* 1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 3) Department of Computer Science, University of North Carolina, Chapel Hill, NC; 4) Department of Biology, University of North Carolina, Chapel Hill, NC.

Allelic differences in protein-DNA binding can influence gene transcription at disease-associated loci. These differences can be detected as allelic imbalance, or enrichment of one allele, in ChIP-seq data. Allele-aware alignments that use genotyping data during mapping more accurately align sequence reads to heterozygous sites than traditional, genotype-independent alignments, and are important for detecting imbalances. We used an allele-aware aligner, GSNAP, and heterozygous sites from complete (sequencing-based), partial (Human1MDuo array + imputation), or no sample-specific genotype data to map CREB1 ChIP-seq reads from GM12878, a human lymphoblastoid cell line. In the case of no genotype data, common variants (1000 Genomes; EUR MAF>.05) were considered heterozygous sites. Both CREB1, a transcription-activating protein, and B-lymphocytes are known to play a role in inflammation. Using complete genotype data and requiring a minimum of 5 sequences containing each allele, we detected 204 sites of allelic imbalance (binomial p<.01). We correctly identified 161 and 146 of these imbalances using partial and no genotypes, respectively, plus an additional 52 and 47 imbalances not detected using complete genotypes. Two sites of allelic imbalance, rs2382818 and rs713875, are in linkage disequilibrium (1000 Genomes EUR; r2>=.8) with inflammatory bowel disease (IBD)-associated variants, and are predicted to be strong enhancers (chromHMM) in GM12878 cells. While the imbalance at rs713875 was detected in the genotype-independent and all three allele-aware alignments, the imbalance at rs2382818 was only detected in allele-aware alignments using complete or partial genotypes. We performed electrophoretic mobility shift assays for these two variants with purified CREB1 protein and observed allelic differences in CREB1 binding of both variants in the directions consistent with the predicted imbalances. Together, these data suggest that differential binding of CREB1 to alleles of these variants may influence IBD by altering the transcription of one or more nearby genes. These findings highlight our ability to create allele-aware alignments in the absence of completely genotyped samples and the utility of these alignments in identifying sites with allelic differences in protein binding at disease-associated loci.

512M

Functional genomics approaches reveal novel role of a key hematopoietic transcription factor in blood disorders. *M. Byrska-Bishop[1], D. VanDorn[2], AE. Campbell[2], PR. Arca[2], Y. Yao[2], CA. Keller[1], BM. Giardine[1], P. Gadue[3], FF. Costa[4], RL. Nemiroff[5], GA. Blobel[2], DL. French[3], RC. Hardison[1], MJ. Weiss[2], ST. Chou[2].* 1) The Pennsylvania State University, Center for Comparative Genomics and Bioinformatics, Department of Biochemistry and Molecular Biology, University Park, PA; 2) The Children's Hospital of Philadelphia, Division of Hematology, Philadelphia, PA; 3) The Children's Hospital of Philadelphia, Department of Pathology and Laboratory Medicine, Philadelphia, PA; 4) University of Campinas, Department of Internal Medicine, School of Medicine, Campinas, Brazil; 5) The University of Pennsylvania School of Medicine, Department of Obstetrics and Gynecology, Philadelphia, PA.

GATA1 (MIM *305371) is a hematopoietic transcription factor important for the development of erythrocytes and megakaryocytes. Normally, a human hematopoietic cell expresses a full-length form of GATA1 and a short form, GATA1s, that lacks the N-terminal transactivation domain (NAD). Germline *GATA1* mutations that result in exclusive production of GATA1s (referred to as "*GATA1s* mutations") were identified in patients with congenital hypoplastic anemia and neutropenia (MIM #300835), as well as Diamond Blackfan anemia (MIM #300835). In individuals with Down syndrome (MIM #190685), somatic *GATA1s* mutations lead to transient myeloproliferative disease (TMD [MIM #190685]) and acute megakaryoblastic leukemia (AMKL [MIM #190685]). We investigated the function of the human GATA1s by creating and analyzing induced pluripotent stem cells (iPSCs) from somatic tissues of patients with Down syndrome-associated TMD, as well as congenital macrocytic anemia due to a germline *GATA1s* mutation. Microarray transcriptome profiles of iPSC-derived progenitors revealed that genes upregulated by GATA1s were enriched for megakaryocytic and myeloid genes, while the downregulated genes were enriched for erythroid genes. Furthermore, compared to controls, erythropoiesis by *GATA1s* iPSCs was markedly reduced, while megakaryopoiesis and myelopoiesis were enhanced. Single cell expression profiling of 91 selected hematopoietic genes revealed small, but significant differences in gene expression distributions for 40 genes between *wtGATA1* and *GATA1s* cells. This suggests that small changes in expression of many genes contribute to the GATA1s-mediated change in lineage bias from erythroid to myelo-megakaryocytic lineage fate. To provide a mechanistic explanation of the observed phenotype, we performed ChIP-seq experiments to study differences in binding patterns between full length GATA1 and GATA1s. ChIP-seq results indicate that GATA1s binding is specifically impaired at erythroid target genes, but normal at megakaryocytic genes, implicating the N-terminus as a chromatin occupancy factor selective for erythroid cells.

513T

Identifying rare, non-coding DNA variants in Systemic Lupus Erythematosus. *S.J. White[1], S. Cantsilieris[1], F.J. Rossello[2], D. Belluoccio[3], E.F. Morand[4].* 1) MIMR-PHI Institute of Medical Research, Monash University, Clayton, Victoria, Australia; 2) Victorian Bioinformatics Consortium, Monash University, Clayton, Victoria, Australia; 3) Agilent Technologies, Mulgrave, Victoria, Australia; 4) Southern Clinical School, Monash University, Clayton, Victoria, Australia.

Systemic lupus erythematosus (SLE) is a debilitating multisystem autoimmune disease. Patients with SLE suffer a marked loss of life expectancy, and severe morbidity, due to autoimmune-mediated inflammation of multiple organs. A classic hallmark of SLE is the presence of autoantibodies. This is a clear biological link between SLE and antibody-producing B-lymphocytes (B-cells), strengthened by the clinical efficacy of anti-B cell therapies in certain patients. Genetic studies have associated SLE susceptibility with a number of Single Nucleotide Polymorphisms (SNPs), many of which are found in or near genes with B-cell functions. However, the known SLE SNPs together only account for a small percentage of the genetic contribution to SLE, and as most SNPs are located in non-coding DNA, the mechanism of their effect is often unclear. Several SLE-associated SNPs are located in such DNA elements that control gene regulation, with sequence changes disrupting the binding of specific transcription factors. In 2012 the ENCODE consortium published papers outlining genome-wide studies of open chromatin in a range of cell types. These studies demonstrated that DNaseI hypersensitive sites (DHS) strongly correlate with known and predicted regulatory sequences. In addition, they are enriched for known and predicted transcription factor binding sites (TFBS). This provides a clear link between sequence variants and functional activity, and rare variants in B-cell regulatory loci (defined by open chromatin in B-cells) would explain genetic susceptibility to SLE in a way not detectable by GWAS. We have designed a custom capture approach using an Agilent SureSelectXT2 Custom Library to screen 49K B-cell DHS in genomic DNA from SLE cases. This covers ~ 17Mb of genomic sequence, with up to 16 samples being pooled in a single lane on the Illumina HiSeq 1500. We are currently analysing the sequence data from a cohort of SLE cases, and plan to perform functional validation using reporter constructs and changes in DNaseI-sensitivity.

† Trainee Award Finalist

514S

Coordinated Regulatory Variation Associated with Maternal Glycemic Traits Regulates HKDC1 Expression. *T.E. Reddy[1,2], C. Guo[1], A.E. Ludvik[3], M.G. Hayes[3], L.L. Armstrong[3], D.M. Scholtens[4], C.D. Brown[5], B.T. Layden[3,6], W.L. Lowe[3].* 1) Institute for Genome Science & Policy, Duke University Medical School, Durham, NC; 2) Department of Biostatistics & Bioinformatics, Duke University Medical School, Durham NC; 3) Division of Endocrinology, Metabolism and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 4) Department of Preventive Medicine, Division of Biostatistics, Northwestern University Feinberg School of Medicine, Chicago, IL; 5) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 6) Jesse Brown Veterans Affairs Medical Center, Chicago, IL.

Genome wide association studies (GWAS) to understand the genetics of complex human traits strongly suggest that non-coding genetic variation is a major contributor to disease risk. One hypothesis is that variation in gene regulation may contribute to altered expression of disease-relevant genes. As one example, in our previous GWAS, we identified SNPs in 10q22 near the putative hexokinase gene HKDC1 that were associated with two hour glucose levels in an oral glucose tolerance test administered at ~28 weeks gestation during pregnancy. None of the associated SNPs were within the protein-coding regions of HKDC1, suggesting that the SNPs instead could have a regulatory effect on HKDC1. To address that hypothesis, we have comprehensively mapped regulatory variation in the 10q22 region associated with maternal glucose levels. We used open chromatin data from the ENCODE project and the Roadmap Epigenome Consortium to define regulatory elements, and measured regulatory function of each element across four different human ancestry groups using reporter assays constructed from DNA from the 1,000 Genomes Project. To disrupt linkage within regulatory elements and to better assay rare and personal variants, we complemented the set of natural haplotypes using site-directed mutagenesis. With this approach, we saturated the common and rare variation in regulatory elements across this gestational hyperglycemia locus. These assays revealed a surprising amount of regulatory variation controlling expression of HKDC1. The effects of regulatory variants were strongly and significantly coordinated across risk-haplotypes. Furthermore, when cloned into a single reporter construct, multiple regulatory elements acted independently. These results suggest that regulatory variants with a coordinated effect across a large haplotype contribute to a reduction in HKDC1 expression that is greater than would be expected for a single causal regulatory variant. We confirmed the reporter assay results using an eQTL analysis that revealed a genetic influence on HKDC1 expression in the same locus and in the expected direction at genome-wide significance. Together, these results provide evidence for a novel model of human disease in which coordinated regulatory effects across a risk haplotype have a magnified effect on endogenous gene expression, leading to subtle but significant effects on phenotype.

515M

HyCCAPP for functional proteomic analysis of promoter variants associated with plasma lipid levels and gene expression in Mexican Americans. *H. Guillen Ahlers[1], D. Perumalla[1], P. Rao[1], A. Jadhav[1], J.M. Proffitt[1], M.P. Johnson[1], J.E. Curran[1], H.H. Göring[1], P.J. Meikle[3], J. Blangero[1], M.R. Shortreed[2], L.M. Smith[2], M. Olivier[1].* 1) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) Department of Chemistry, University of Madison, Madison, WI; 3) Metabolomics Laboratory, Baker IDI Heart and Diabetes Institute, Melbourne, Australia.

Lipid and lipoprotein levels influence risk of cardiovascular disease; however, studies have traditionally focused on the assessment of plasma cholesterol and triglyceride levels. The plasma lipidome consists of many thousands of lipid metabolites, with large structural diversity. As part of the San Antonio Family Heart Study, lipidomic profiles (quantification of 319 lipid species) were generated from 1,240 Mexican American individuals from 30 extended families using electrospray ionization-tandem mass spectrometry. Individual lipid measures were correlated with gene expression profiles for 20,413 genes from PBMCs, and promoter variants (cis eQTL) were identified that were strongly associated with both expression and individual lipid profiles. Three variants were further characterized. Variants, rs8736 within 1000 bp of the TSS 5' of the TMC4 gene and rs887316 in the proximal promoter of the RAD9A gene, were associated with a phosphatidylinositol lipid levels ($p = 1.0\times10^{-5}$ and 4.2×10^{-7}). The rs606458 variant in the SF1 gene promoter was associated with levels of a phosphatidylcholine levels ($p = 1.7\times10^{-5}$). We tested the two alleles of each variant in luciferase reporter assays introducing promoter-containing plasmids in K562 cells. All variants resulted in differential activity in the reporter assay, suggesting altered gene expression for each allele. However, ChIP-seq data from ENCODE did not reveal the likely DNA-binding protein mediating the observed effect. Therefore, we adapted our novel methodology, Hybridization Capture of Chromatin-Associated Proteins for Proteomics (HyCCAPP) to isolate the desired promoter regions from crosslinked cells using hybridization capture with complementary oligonucleotide probes, as initially developed for the analysis of genomic regions in yeast. The methodology was developed to isolate individual chromatin regions after protein-DNA crosslinking, and use mass spectrometry to characterize and quantify the bound proteins. In our adaptation to the luciferase reporter system, the targeted promoter regions were captured with over 3% efficiency, and the isolated material was amenable to proteomic analysis. HyCCAPP now allows the effective characterization of DNA-binding proteins for promoter variants that show differential activity in luciferase reporter assays, and will help elucidate the underlying mechanisms by which the variants identified in our cohort modulate gene expression and plasma lipid levels.

516T

Single cell allele specific expression (ASE) in Down syndrome and common aneuploidies. *G. Stamoulis[1], P.G. Fereirra[1], P. Makrythanasis[1,2], F. Santoni[1], M. Guipponi[1,2], M. Garieri[1], O. Delaneau[1], E. Falconnet[1], P. Ribaux[1], E.T. Dermitzakis[1,3], C. Borel[1], S.E. Antonarakis[1,2,3].* 1) University of Geneva, Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva 4, Switzerland; 2) Geneva University Hospitals-HUG, Service of Genetic Medicine, 1211 Geneva 4, Switzerland; 3) iGE3 Institute of Genetics and Genomics of Geneva, University of Geneva, 1211 Geneva 4, Switzerland.

Trisomy 21 is a model disorder of altered gene expression. Several studies have addressed the transcriptome differences between normal and affected individuals; however all of these studies suffer from the presence of ''noise'' due to gene expression variation among different individuals. We have previously used a pair of monozygotic twins discordant for trisomy 21 in order to study the global dysregulation of gene expression (Nature: 508; 7496,2014). The majority of previous studies focused on aneuploidies were conducted on cultured cell populations or tissues, but studies focusing on gene and allelic expression behaviour at the single cell level are lacking. In this study we explore the allele specific expression in Trisomy 21 using transcriptome studies in single cells. We have used 40 normal cells and 48 trisomic cells from the fibroblasts of the monozygotic twins discordant for trisomy 21 and compared the ASE (allele specific expression), and their transcriptional metrics in these two cell groups. Remarkably we have observed extensive monoallelic expression of genes in all autosomes including on chromosome 21. In addition a series of samples from mosaic trisomy 21, trisomy 13 and trisomy 18 are in different stages of investigation. These studies in single cells will provide a fundamental understanding of the gene expression dysregulation and allele specific expression in aneuploidies and may contribute to the understanding of the phenotypic heterogeneity of these syndromes.

517S

Analysis of long-range interactions in primary human cells reveals *CFTR* new regulatory elements. *S. Moisan[1,2,3], S. Berlivet[4,5], J. Dostie[5], C. Férec[1,2,3].* 1) INSERM U1078 Laboratoire de Génétique, Génomique Fonctionnelle et Biotechnologies, Brest, France; 2) Université de Bretagne Occidentale, Faculté de Médecine et des Sciences de la Santé, Brest, France; 3) CHU Brest, Hôpital Morvan, Laboratoire de Génétique Moléculaire, Brest, France; 4) Unité de Génétique et Développement, Institut de Recherches Cliniques de Montréal, Québec, Canada; 5)478 Departement of Biochemistry and Goodman Cancer Research Center, McGill University, Montréal, Québec, Canada.

The cystic fibrosis transmembrane conductance regulator (*CFTR*) gene was identified in 1989. Twenty five years later, the regulatory mechanisms controlling its complex expression are still not fully understood. Although, 1970 mutations have been identified, many cases of cystic fibrosis or CFTR Related Disorders remain still of unknown origin. The promoter which binds transcription factors and drives some aspects of *CFTR* gene expression, cannot alone account for tissue specific control. This implicates other distal cis- or trans-acting elements in cell-type-specific regulation of *CFTR* expression. The aim of our project is to study long-range regulatory mechanisms of the *CFTR* gene. Interestingly, the majority of the human genome is composed of non coding DNA whose function has not been thoroughly investigated. A significant number of conserved non-coding sequences (CNCs) are found in gene-poor regions, these large intergenic regions must have kept a function throughout evolution. We first developped the Chromosome Conformation Captures (3C) approach to map these potentials regulatory elements which could interact specifically with the *CFTR* gene by tridimensional folding mechanism. Subsequently, we enhanced our analyses with a high-throughput adaptation of 3C: the 3C-Carbon Copy (5C) technology. This approach allows the analysis of millions chromatin interactions. Thus, we have analyzed the spatial organization of a ~790kb region, comprising the *CFTR* gene, with 145 5C primers. Interactions between this locus and the *CFTR* promoter have been analysed by next generation sequencing with the Ion PGM™, in primary epithelial cells, which express the gene and primary skin fibroblasts, which do not express the gene. We compare chromatin conformation in order to identify uncharacterized regulatory elements that act especially in CFTR-expressing cells. Our approach is validated by the identification of previously characterized regulatory elements. Moreover, we identify novel chromatin contacts of the *CFTR* promoter with chromosomal regions, which could potentially be involved in *CFTR* gene expression regulation. Thanks to 3C and 3C-derivated analyses, we could identify new possible mutations far from the gene, which may lead to its dysfunction by modifying the chromatin conformation. These analyses will be pursued on patients affected by cystic fibrosis or by CFTR Related Disorders, in whom either a single mutation or none was found in the *CFTR* gene coding sequence.

518M

Assessing the functional significance of disease-associated cis-regulatory variants in vivo using a versatile dual colour transgenesis strategy in zebrafish. *S. Bhatia, DA. Kleinjan, VV. Heyningen.* MRC Human Genetics Unit, University of Edinburgh, Edinburgh, United Kingdom.

Disruption of gene regulation is a significant cause of human diseases. A class of noncoding region of the genome called cis-regulatory elements (CREs) or enhancers function as the primary determinants of precise spatial and temporal regulation of their target genes during development by serving as docking sites for tissue-specific transcription factors. Sequence variation in CREs disrupt transcription factor binding sites causing mis-regulation of their target genes thereby leading to a plethora of human diseases. The importance of these regulatory elements for human health is further highlighted by the finding that >70% of single nucleotide polymorphisms (SNPs), identified in genome-wide association studies (GWAS) to be associated with common and complex human diseases, are located in noncoding regions without any currently known function, but which are likely to be CREs. While technological advances in sequencing methods have led to the identification of a large number of putative disease-associated CRE variants, robust methods for discerning the functional CRE variants from background sequence variation are largely lacking. We have developed an efficient approach for the in-depth characterisation of putative regulatory variants using dual-colour reporter transgenics in zebrafish, allowing detailed in vivo comparison of spatial and temporal differences in regulatory activity between variants in the same animal. The method also allows for simultaneous assessment of multiple separate elements for subtle differences in spatio-temporal overlap, and the validation of putative binding factors by the effect of morpholino depletion on CRE activity. We validate the method on a number of known human disease-associated elements from the SHH, PAX6 and IRF6 locus, and use the approach to characterise a set of novel long-range SOX9 enhancers implicated in Pierre Robin Sequence. Our results provide compelling evidence for establishment of a robust in vivo method for functionally characterising disease associated CRE variants. The data reveals novel cis-regulatory elements controlling SOX9 expression during craniofacial development and simultaneously highlight how the functional activity of these elements can be affected by a variety of mutant enhancer alleles in patients with Pierre Robin Sequence.

519T

Systemic Lupus Erythematosus-associated functional variants influence the gene expression of UBE2L3 through the regulation of promoter and enhancer activities. *S. Wang[1], B. He[2], P. Gaffney[1,2].* 1) Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma city, OK; 2) Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Genetic variants in the UBE2L3 region have been associated with multiple autoimmune diseases including Systemic Lupus Erythematosus (SLE), Crohn's disease, celiac disease, and rheumatoid arthritis. We previously reported an SLE-associated functional risk haplotype in the UBE2L3 region that led to significant increases in both mRNA expression of UBE2L3 and protein expression of UBCH7. However, the mechanism by which the risk haplotype influences UBE2L3 expression has not yet been described. In this report, we employed a combination of approaches including EMSA, Luciferase reporter assay, Western blotting, and RT-qPCR to evaluate the functional potential of SLE-associated variants in the UBE2L3 region and to characterize the mechanism by which these functional variants influence UBE2L3 gene expression. Through the use of EMSA, we observed functional evidence for two variants in and around the UBE2L3 promoter region. A functional variant (rs140490), located150 bps upstream of the UBE2L3 transcription start site increases nuclear protein binding (P=0.0031). Luciferase reporter assays demonstrated that the risk allele of this functional variant increases UBE2L3 promoter activity (P=0.0038) in EBV-transformed B cells. In contrast, another functional variant (rs2266959), located in the first intron of UBE2L3, reduces nuclear protein binding (P=0.0014). Luciferase reporter activity assays demonstrated that rs2266959 falls within an enhancer element, the risk allele of which leads to a significant increase in the enhancer activity (P=0.0118). These results are consistent with our previous finding in which the risk haplotype carrying these functional variants is associated with increased UBE2L3 gene expression. In conclusion, we identified two functional variants, rs140490 and rs2266959, carried on the SLE-associated risk haplotype of UBE2L3. The variant rs140490 increases binding of nuclear proteins to the promoter, while the variant rs2266959 reduces binding of nuclear proteins to the enhancer. However, both functional variants lead to significantly increased transcriptional activity of the promoter and the enhancer of UBE2L3. Our data suggest that these functional variants may be responsible for the gene expression phenotype produced by the risk haplotype of UBE2L3. Further mechanistic characterization of the functional variants of UBE2L3 is required to understand the role of this risk haplotype in autoimmune disease pathogenesis.

520S

Comparative genomics and abstraction of DNA sequences allow the identification of cis-regulatory signatures. *P.EM. Guimaraes[1], M.A.A. Almeida[2], T.J.P. Sobreira[3], H.A. Castillo[4], J.X. Neto[4], P.S.L. Oliveira[3].* 1) Laboratório de Evolução de Genes e Genomas, INGEB, UFU, Patos de Minas, MG, Brazil; 2) AT&T GCC, Department of Genetics, Texas Biomedical Research Institute, San Antonio TX USA; 3) Laboratório de Bioinformática, LNBio, CNPEM, Campinas, SP, Brazil; 4) Laboratório de Modificação do Genoma, LNBio, CNPEM, Campinas, SP, Brazil.

Gene expression fine regulation is achieved by the complex interaction between transcription factors (TF) and their respective binding sites (TFBS). Identify these short DNA motifs is a major current goal in modern molecular biology. However, the efficiency of current prediction algorithms is limited due to the sequence variation between TFBS and the complexity of transcription regulatory networks. As an endeavor to modify this scenario, here we present and validate a new computational routine to improve the prediction of functional TFBS. This pipeline uses a "phylogenetic footprint" filter and converts predicted TFBS to an alphabet, which describes evolutionary-functionally related families of TF. The "phylogenetic footprint" works as a selection step for orthologous sequences. Further, TFBS were predicted in syntenic blocks, for each species evaluated, and converted into strings of families of functional related transcription factors. The family clustering reduces TFBS variability and prediction redundancy. Finally, only TFBS families predicted at the same gene relative position and present in the set of selected species were considered. We carried our routine in order to identify functional cis-regulatory elements acting in *aldh1a2* gene, searching for syntenic sequences in a 30 kb window flanking the coding sequence of the mouse gene, including all intronic sequences, comprising a total of 138,191 bases. Orthologous segments were identified comparing the mouse to human, rat and chicken genomes. A total of 1,318 predicted TFBS families were considered constrained by our "spatial and evolutionary filter", reducing in 99.6% the original universe of 382,750 predicted elements. Despite the reduction of predicted TFBS, it could be accomplished of one undesired side effect: the lost of functional binding sites. To evaluate the efficiency of our routine we compared the final set of predicted TFBS to those previously experimentally identified as real regulators of *aldh1a2* gene. All validated TFBS were represented in our set, indicating that even with the considerable reduction of predicted TFBS we still clamped the real ones. Taken together, our data suggest that the routine can be applied to high-throughput analyses and help to draw more focused experimental validation protocols. The same strategy can be applied to identification of miRNA seed pairing, splicing enhancer elements and others regulatory signatures and structural elements.

521M

Detecting gene-by-environment interactions using allele specific expression. *D.A. Knowles[1], S.B. Montgomery[1], A. Battle[2].* 1) Stanford University, Stanford, CA; 2) Johns Hopkins University, Baltimore, MD.

The impact of environment and lifestyle on human health is dramatic, with major risk factors including substance use, pollution, diet and exercise. However, the interaction between environment and individual genetic background is poorly understood; detecting gene-by-environment interactions is both statistically and computationally challenging. Using RNA sequencing of primary tissue (whole blood) from 922 individuals in the DGN cohort, combined with extensive annotation of environmental factors including drug use and behavioral factors, we evaluate GxE effects at a cellular level. Even in this study, standard association methods detect only a handful of gene-by-environment interactions at a lenient FDR of 0.1 across 30 environment variables. RNA-seq offers an alternative strategy however: we look for genes whose allelic expression is associated with an environment variable of interest. Measuring two different alleles within a single individual and time point offers a highly controlled comparison of environmental modulation of genetic effects, not subject to the many biological and technical variables that differ between samples, and thus provides additional power and specificity to identify causal factors. To test for such associations, we have developed a hierarchical Bayesian model of allele-specific read counts that takes into account local cis-regulation, read depth, and technical influences on over-dispersion of RNA-seq data. This methodology significantly increases power by leveraging the controlled, within individual, nature of allelic expression and by integrating over potential causal cis-variants. Applying this method across 30 environment variables, we find dozens of GxE interactions at FDR 0.1, a substantial increase over the standard approach. Pathway analysis of genes responding to each environmental factor further uncovers relevant shared function, including enrichment in a platelet degranulation pathway for use of ACE inhibitors, the PYK2 pathway for exercise, and the Toll-like receptor signaling pathway for time of day. Individually significant genes include SESN3 for exercise, which regulates blood glucose levels, IL10RA for smoking, and CDH23 for decongestant medication, which is expressed in the neurosensory epithelium. In conclusion, we show it is possible to leverage the novel information provided by RNA-seq beyond total expression to unravel the influence of disease-associated environmental factors on gene regulation.

522T

An investigation of the correlation between cis-genetic and environmental factors regulating gene expression between cell types, exposures, and populations. *N. Zaitlen[1], J. Yang[2], A. Bakshi[2], A.L. Price[3,4], A. Regev[4,5,6], C.J. Ye[4].* 1) Medicine, University of California San Francisco, San Francisco, CA; 2) University of Queensland, Brisbane, Australia; 3) Harvard School of Public Health, Boston, MA; 4) Broad Institute, Cambridge, MA; 5) Massachusetts Institute of Technology, Cambridge, MA; 6) Howard Hughes Medical Institute.

RNA is the first intermediary between genetic variants and their downstream effects on human disease. Recent work has shown that SNPs altering gene expression as opposed to coding SNPs may be the major drivers of complex phenotypic variation. Therefore, characterization of how regulatory SNPs (eQTLs) change between tissues, environments, and populations is central to understanding of the genetic basis of human disease. In this work we leverage a unique data set of expression from 19,360 genes collected in CD4+ T cells and CD14+ monocytes, from 183 Europeans (Eur), 91 African Americans (AfAm), and 74 Asians (Asn) genotyped at 700K SNPs. We also profiled 415 genes in CD14+/CD16- dendritic cells after exposure to IFNβ, influenza, and LPS. We estimate components of heritability of genotyped SNPs within and between contexts averaged across all genes; determining cis-genetic heritability ($h^2_{g,cis}$) in each context as well as the correlation of cis-genetic and environmental effects betweens cell types ($\rho_{g,cel}$,$\rho_{e,cel}$) and exposures ($\rho_{g,exp}$,$\rho_{e,exp}$). Finally, we develop new theory to estimate a lower bound of cross-population cis-genetic correlation of gene expression for each pair of populations ($\rho_{g,pop}$). We estimate the $h^2_{g,cis}$ in (Eur,AfAm,Asn) at (0.039,0.041,0.031) for CD4s and (0.048,0.053, 0.045) for CD14s (s.e. < 0.001), observing higher $h^2_{g,cis}$ for CD14s than CD4s and an ordering of $h^2_{g,cis}$ in according to population genetic diversity (AfAm>Eur>Asn). The cross cell type genetic correlations $\rho_{g,cel}$=(0.353, 0.251,0.244), implying different sets of eQTLs between cell types. Upon exposure to LPS, flu and IFNβ, $h^2_{g,cis}$ estimates for Eur CD14 increased by (34.4%, 44.5%, 30.0%), with similar patterns in AfAm and Asn. In contrast to $\rho_{g,cel}$, the cross exposure correlations were dramatically higher with $\rho_{g,exp}$= (0.93,0.542,0.779). The higher value of $\rho_{g,exp}$ than $\rho_{g,cel}$, coupled with the increases in $h^2_{g,cis}$ are consistent with a model where upon exposure, the same causal variants regulate expression with larger effect sizes. In CD14 across populations we observe $\rho_{g,pop}$(Eur,AfAm)=0.56, $\rho_{g,pop}$(Eur,Asn)= 0.52, and $\rho_{g,pop}$(AfAm,Asn)=0.40. The cross population genetic correlations are higher than cross tissue and lower than cross exposure correlations and decrease as a function of population genetic distance. In all, our results suggest cross tissue analysis will uncover distinct regulatory regions while stimulation will increase the power to detect eQTLs.

523S

Studies on the function of HCMV-encoded essential gene UL49. *H. LI, D. YANG, L. FENG, Y. RAN, H. LIU, T. ZHOU, Z. DENG.* Dept. Biotechnology, Jinan University, Guangzhou, Guangdong, China.

Human cytomegalovirus (HCMV), is a clinically important and ubiquitous herpesvirus, and affects individuals whose immune functions are compromised or immature. This virus is a leading cause of retinitis-associated blindness and other debilitating conditions such as pneumonia and enteritis among AIDS patients. Moreover, HCMV causes mental and behavioral dysfunctions in children that were infected in utero. HCMV- encoded UL49 gene is essential for viral growth and might serve as a potential target for combat HCMV infection. RNase P complexed with external guide sequence (EGS) represents a novel nucleic-acid-based gene interference approach to modulate gene expression and is used in this study to explore a potential therapeutic approach for inhibiting UL49 gene expressing and replication of HCMV. A functional EGS RNA was constructed to target the UL49 mRNA . The EGS RNA was shown to be able to direct human RNase P to cleave the target mRNA sequence efficiently in vitro. A reduction of 68 % in the mRNA and protein expression levels and a reduction of 1000-fold in viral growth were observed in human cells that expressed the functional EGS. Further studies of the function of UL49 gene indicate that the reduction of UL49 gene expression has no effect on HCMV genome replication and gene expression of IE2 and gB. However, it leads to the down-regulation on the gene expression UL99, which is highly immunogenic and play roles in the assembly and egress of virus particles. This observation has provided us a valuable new clue to understand the role of UL49 gene during HCMV lytic infections.

524M

Deep genetic, transcriptomic and epigenetic maps of human blood elements. *L. Chen[1,2], S. Busche[3], A. Datta[4], O. Delaneau[5], K. Downes[2], S. Ecker[6], H. Kerstens[7], M. Kostadima[4,2], D.S. Paul[8], OBO. BLUEPRINT CONSORTIUM[9].* 1) Human Genetics, Wellcome Trust Sanger Institute, Cambridge, Cambridge, United Kingdom; 2) Department of Haematology, University of Cambridge, NHS Blood and Transplant, Cambridge Biomedical Campus, Cambridge, United Kingdom; 3) McGill Epigenome Mapping Centre, McGill University, Montreal, Quebec, Canada; 4) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom; 5) Département de Génétique et Développement (GEDEV) 1, Michel-Servet, CH-1211 Genève 4; 6) Structural and Computational Biology Group, Spanish National Cancer Research Center (CNIO), Madrid, Spain; 7) Department of Molecular Biology, Radboud Institute for Molecular Life Sciences, Radboud University, Nijmegen, The Netherlands; 8) UCL Cancer Institute, University College London, London, United Kingdom; 9) http://www.blueprint-epigenome.eu/.

Haematopoiesis, the formation of blood elements, has been studied as a paradigm of stem cell biology and development. The BLUEPRINT project (http://www.blueprint-epigenome.eu/) is generating deep genetic, transcriptomic and epigenetic maps of over 30 human blood cell types and their malignant counterparts that will be available to the scientific community. A first stream of research has focused on cataloguing the transcriptional and alternative splicing changes in rare blood progenitor cells. We performed high-throughput RNA transcriptome analysis of eight highly-purified cell populations (HSC, hematopoietic stem cells; MPP, multipotent progenitor cell; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte monocyte progenitor; MEP, megakaryocyte erythrocyte progenitor;EB, erythroblasts and MK, megakaryocytes). We identified (1) 7,881 novel splice junctions that were specific to these progenitor cell populations; (2) 20,924 high-confident alternative-splicing junctions, where approximately 11% were cell-type-specific and included categories of genes such as RNA splicing and transcription factors; (3) two differentially spliced transcription factors that were important in megakaryopoiesis by using in vitro experiments. These results suggest that alternative splicing is a critical post-transcriptional regulatory mechanism in hematopoietic cell differentiation. A second stream of research has focused on the effect of common sequence variation on the epigenetic landscape using detailed and integrated analyses. In three peripheral blood cell types, monocytes, neutrophils and naïve CD4 T cells, we have generated data for whole genome sequencing, RNA-seq, ChIP-seq of two histone modifications and DNA methylation by 450K arrays in an initial set of 48 healthy individuals with European-ancestry. Here we describe the first epigenetic quantitative trait loci (QTL) analyses of this integrated data set to further understand the effects of genotype-phenotype correlations via epigenetic mechanisms in these immune cells.

525T

Analysis of monoallelic expression in human individual cells revealed novel imprinting genes. *F. Santoni, C. Borel, M. Garieri, E. Falconnet, P. Ribaux, S.E. Antonarakis.* Department of Medical Genetics and Development, University of Geneva, Geneva, Geneva, Switzerland.

Genomic imprinting in mammals results in the expression of one allele of a given gene depending on its maternal or paternal inheritance. There has been considerable interest in identifying imprinted genes since they have been associated both with rare syndromes, and with common diseases including diabetes and cancer. So far, less than 100 human imprinted genes have been identified; however, the total numbered of them has been estimated to be a few hundreds. The aim of our study is to identify novel imprinting genes using an innovative approach based on single-cell RNA sequencing. We captured and RNA sequenced 203 individual fibroblasts from 2 different individuals. Additionally, the DNA of these individuals has been whole-genome sequenced at 8x depth. In total, more than 400'000 heterozygous single nucleotide variants (hetSNVs in the coding genes) were identified for further allelic specific expression analysis using an in-house pipeline. We selected all the genes with at least one cell with a coverage superior or equal to 16 reads at the hetSNV position and showing a monoallelic ratio (ref/total) of 0-0.1 or 0.9-1 for all the analyzed single cells. We were able to detect 1007 putative imprinted genes for the first individual and 1926 for the second individual. By retaining only the genes in common, we were able to identify 198 putative imprinted genes in fibroblast cells. This included novel and 8 known imprinted genes. We only detected and validated the monoallelic expression of 42 genes with bulk cell samples. Our single-cell approach allowed the detection of low expressed imprinted genes that are undetectable in standard RNAseq protocol. We confirmed their status of imprinting genes by investigating their parent of origin-dependent transmission of allelic biases using sequenced data from independent parent-offspring fibroblast trios and from Caucasian and Yoruban parent-offspring trios as part of the 1000 genome project. We also documented their tissue specificity by analyzing their monoallelic expression in datasets from multiple tissues. C.B. anf F.S. contributed equally.

526S

Somatic Rhodopsin transcriptional repression by artificial DNA-binding proteins targeted to cis-regulatory elements. *S. Botta[1], E. Marrocco[1], N. de Prisco[1], F. Curion[1], M. Sofia[1], M.L. Bacci[3], S. Rossi[2], F. Simonelli[2], E.M. Surace[1,4].* 1) Tigem (Telethon Institute of Genetics and medicine), Napoli, Italy;; 2) Dipartimento di Oftalmologia, Seconda Università degli Studi di Napoli, Napoli, Italy;; 3) Dipartimento di Scienze Mediche Veterinarie, Università di Bologna, Bologna, Italy; 4) Dept. of Translational Medicine, "Federico II" University, Napoli, Italy.

Artificial DNA-binding proteins are becoming a novel and versatile system to development therapeutics. The ZF protein (ZFPs) modular structure enables both the sequential assembling of multiple ZFs to generate a DNA-binding domain (DBD) with different target specificities and the use of various effector domains including transcriptional regulation domain or nucleases for either genomic regulation or editing. We showed that a selected artificial transcriptional repressor (Zinc-Finger 6 repressor, ZF6R) delivered by AAV2/8 vector down regulates selectively mutated human RHO expression levels, ameliorating the retinal phenotype in a mouse model of autosomal dominant Retinitis Pigmentosa (adRP). We speculate that a protein, which contains exclusively the DBD (ZF6-DBD) domain without the effector domain (KRAB), may confer a higher therapeutic efficacy and tolerability genome-wide (off target) than ZF6-KRAB. Therefore, we removed from the construct the KRAB domain. The adRP mouse model eyes treated with AAV8-CMV-ZF6-DBD demonstrate robust and consistent recovery of the ERG a-wave and b-wave responses along a wide range of luminance in both scotopic and photopic conditions when compared to EGFP treated eyes. In addition, ZF6-DBD treated eyes show significant higher responses compared to AAV8-CMV-ZF6-KRAB. Based on sequence identity between the human and the porcine ZF6-DBD target site, we next assess the functional ability of the construct in the porcine retina. To assess "pure" transcriptional effects and not possible secondary degeneration owed to rhodopsin knock-down, we performed an early sacrifice. Fifteen days after vector administration the retina regions transduced were harvested and RT-PCR studies were performed. Transcriptional levels of Rho showed that retinas that received ZF6-DBD demonstrate more than 40%; of transcriptional repression of Rho. Furthermore, next generation sequencing (NGS; RNA-seq) was used to assess the differential transcriptional profile of AAV8-CMV-ZF6 and AAV8-CMV-ZF6-KRAB treated eyes compared to controls. The dataset generated show specific down-regulation of Rho expression. Ongoing experiments are showing remarkably similar results with Transcription activator-like effector (TALE) technology. In summary, the data support the use of artificial DNA-binding proteins as therapeutics.

527M

Epigenetic promoter silencing in Friedreich ataxia is dependent on the length of the GAA triplet-repeat mutation. *Y. Chutake[1], C. Lam[1], W. Costello[1], M. Anderson[3], S. Bidichandani[1, 2].* 1) Department of Pediatrics, Univ Oklahoma HSC, Oklahoma City, OK; 2) Department of Biochemistry, Univ Oklahoma HSC, Oklahoma City, OK; 3) Department of Biostatistics & Epidemiology, Univ Oklahoma HSC, Oklahoma City, OK.

Friedreich ataxia (FRDA) is an autosomal recessive disease caused by an expanded GAA triplet-repeat (GAA-TR) mutation in intron 1 of the *FXN* gene. Most patients are homozygous for expanded alleles containing 100–1700 triplets, which results in a severe deficiency of *FXN* transcript. Transcriptional deficiency and phenotypic severity are significantly correlated with the length of the shorter of the two expanded alleles. We show that repressive chromatin extends from the expanded GAA-TR in intron 1 to the upstream regions of the *FXN* gene, involving the *FXN* transcriptional start site (*FXN*-TSS). Using a chromatin accessibility assay and a high-resolution nucleosome occupancy assay, we found that the *FXN*-TSS, which is normally in a nucleosome-depleted region is rendered inaccessible by altered nucleosome positioning in FRDA. Consistent with the altered epigenetic landscape, the *FXN* gene promoter was found to be in a transcriptionally non-permissive state in FRDA. Both metabolic labeling of nascent transcripts and an unbiased whole transcriptome analysis revealed a severe deficiency of transcriptional initiation in FRDA. *FXN* transcriptional deficiency, measured by qRT-PCR in patient cell lines bearing a range of expanded alleles (200–1122), was significantly correlated with the length of the shorter of the two expanded alleles. Importantly, this was noted both upstream (R^2=0.84; p=0.014) and downstream (R^2=0.89; p=0.002) of the expanded GAA-TR mutation, suggesting that *FXN* promoter silencing in FRDA is related to repeat length. A bilinear regression model revealed that length-dependence was strongest when the shorter of the two expanded alleles contained <400 triplets. Direct measurement of *FXN* promoter function in patients with expanded alleles containing <400 versus >400 triplets in the shorter of the two expanded alleles revealed a significantly greater deficiency in individuals with longer GAA-TR alleles. Thus, deficient transcriptional initiation caused by epigenetic silencing of the *FXN* promoter is a major cause of *FXN* transcriptional deficiency in FRDA, and it is dependent on the length of the expanded GAA-TR mutation.

528T

Characterizing the predictive features of regulatory genetic variants. *N.M. Ioannidis[1], N. Larson[2], Y. Zhang[2], A. Battle[1], S. Montgomery[1], S.N. Thibodeau[2], W. Sieh[1], A.S. Whittemore[1], C.D. Bustamante[1].* 1) Stanford University School of Medicine, Stanford, CA; 2) Mayo Clinic, Rochester, MN.

Whole-genome and whole-exome sequencing technologies are increasingly enabling studies of genetic variation in large numbers of healthy and diseased individuals; however, interpreting the clinical significance of the many genetic variants identified in these studies remains a critical challenge. One mechanism by which genetic variants can achieve clinical significance—even those that do not affect protein structure, such as noncoding, intronic, and synonymous variants—is by regulating gene expression. Here we use a machine learning approach to identify genomic features predictive of genetic variants that are associated with gene expression regulation. We test a large set of features including sequence conservation, overlapping functional elements from ENCODE and Ensembl, and position relative to the nearest transcription start site and splice site. We identify predictive features by training random forest models on known cis-expression quantitative trait loci (cis-eQTLs) in lymphoblastoid cell lines (LCLs). These models achieve an area under the receiver operating characteristic curve (AUC) of between 0.75 and 0.9 depending on the stringency with which cis-eQTLs are defined. We also characterize the tissue-specificity of the genomic features of regulatory genetic variants by comparing the predictive performance of our random forest on cis-eQTLs in LCLs versus cis-eQTLs found in normal prostate tissue from an independent sample of around 500 prostate cancer patients. We anticipate that our results will be useful in future studies of regulatory variation and for prioritizing the likely clinical significance of rare and novel genetic variants identified in large-scale clinical sequencing studies.

529S

Characterization of enhancer gene interactions using DNaseI and gene expression data across 110 cell types. *P. Kheradpour[1,2], M. Kellis[1,2], Roadmap Epigenomics Consortium.* 1) Massachusetts Institute of Technology, Cambridge, MA; 2) Broad Institute, Cambridge, MA.

Recent efforts to characterize diseases through genome-wide association studies and annotate the genome using ChIP-seq experiments have led to a dramatic increase in putative functional genomic regions. While most of the implicated loci have fallen outside coding regions and are thought to be regulatory in nature, efforts to link these regions to their target genes, thereby permitting a better understanding of their importance, has lagged considerably. Generally, experimental linking techniques only permit the interrogation of a small number of specific regions or produce a genome-wide linking at very low resolution.

We utilize DNaseI hypersensitivity and expression data from 110 cell types produced by the ENCODE and Roadmap Epigenomics projects to produce linking confidences between hypersensitive regions and nearby genes. We find that high confidence links are supported by independent datasets such as eQTL annotations. Moreover, we find that after controlling for distance several factors, including the orientation of the upstream gene and the presence of an interleaving CTCF, greatly affect the linking confidence.

In addition to predicting links between genes and nearby putative enhancers, we are interested in answering fundamental biological questions such as the number of enhancers per gene and where these enhancers fall relative to a gene's transcription start site. We address these questions through a careful analysis of the distribution of confidence scores for each gene and by performing a scaling analysis on the number of cell types.

530M

Allele-specific gene expression analysis of RNA-seq data in 471 healthy prostate tissue samples identifies extensive aseQTLs in the human prostate transcriptome. *N.B. Larson, S. McDonnell, A.J. French, Z. Fogarty, J. Cheville, S. Middha, S. Riska, S. Baheti, A.A. Nair, L. Wang, J. Cunningham, D. Schaid, S.N. Thibodeau.* Mayo Clinic, Rochester, MN.

Imbalance in allele-specific expression (ASE) of homologous copies of a gene may result from heterozygous SNP cis-acting eQTLs. Identifying these regulatory variants provides insight into the dynamics of the transcriptome and may highlight genetic variation relevant to complex traits. RNA sequencing is a powerful tool for quantifying ASE by directly capturing mRNA transcript sequence information. It is important to characterize such regulatory variation in a tissue-specific manner, as most publically available eQTL results are from lymphoblastoid cells and tissue-specific regulatory variation may go undetected. We recently conducted a genome-wide allele-specific eQTL (aseQTL) study on 471 samples of normal prostate tissue (from patients with prostate/bladder cancer) by generating gene expression (RNA-seq) and chromosomally-phased genotype (Illumina Infinium 2.5M) datasets. We additionally imputed common (MAF >1%) exonic SNPs using 1000 Genomes data to increase the proportion of ASE-informative reads, resulting in ~2.7% of all aligned reads being allele-specific. From ~1.6M QC-passed SNPs and 16,607 expressed genes, we evaluated 7,435,266 aseQTL (SNP × gene) associations, restricting our analyses to common autosomal SNPs within 500kb of gene transcription start/stop sites. We tested aseQTL associations by jointly modeling the total and ASE gene read counts, and inconsistent results (e.g., trans-acting or imprinting effects) were excluded. We identified 393,046 significant aseQTLs (FDR = 0.05), with 10,870 genes corresponding to ≥1 putative aseQTL. From the set of most significant aseQTLs per gene, 172 SNPs corresponded to multiple genes. Functional annotation for these SNPs was enriched for ENCODE enhancer histone marks, suggesting evidence of cis co-regulation. The effect size estimates were inversely correlated with MAF (Spearman ρ = -0.38), indicating purifying selection pressure on high-impact regulatory variation. Large-effect aseQTLs were highly concentrated near the TSSs of the regulated genes and tended to co-localize with ENCODE promoter histone marks. Our findings also highlight dynamic expression in genes preferentially expressed in prostate tissue. Of the 102 prostate-specific genes listed in the TiGER tissue-specific gene expression database within our expressed gene set, 81 were significantly associated with aseQTLs. Overall, our results demonstrate widespread and complex cis-acting functional variation in the prostate transcriptome.

531T

Uncovering SMCHD1 regulation of the Protocadherin Cluster. *A.G. Mason[1], J. Balog[1], R.J.L.F. Lemmers[1], R.C. Slieker[2], B.T. Heijmans[2,3], R. Tawil[4], S.J. Tapscott[5], S.M. van der Maarel[1].* 1) Human Genetics, Leiden University Medical Center, Leiden , Netherlands; 2) Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; 3) Netherlands Consortium for Healthy Aging, PO Box 9600, Leiden 2300, RC, The Netherlands; 4) Neuromuscular Disease Unit, Department of Neurology, University of Rochester Medical Center, Rochester, NY, USA; 5) Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

Facioscapulohumeral muscular dystrophy (FSHD) is a rare disease affecting 1;15,0000–20,000 individuals and is clinically characterized by progressive weakness and wasting of the facial and upper extremity muscles. In FSHD individuals, due to the contraction of the D4Z4 macrosatellite repeat array on chromosome 4 (FSHD1), or due to *SMCHD1* mutations (FSHD2), the somatic repression of the retrogene *DUX4* is incomplete leading to the variegated expression of the germline transcription factor in muscle fibers. *SMCHD1* encodes a protein that regulates chromatin repression at different loci, and it has been shown to be important for the regulation of the protocadherin (*Pcdh*) gene cluster in mice. Protocadherins are important for cell adhesion and are predominantly, but not exclusively, expressed in the nervous system. *PCDH* genes create cell-surface diversity, are monoallelically expressed in neurons, and are important for neuronal connectivity, self-recognition, and self-avoidance. Studies in SMCHD1 knockout mouse embryos show that the Pcdh cluster is in fact dysregulated. We are interested in studying the role that SMCHD1 has on the *PCDH* cluster in a neuronal cell context as well as in FSHD muscle cells to fully decipher the normal function of SMCHD1. To investigate the regulation effects that SMCHD1 has on the *PCDH* gene cluster a series of knockdown experiments are being carried out followed by RNA-sequencing wherein we can interrogate the regulation of the *PCDH* gene cluster. For these studies we are using the neuronal like cell type SK-N-SH, which has been shown to express a specific set of protocadherins. In addition a series of quantitative chromosome immunoprecipitation followed by sequencing (ChIP-seq) and ChIP-qPCR experiments are being done to interrogate the regulation of the *PCDH* gene cluster in FSHD2 patient primary muscle cells and, as a reference, in neuronal SK-N-SH cells. Thus functional genomics strategies are being used to dissect the function and hierarchy of SMCHD1 in *PCDH* gene cluster regulation in human cells, and functional studies of muscle cell to establish the role of *PCDH* gene cluster proteins in a non-neuronal context. The long-term goal of these lines of investigation is to uncover the consequences of FSHD2 mutations on *PCDH* gene cluster regulation and to understand how these alterations in transcription are leading/contributing to FSHD phenotypes.

532S

Allelic expression in single human primary fibroblasts. *C. Borel[1], P.G. Ferreira[1], F. Santoni[1], O. Delaneau[1], A. Fort[2], K.Y. Popadin[1], M. Garieri[1], E. Falconnet[1], P. Ribaux[1], M. Guipponi[1], I. Padioleau[1], P. Carninci[1], E.T. Dermitzakis[1], S.E. Antonarakis[1].* 1) Dept Genetic Medicine, Univ Geneva Medical Sch, Geneva, Switzerland; 2) RIKEN Center for Life Science Technologies, Division of Genomic Technologies, Yokohama, Japan.

The study of gene expression in mammalian single cells using genomic technologies is now possible and the rules of allelic gene expression are now being investigated. We have used single-cell RNA sequencing to detect the allele-specific mRNA level in 203 single human primary fibroblast cells over 133,633 unique heterozygous SNVs (hetSNVs). We have observed that at the time point of analyses each cell contains mostly transcripts from one allele from the majority of genes. We revealed that 76.49 % of the hetSNVs with a RPSM superior to 20 displayed a monoallelic ratio (RPSM: reads at a single nucleotide position per sequencing read length and per million mapped reads). The allelic specific expression in single cells correlated with the rate of transcription initiation, mRNA half-life, abundance of the cellular transcript. Haplotype analysis suggests that two neighboring genes show independent allelic expression (i.e. not in the same haplotype). These results have implications in cellular phenotypic variability. C.B. and P.G.F. contributed equally.

533M

Exploring the Human Genome for Functional, Non-Coding Variation: Implications for Diseases of the Peripheral Nerve. *W. Law[1], A. Antonellis[1,2].* 1) Human Genetics, University of Michigan, Ann Arbor, MI; 2) Neurology, University of Michigan, Ann Arbor, MI.

Sequence variations within transcription factor binding sites (TFBS) can result in dysregulation of gene expression and directly cause or alter the severity of human disease. The implicated variants are often single-nucleotide polymorphisms (SNPs) within TFBS that alter the DNA binding affinity. We hypothesize that such regulatory SNPs (rSNPs) impact the function of genes important for peripheral nerve development and function, and represent excellent candidate modifier loci for peripheral nerve disease. To address this hypothesis, we developed a pipeline to computationally identify and functionally characterize novel rSNPs impacting the peripheral nervous system. Specifically, we computationally identified candidate TFBS across the human genome based on: conservation of 5 base pair (or longer) genomic segments between human, mouse, and chicken; overlap with a known, validated SNP; and localization to non-protein-coding sequences. These efforts revealed 6,164 conserved SNPs in the human genome. To evaluate the efficacy of this approach, we characterized a pilot set of 160 SNPs on chromosomes 21, 22, and X. We also tested the ability of our computational pipeline to identify rSNPs in a transcription-factor centric manner focusing on SOX10, which is essential for peripheral nerve function. We filtered the above 6,164 SNPs to find those that reside within a SOX10 consensus binding site or a SOX10 ChIP-Seq binding peak. These efforts revealed 24 putative SOX10-relevant rSNPs. To evaluate the regulatory potential of our predicted rSNPs, a genomic segment surrounding each rSNP was cloned upstream of a luciferase reporter gene and transfected into immortalized Schwann cells, motor neurons, and muscle cells to approximate a peripheral nerve motor unit. Luciferase assays were performed to identify genomic segments with at least a 5-fold increase in activity over a control vector. This revealed 32 novel genomic segments with enhancer activity in at least one cell line. Of the 32 genomic segments, nine showed a significant difference in luciferase activity between the major and minor alleles. Currently, we are further characterizing a subset of prioritized rSNPs by deleting the genomic segment in cultured cells using CRISPR/Cas9 technology followed by RNA expression studies to identify the gene(s) regulated by the elements. Any SNP associated with altered gene expression will represent candidate modifiers of peripheral nerve-related diseases.

534T

Sex-specific Association of Alu Retrotransposons with Gene Expression. *S. Linker[1], D. Hedges[2].* 1) HIHG, University of Miami, Miami, FL; 2) The Ohio State University, Columbus, OH.

Transcriptional regulation is a highly coordinated process necessary for the proper function of an organism. Although many mediators of gene expression have been identified (e.g.: CpG methylation, promoter accessibility), much of the repetitive content of the genome remains unexamined for its direct association with expression. Retrotransposons (RT) have the potential to alter gene regulation. However, to date there have been limited studies analyzing that association of RTs with alterations in gene expression in humans. RTs are a major cause of variation both between individuals as well as within an individual, due to somatic retrotransposition. This indicates that if RTs are indeed modifiers of gene expression, then they can influence phenotypic variability among as well as within humans. It is therefore essential for disease studies to identify if RTs have an effect on gene expression. The research presented here examines the association of RTs with differential gene expression. Our results suggest that the presence of a RT within a gene is indeed associated with decreased expression of the same gene. More importantly, this effect of RTs on gene expression was sex-specific, and largely driven by effects in males. Lastly, the genes found to contain variations in RTs are enriched for Autism susceptibility genes, which are also enriched for differences in expression between the sexes. These results provide evidence for RTs in mediating sex-specific effects of human gene expression.

535S

MicroRNA Expression, *APOE* and *TOMM40*. *L.M. Bekris[1,2], Y. Shao[1], M. Shaw[1], B. McCue[1], J.B. Leverenz[3].* 1) Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH; 2) Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University; 3) Lou Ruvo Center for Brain Health, Cleveland Clinic, Cleveland, OH.

MicroRNA (miRNA) play an essential role in post-transcriptional gene regulation in the brain. Differences in brain miRNA expression between Alzheimer's disease (AD) and cognitively normal controls have been reported. Genetic variants located at the *APOE* locus are the strongest risk factors for late-onset AD described to date. A complex regulatory structure exists at the *APOE* locus that includes putative regulatory sites within the *TOMM40* gene. Little is known about the influence of miRNA post-transcriptional regulation at this locus. The aim of this investigation was to demonstrate that brain miRNA correlate with *APOE* and *TOMM40* brain expression differentially according to brain region or the presence of AD. In addition, it was demonstrated, using bioinformatics tools, that a subset differentially expressed miRNA that correlated with *APOE* or *TOMM40* expression are predicted to target *APOE* or *TOMM40*. To achieve this aim, miRNA array and miRNA quantitative RT-PCR were used to measure miRNA expression in post-mortem brain from AD (n=21) or cognitively normal age-matched control (n=21) hippocampus (HP) and cerebellum (CB). Quantitative RT-PCR was used to measure *APOE* and *TOMM40* mRNA. Western blotting was used to measure brain ApoE and Tomm40 protein. Linear regression was used to determine if *APOE* and *TOMM40* mRNA expression correlate with protein expression in AD (n=8) compared to cognitively normal age-matched controls (n=8) or in HP compared to CB. Linear regression analyses were used to determine if miRNA qRT-PCR levels differentially correlate with *TOMM40* or *APOE* expression in AD (n=6) compared cognitively normal age-matched controls (n=6). Post-mortem brain miRNA were identified that significantly correlate with AD or control brain *APOE* or *TOMM40* expression levels. In addition, a subset of miRNA were predicted to target *APOE* or *TOMM40*. These results suggest that an intricate *APOE* locus regulatory structure may be further fine-tuned by miRNA post-transcriptional modulation according to brain region or disease status.

536M

Gene silencing mediated by endogenous miRNAs under heat stress conditions in mammalian cells. *M. Fukuoka, M. Yoshida, A. Eda, M. Takahashi, H. Hohjoh.* Department of Molecular Pharmacology, National Institute of Nuroscience, National Center of Neurology and Phychiatry, Kodaira, Tokyo, Japan.

MicroRNAs (miRNAs) are 21~23-nucleotide-long small non-coding RNAs that are processed from longer transcripts forming stem-loop structure by digestion with RNase III enzymes, Drosha in the nucleus and Dicer in the cytoplasm. The processed miRNAs are incorporated into the RNA-induced silencing complex (RISC) and functions as mediators in gene silencing. MiRNAs play essential roles in gene regulation by inhibiting translation of messenger RNAs (mRNAs), and by digestion of mRNAs, or by RNA interference, during various vital function and phenomenon. Heat shock is an external stress, and cells that are subjected to heat shock immediately respond to such an environmental stress for protecting themselves and for maintenance of homeostasis. A major response to heat shock is the expression of a certain set of proteins, referred to as heat-shock proteins, and heat-shock proteins play important roles in cells under heat stress conditions. A recent study (Mol Cell, 39, 292-299, 2010) suggested that HSP70 and HSP90 likely associated with the Ago2 protein that is a major component of RISC, and would participate in loading of small RNAs into RISCs, leading us to the possibility that there may be some sort of relationship between heat shock and gene silencing. In this study, we investigated gene silencing mediated by miRNAs in mammalian cells exposed to a mild hyperthermia (40 °C) by means of miRNA activity assay using a luciferase reporter gene as well as miRNA expression analysis using a DNA microarray. Our data indicated that the gene silencing activities involving miRNAs were enhanced without increasing in their expression levels under such a heat-stress condition. Additionally, the gene silencing activity appeared to be independent of the cytoprotective action involving heat shock proteins that are immediately activated in heat-shocked cells and that function as molecular chaperons for restoring heat-denatured proteins to normal proteins. Our findings suggested the possibility that gene silencing involving endogenous miRNAs might play a subsidiary role in heat-shocked cells for an aggressive inhibition of the expression of heat-denatured proteins.

537T

MicroRNA profiling of human lymphoblastoid, iPS and neural stem cell lines shows overlapping but distinct expression patterns. *S. Kumar, J.E. Curran, J. Blangero.* GENETICS, TEXAS BIOMEDICAL RESEARCH INSTITUTE, SAN ANTONIO, TX.

EBV immortalized lymphoblastoid cell lines (LCL) have been widely banked for studying a variety of diseases, including complex genetic disorders. These cell lines with the emerging induced pluripotent stem cell (iPSC) technology represent an important resource for disease modeling particularly in providing specific cells/tissue type to experimentally test the hypotheses developed by statistical genetics methods. Accumulating evidence now provides support that microRNAs (miRNAs) play a critical role in transcription factor-induced reprogramming of iPSCs, maintenance of pluripotency and differentiation. Deciphering the role of miRNAs in pluripotency, self-renewal and differentiation will reveal new molecular mechanisms involved in induction and maintenance of the pluripotent state as well as triggering the reprogramed iPSCs differentiation into a specific cell type of interest. We used episomal plasmids encoding reprogramming factors (OCT3/4, SOX2, KLF4, L-MYC and LIN28) to induce pluripotent stem cells from six LCLs of our San Antonio Family Heart Study (SAFHS) subjects. The reprogrammed iPSCs were then differentiated into neural stem cells (NSC). Both iPSC reprogramming and NSC inductions were confirmed by immunocytochemistry. Small RNAs from six LCLs and their generated iPSC and NSC lines underwent deep sequencing. The differential expression analysis of the miRNA across these three cell types (LCLs, iPSCs and NSCs) from six SAFHS individuals identifies 41 miRNAs that show statistically significant (*corrected p value < 0.05*) change in expression. Comparison of the average expression from LCLs to iPSCs shows ≥2 fold up-regulation of 6 miRNAs (miR-92b, miR-151a, miR-93, miR-182, miR-222, miR-17) and ≥2 fold down-regulation of 13 miRNAs (let-7a-3, miR-16-2, miR-181b-1, miR-181a-1, miR-193b, miR-10a, miR-21, miR-181a-2, miR-181b-2, miR-361, let-7a-2, miR-342, miR-16-1). Comparing iPSCs to NSCs, 4 miRNAs were ≥2 fold down-regulated (let-7a-3, miR-148a, miR-182, let-7a-2) and 23 were ≥2 fold up-regulated (miR-130b, miR-103a-2, miR-103a-1, miR-26a-1, miR-30c-1, miR-197, miR-92b, miR-484, miR-10a, miR-27b, miR-30d, miR-151a, miR-25, miR-93, miR-182, miR-222, miR-532, miR-92a-2, miR-30c-2, miR-148b, miR-26a-2, miR-17, miR-92a-1). In summary, our miRNA expression data encompassing human LCLs, iPSCs and NSCs shows a broad resetting of the miRNA profiles during reprogramming of pluripotent cells and their differentiation into neural lineages.

538S

The AGO3 Slicer controls the dynamic mitochondria-nuage localization of Armitage and contributes to piRNA biogenesis. *Y. Li[1], H. Huang[2], K. Szulwach[1], G. Zhang[2], D. Chen[2], P. Jin[1].* 1) Dept Human Gen, Emory Univ, Atlanta, GA; 2) 1State Key Laboratory of Reproductive Biology Institute of Zoology, Chinese Academy of Sciences.

Piwi-interacting RNAs (piRNAs) silence transposons in animal germ cells. In Drosophila, although the reciprocal “Ping-Pong” cycle of piRNA-directed RNA cleavage, catalyzed by the PIWI proteins Aubergine (Aub) and Argonaute3 (AGO3) through their Slicer activity, has been proposed to expand the population of antisense piRNAs in response to transposon expression. It remains unclear whether and how the Slicer activity of AGO3/AUB promotes the process of the secondary piRNA amplification. Here we generate the transgenic flies that could express AGO3 Slicer mutant forms, and show that the Slicer activity of AGO3 is essential for germline development and piRNA amplification in general. Surprisingly, we find that the piRNA-binding but not the Slicer activity of AGO3 is required for its Nuage localization. Furthermore, we show that the AGO3 inhibits the Aub:Aub Ping-Pong process in a Slicer-independent manner. Finally, we observe that the AGO3 Slicer activity is essential for the proper sub-cellular localization of the key component in piRNA pathway, Armitage, between mitochondria and nuage. Thus, our findings reveal a novel mechanism that couples mitochondria and nuage to regulate the secondary piRNA amplification.

539M

The functional role of lncRNAs in breast tissue-specific gene regulation. *E. Wagner[1,2], Y. Hao[3], Y. Liu[3], C. He[1,2].* 1) Dept. of Epidemiology, Indiana University School of Public Health, Indianapolis, IN; 2) Indiana University Melvin and Bren Simon Cancer Center, Indianapolis, IN; 3) Dept. of Medical & Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN.

Long noncoding RNAs (lncRNAs) are a class of RNAs that are more than 200nt but do not code for protein. Increasing evidence suggested that lncRNAs play a key role in gene regulation. However, few studies have investigated the role of lncRNAs in gene regulation at genome-wide level, especially in target tissue. In order to determine the functional role of lncRNAs in tissue-specific gene regulation, we have conducted a genome-wide association study of co-expression of lncRNAs and mRNAs in 119 breast tumors and 23 normal breast tissue samples using RNA-seq data (2X100nt). After QC and normalization, the expression levels of 4,777 known lncRNAs and 21,128 mRNA were estimated. We used linear regression to examine the association between each mRNA and lncRNA expression, adjusting for age, race, tissue type, and sequencing batch. Multiple testing was corrected using Bonferrroni correction. We considered two possible underlying mechanisms: 1) differential co-expression mechanism in which the association between lncRNA and mRNA expression differs in normal and tumor breast tissue; 2) dose-response co-expression mechanism in which there is an association between lncRNA and mRNA expression and the association is similar in normal and tumor breast tissue. Accordingly, we tested an interaction term between lncRNAs and tissue type in regression model in order to detect the differential co-expression. We found that 694 pairs of lncRNA-mRNA (comprising of 524 lncRNAs and 76 mRNAs) are differentially co-expressed in tumor and normal breast tissue (p for interaction<1×10–12). Genes such as TFE3, DEDD, and DNAL1 were differentially co-expressed with multiple lncRNAs. These genes are involved in chromosomal translocation, apoptosis, and cellular movement, functions implicated in tumorigenesis. For dose-response co-expression, we found 804 pairs of lncRNA-mRNA (comprising of 13 lncRNAs and 706 mRNAs) are similarly co-expressed in tumor and normal breast tissue. To identify cancer-related co-expression, we further restricted the results by differentially expressed lncRNAs and mRNAs in tumor and normal breast tissue, and identify genes including GABARAPL1, DCTPP1, KIFC1, and MAGEF1. Ingenuity pathway analysis revealed that the identified genes are primarily clustered in inflammatory response, cell death and survival, and cellular movement. Our study suggested that lncRNAs play a key role in regulation of genes that are important for breast cancer initiation and progression.

540T

Genome-wide discovery of tissue specific properties of human small non-coding RNAs. *Y. Leung, U. Lyle H, G. Brian D, W. Li-San.* University of Pennsylvania , Philadelphia, PA19104, U.S.A.

Small non-coding RNAs (sncRNAs) comprise many classes of distinct RNAs with diverse and unknown functions. Thus, a complete and accurate catalog of sncRNAs is essential for elucidating the non-coding repertoire. Here we report the first genome-wide study to explore the tissue specificity properties along with other characteristics of sncRNAs such as cleavage specificity and secondary structure. We collected publicly available small-RNA sequencing data from adipose, B-cell, blood, brain, colon, heart, kidney, liver, lung and skin samples from 275 human normal, non-diseased subjects. We first performed stringent data quality control steps to account for biological and inter-laboratory variation. We then quantified the expression values across all tissues and identified 2215 sncRNAs, 83% (85% in GENCODE) of which belong to the miRNA, tRNA, snRNA, snoRNA and rRNA classes. Overall, we find 33% of sncRNAs to be expressed in at least one tissue type, with no tissue expressing more than 46% of all possible sncRNAs. Distributions of expression levels were very similar across tissues, with miRNA having the highest absolute expression levels, followed by snoRNA C/D box, H/ACA box and tRNA. Clustering of overall expression levels of these tissues mimics the stem cell potency groups: liver and lung (endoderm), brain and skin (ectoderm), heart and kidney (mesoderm). We then analyzed the tissue specificity of sncRNAs using Shannon entropy introduced by Schug et al. Unexpectedly, the tRNA (58%), scRNA (51%), and snRNA(48%) classes were highly tissue specific even though they are generally thought to perform housekeeping cellular functions. Over half of the tissue-specific tRNAs were identified in B cells, similar to tRNA-derived miRNAs reported recently, and some of the scRNAs are residing in the same compartment as lincRNAs, which may suggest why scRNAs have diverse roles across tissues. In contrast, miRNAs and lincRNAs were found to be less tissue-specific (38% and 32%, respectively). We also found that tissue-specific miRNAs are enriched for functions in organ morphogenesis and development, pattern specification processes, regionalization and regulation. Tissue-specific miRNAs also have a more precise 5' end cleavage position as compared to ubiquitous miRNAs. Taken together, these results suggest that sncRNAs typically thought to function in basic biological processes may work through different mechanisms depending on tissue type.

541S

A tissue-specific lincRNA in the *TRAF1-C5* risk locus as a putative *cis*-regulator bridging genetics and disease. *T. Messemaker[1,2], R.B. Marques[1], T.W. Huizinga[1], A. Adriaans[1,2], A.M. Bakker[1], A.W.G. Berendsen[1], N. Daha[1], R.E.M. Toes[1], H.M.M. Mikkers[2], F. Kurreeman[1,3].* 1) Department of Rheumatology, Leiden Universitiy Medical Centre, Leiden, Netherlands; 2) Department of Molecular Cell Biology, Leiden Universitiy Medical Centre, Leiden, Netherlands; 3) Department of Human Genetics, Leiden Universitiy Medical Centre, Leiden, Netherlands.

In the last decade genome wide association studies (GWAS) have identified genetic polymorphisms that associate with Rheumatoid arthritis (RA). However, the way these genetic regions contribute to disease remains ill defined. We previously identified the *TRAF1-C5* locus as a predisposing risk factor to the development of RA. In the present study we investigated functional consequences of this risk locus. Using expression quantitative trait loci (eQTL) datasets, we observed an association between the risk allele and expression of *TRAF1* and *C5* at the mRNA level in various blood-related cell types. As part of an underlying mechanism we identified a novel large non-coding RNA intergenic of *TRAF1* and *C5* (*C5T1-lincRNA*). The lincRNA is transcribed in the same orientation as *TRAF1* and *C5* by RNA polymerase II, is highly transcribed in liver, and its expression is rapidly induced in different immune cells by specific immune stimuli. Expression of *C5T1-lincRNA* correlated with either *C5* or *TRAF1* expression in a tissue specific manner. In addition, knockdown of the intergenic transcript in a hepatocyte cell line resulted in decreased C5 levels. Together our data support the involvement of a novel lincRNA in regulating *C5* and *TRAF1* expression. We propose that this lincRNA, which is fully located within the associated region, is responsible for the RA associated altered RNA levels of TRAF1 and C5 and plays a role in RA susceptibility.

542M

Conservation and Novel Functions of Non-coding RNAs. *T.R. Gingeras.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

The approximately three billion base pairs of the human DNA represent a storage devise encoding information for hundreds of thousands of processes that can go on within and outside of a cell. This information is partially revealed in distinguishable transcripts that are potentially derived from 12 billion nucleotides, considering strandedness and the diploid nature of the most genomes. Results stemming from the efforts to catalogue and analyze the RNA products made by many cells in the Encyclopedia of DNA Elements (ENCODE), in human, fly-worm (modENCODE) and mouse ENCODE projects have contributed to help shed light on both the functional content, how this transcriptional information is organized and its evolutionary conservation. Currently for the human genome, there are annotated a total of 196,520 transcripts (41 % coding) within 57,820genic regions (35% coding) based on v19 in Gencode. Analyses of both the regulatory and transcriptome data sets of phase 2 of the human ENCODE projects underscore several underappreciated lessons concerning the conservation and novel functions of long and short non-coding (nc)RNAs. These lessons include: a) specific short ncRNA classes are enriched in micro-vesicles and used for intercellular communication b) the conservation of long ncRNAs between mouse and human of several transcriptional features that appear to be independent of the degree of sequence similarity and c) a multiplicity of cap modifications present on short ncRNAs. These and other insights drawn from the multiple ENCODE and other genome-wide data sets assist in understanding what is often seen as dauntingly complex but elegantly organized genome and continues to prompt a reconsideration of the definition of a gene.

543T

The landscape of retrotransposon expression in human lung adenocarcinomas. *A. Biton[1], B. Xue[2], D. Risso[3], T. Speed[3,4], L. He[2], Biton.* 1) Department of Medicine, Lung Biology Center, University of California, San Francisco, San Francisco, California 94110 USA; 2) Department of Molecular and Cell Biology, Division of Cellular and Developmental Biology, University of California at Berkeley, Berkeley, CA 94720, USA; 3) Department of Statistics, University of California, Berkeley, CA 94720, USA; 4) Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, Parkville VIC 3052, Australia.

Nearly half of the human genome consists of repetitive elements. The largest class of mammalian repetitive elements are retrotransposons, a class of transposable elements that can duplicate in the genome through RNA intermediates using a “copy and paste” mechanism. Retrotransposons are considered a major source of genetic variation and de novo insertion of retrotransposons in the genome can introduce mutations and modify gene splicing and expression. Normally silenced in adult somatic cells, aberrant retrotransposon activation has been associated with cancer. For example, using exome sequencing data, de novo insertions of L1 elements have been detected in tumor suppressor genes of various cancer types (Helman E. 2014 Genome Res). In this study, we are interested in detecting potential aberrant retrotransposon expression in cancer genomes and determining their potential effect on proximal genes.

We studied the expression of retrotransposons (LINEs, LTRs, and SINEs) in RNA-Seq data from tumor/normal pairs of lung adenocarcinoma from public database including The Cancer Genome Atlas. The analysis of the retrotransposon expression was performed at two levels. First, at the level of the consensus sequence, by mapping the reads to the consensus sequences listed in RepBase, to crudely evaluate the abundance of each retrotransposon class. Second, at the locus-specific level, by using the RepeatMasker annotation in the genome and investigating the potential effect of the retrotransposons in regulating the adjacent genes.

We observed that lung adenocarcinomas have distinct retrotransposon expression profiles compared to the normal samples. Indeed we found that tumor and normal samples could be distinguished solely based on their retrotransposon expression profiles. Interestingly, in addition to LINEs and SINEs, we observed aberrant expression of LTRs (including but not limited to HERV18 or HERVK). Using the locus-centric approach, we found that derepression of some retrotransposons could be involved in inhibition of their adjacent gene. We also observed examples of concurrence of exonization of the retrotransposon and differential exon-usage within the same gene. Altogether, these results suggest that retrotransposons may play roles in cancer development.

544S

Overexpression of the *FMR1* mRNA in premutation carriers is isoform specific. *C. Yrigollen[1], D. Pretto[1], J. Eid[2], H-T. Tang[1], E. Loomis[1], C. Raske[1], B. Durbin-Johnson[3], P. Hagerman[1, 4], F. Tassone[1, 4].* 1) Department of Biochemistry and Molecular Medicine, University of California Davis, Sacramento, CA; 2) Pacific Biosciences, Inc., Menlo Park, CA; 3) Department of Public Health Sciences, University of California Davis, School of Medicine, Davis, CA; 4) MIND Institute, University of California Davis Medical Center, Sacramento, CA.

Premutation carriers of the *FMR1* gene are at risk to develop multiple fragile X associated disorders, including the neurodegenerative disorder fragile X-associated tremor/ataxia syndrome (FXTAS). Elevated *FMR1* mRNA levels have been observed in premutation carriers; whether all *FMR1* isoforms are differentially expressed or only specific ones, is currently unknown.

Thus, using *single* molecule real time (SMRT®) sequencing and qRT-PCR we have determined which of the potential transcribed *FMR1* isoforms are expressed and, more importantly, determined if all or only some of the transcripts are overexpressed in premutation carriers compared to normal controls.

The SMRT® sequencing approach has provided us complete cDNA sequence reads from exon 1 through exon 9 and exon 9 - exon 17 within a single *FMR1* transcript. Our results showed that out of the 24 predicted isoforms in human, only 16 *FMR1* isoforms are expressed in peripheral blood monocytic cells, brain, and primary fibroblast cells. Interestingly, although the isoforms missing exon 12 were the most abundant in both premutation carriers and controls in all three tissues, only two were differentially expressed in carriers.

Specifically, of the 16 observed isoforms, *Iso10* and *Iso10b*, both with exons 12 and 14 spliced, and differing only in the splicing acceptor site used at exon 17, were significantly overexpressed.

These findings demonstrate that the differential expression of *FMR1* observed in premutation carriers is isoform specific. The loss of exon 12 results in a shortened unstructured variable loop between two β sheets in the FMRP KH2 domain while the splicing of exon 14 results in the loss of a nuclear export signal and introduce a frameshift that alters the amino acid sequence from exon 15 - exon 17, giving rise to truncated proteins.

Why *FMR1* alternative splicing regulation is compromised in premutations is not yet understood, nor the functional consequences that result from abundance of *Iso10* and *Iso10b*, but these changes in expression of specific isoforms are likely to play a relevant role for understanding the pathogenesis of FXTAS and of the *FMR1* associated disorders.

545M

High throughput identification of exonic sequence variation that exhibits allelic imbalance in RNA splicing. *R. Soemedi[1,2], K. Cygan[1], A. Leblang[3], W. Fairbrother[1,2,4].* 1) Center for Computational Molecular Biology, Brown University, Providence, RI; 2) Molecular and Cellular Biology and Biochemistry, Brown University, Providence, RI; 3) Department of Computer Science, Brown University, Providence, RI; 4) Center for Biomedical Engineering, Brown University, Providence, RI.

Introduction: The rate of discovery of causative mutations for human diseases has accelerated in recent years, but a significant proportion of published causality lacks direct evidence of pathogenicity of the observed variants. Recent studies have predicted that ~15–30% of exonic variants can alter RNA splicing, which may result in exon skipping and aberrant splice site usage; both can potentially lead to more deleterious changes due to frameshift. We have developed High Throughput Splicing Assays (HTSAs) capable to test thousands of wild-type/mutant pairs of splicing substrates simultaneously for the presence of allelic imbalance in RNA splicing. **Methods:** We designed a library of 200-mer oligonucleotides consisting of 5,192 pairs of mutant and wild-type version of exons (median exon length = 76 nt) and their flanking introns derived from reported missense and nonsense mutations at Human Gene Mutation Database (n=4964) and common SNPs (n=228). HTSA *in vivo* was done by incorporating the library into three-exons splicing constructs under CMV promoter that were subsequently transfected into Hek293 cells. HTSA *in vitro* was performed in HeLa nuclear extract using in vitro-transcribed library containing two-exons splicing constructs. Original pool of substrates and spliced RNAs were extracted, made into Illumina libraries and subjected to parallel sequencing in HiSeq2500. **Results:** We observed allelic imbalance in RNA splicing in 1,036 (21%) rare exonic mutations and 17 (7%) SNPs that were tested in both HTSA *in vivo* and *in vitro* (>1.5 fold-change and $P < 0.05$ by two-sided Fisher's exact test, followed by Benjamini-Hochberg FDR for multiple testing correction). We further identified alternative splice site usage and intron retention events that occur in mutant sequences exclusively in 373 (7%) and 49 (1%) mutant/wild-type pairs of the library substrates, respectively ($P< 0.001$ by two-sided Fisher's exact test). The resulting datasets were used to identify factors attributed to sequence variation that impact splicing. **Conclusions:** Our study facilitates a better characterization of exonic sequence variation of potential deleterious consequence resulting from aberrant splicing. Our findings indicate the urgency of integrating splice-altering variant hypothesis in characterizing causal variants in human disease. Machine learning approaches can be further applied to generate a superior prediction tool for identifying variants that lead to gene processing defects.

546T

Impact of microRNA binding site polymorphisms on gene expression variation. *T. Annilo[1], U. Võsa[1], T. Esko[1], A. Metspalu[1,2].* 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia.

Functional interpretation of GWAS-identified loci presents a major challenge since only a small fraction of variants has a direct effect on protein-coding regions. MicroRNAs (miRNA) bind to specific motifs called miRNA response elements (MREs) in the 3'-UTRs of mRNAs, repressing the activity of their targets by affecting mRNA stability and/or protein translation. Rapidly increasing number of known human miRNAs and SNPs in MREs provides an opportunity to systematically investigate the potential impact of common genetic variants on regulatory interactions between miRNAs and their target mRNAs.

We hypothesized that polymorphisms within the MREs that disrupt or generate a binding site can affect the expression of target genes, leading to allele-specific expression modulation. Cis-eQTLs (variants associated with gene expression changes) from the meta-analysis of 7 cohorts (~5 500 individuals) were retrieved from bloodeQTLBrowser (http://genenetwork.nl/bloodeqtlbrowser/). SNPs and proxies (R^2=1) were mapped to the 3' UTRs of the cis-eQTL-affected genes. Putative effect of the SNP on the binding of miRNA was estimated using polymiRTS 3.0 database.

In total, 4829 individual SNPs and proxies were found in the putative MREs of 2166 REFSEQ transcripts. Out of 2578 miRNAs present in miRBase v20, 2573 were predicted to be influenced by at least one SNP. All together, 24 858 SNP-transcript-miRNA associations were reported. To restrict the analysis to blood-specific miRNAs only, a list of 123 miRNAs was identified using 11 publicly available small RNA sequencing datasets using a threshold of 10 reads in at least 6 studies. In addition, to add confidence to miRNA binding site prediction, MRE SNPs were filtered against 36 independent AGO2, AGO3 and AGO4 PAR-CLIP and HITS-CLIP datasets from starBase database. The filtering using a consensus list of blood-specific miRNAs revealed a set of 1128 SNP-transcript-miRNA associations, consisting of 123 miRNAs, 771 SNPs and 614 transcripts. After filtering against AGO binding-sites, a conservative set of 331 SNP-transcript-miRNA associations remained, consisting of 101 miRNAs, 206 SNPs and 192 transcripts. We suggest that our analysis enables identification of miRNA-associated eQTLs for future functional studies and helps to prioritize genetic variants that are important in complex diseases or traits.

547S

High-Resolution Genomic Analysis of Human Mitochondrial RNA Sequence Variation Reveals Genetic Determinants of Post-transcriptional Modification and Interplay with the Nuclear Genome. *A. Hodgkinson[1], Y. Idaghdour[1,2,3], E. Gbeha[1], J.-C. Grenier[1], E. Hip-Ki[1], V. Bruat[1], J.-P. Goulet[2], T. de Malliard[1,2], P. Awadalla[1,2].* 1) CHU Sainte-Justine Research Centre, Department of Pediatrics, Faculty of Medicine, Université de Montreal, 3175 Chemin de la Côte-Sainte-Catherine, Montreal, Quebec H3T 1C5, Canada; 2) CARTaGENE, 3333 Queen Mary Road, Office 493, Montreal, Quebec H3V 1A2, Canada; 3) Department of Biology, Division of Science and Mathematics, New York University Abu Dhabi, Post Office Box 129188, Abu Dhabi, United Arab Emirates.

Mutations in the mitochondrial genome have been associated with a wide range of diseases, as well as fundamental biological processes such as ageing. Previous studies have found that sequence variation across mitochondrial genomes within a single individual (heteroplasmy) is reasonably rare, however little is known about the extent and distribution of sequence variation in the mitochondrial transcriptome. By ultra-deeply sequencing mitochondrial RNA from the whole blood of ~1000 individuals at an unprecedented level of coverage (>6000X) from the CARTaGENE project, we find a remarkable level of variation across individuals, as well as sites that show consistent patterns of post-transcriptional modification. Using a genome-wide association study we find that the rate of post-transcriptional modification at functionally important sites in mitochondrial transfer RNAs is under strong genetic control, largely driven by a missense mutation in MRPP3 that explains ~22% of the variance. Furthermore, we find nuclear genetic variants that are associated with modifications in other mitochondrial genes, and explore the relationship between mitochondrial post-transcriptional modification and nuclear gene expression. These results reveal a major nuclear genetic determinant of post-transcriptional modification in mitochondria, develop our understanding on the level of cross-talk between the nuclear and mitochondrial genomes, and suggest that tRNA post-transcriptional modification may impact cellular energy production.

548M

RCARE: RNA-Sequence Comparison and Annotation for RNA Editing. *SY. LEE[1], JG. Joung[2], CH. Park[1], JH. Park[3], JH. Kim[4].* 1) Seoul National University Biomedical Informatics (SNUBI) and Systems Biomedical Informatics Research Center, Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul 110799, Korea, MS; 2) Seoul National University Biomedical Informatics (SNUBI) and Systems Biomedical Informatics Research Center, Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul 110799, Korea, PhD; 3) Seoul National University Biomedical Informatics (SNUBI) and Systems Biomedical Informatics Research Center, Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul 110799, Korea, BS; 4) University College of Medicine, Seoul 110799, Korea, MD, PhD, MS.

Posttranscriptional sequence modification of transcripts through RNA editing is an important mechanism for regulating protein function and is correlated with human disease phenotypes. The identification of RNA-editing sites is a fundamental step in the study of RNA editing. Determining the location of condition-specific RNA-editing sites and elucidating their functional roles will help toward understanding various biological phenomena that are mediated by RNA editing. In the present study, it is proposed that RCARE (RNA-sequence comparison and annotation for RNA editing), which searches, annotates, and visualizes RNA-editing sites using thousands of previously known editing sites, can be used for comparative analyses between multiple samples. This a new tool, RNA-sequence comparison and annotation for RNA editing (RCARE), which determines condition-dependent RNA-editing sites, provides rich and systematic annotations, and delivers 'executive summary' plots for the annotation and comparison of multiple samples with RNA-Seq data through a user-friendly web interface. RCARE provides a Python script that users can use to preprocess raw RNA-Seq data and convert FASTQ and BAM files into RNA VCF (variant call format) files in a practical time frame using a desktop computer. RCARE is freely available at http://www.snubi.org/software/rcare/.

549T

The RNA Editing Landscape in Acute Myeloid Leukemia. *E. Meduri[1,2], B. Huntly[1,2].* 1) Department Of Haematology, Cambridge Institute For Medical Research, Univeristy Of Cambridge, Cambridge, United Kingdom; 2) Wellcome Trust—Medical Research Council Cambridge Stem Cell Institute, Cambridge, United Kingdom.

Recent next generation sequencing efforts have annotated the role of somatic DNA mutations in the generation of malignancies such as acute myeloid leukemia (AML). However, growing evidence points to epigenetic variation as a further and perhaps equal contributor to the malignant phenotype. We present paired transcriptomic and whole genome/exome sequencing (WGS/WES) analysis for 172 AML patients from the Cancer Genome Atlas (TCGA) to examine the role of RNA editing in this epigenetic variation. The RNA editing gene Adenine deaminase, RNA specific (ADAR) is highly expressed across all samples, whereas ADARB1 and ADARB2 are expressed at very low levels. Analysis of WGS/RNA-Seq from a subset of 33 patients was performed, using a stringent pipeline to identify RNA editing events (REE). Surprisingly, we observed a marked heterogeneity in the number of editing events per sample, which ranged from <1000 to >13,000 events and these events demonstrated minimal overlap with available REE databases (RADAR, DARNED). The events were predominantly A>G and T>C, reflecting ADAR mediated A>I editing. Restricting our analysis to recurrent REE (arbitrarily defined as occurring in >20% of the 33 samples), we found editing events targeted ~2,300 protein-coding genes. As previously described the majority of the REE (>90%) occur in non-coding DNA, namely intergenic, intronic and UTR regions. However, of interest, for the events in coding regions the majority are predicted to alter the amino acid sequence of the protein (66% non-synonymous vs. 34% synonymous). Comparing the edited genes with gene ontology processes, we found significant enrichment in functions including regulation of transcription, gene expression and protein modification processes. We are currently extending our analysis to the remaining 139 samples in the TCGA database to further identify commonly edited genes and will assess the functional consequences of the REE in these genes including alterations of protein coding, RNA abundance and RNA splicing. We are also comparing these tumor specific patterns to normal tissue patterns of RNA editing and are correlating the patterns of REE with known prognostic and outcome measures in AML, such as cytogenetic and mutational status and overall survival.

550S

Global identification of binding sites for the splicing regulatory factors SRSF5 and hnRNPA1. *G.H. Bruun[1], T.K. Doktor[1], S. Brøner[1], A. Masuda[2], B. Palhais[1], A.R. Krainer[3], K. Ohno[2], B.S. Andresen[1].* 1) Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark; 2) Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan; 3) Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA.

RNA splicing is a fundamental process required for correct expression of almost all protein coding genes. It is critically dependent on numerous splicing-regulatory elements (SREs) throughout our genes and on correct expression and activity of the splicing regulatory factors (SRFs) that function by binding to the SREs. Thus, mutations/SNPs that affect SREs, may disrupt splicing and a rapidly increasing number of human diseases are now known to be caused by aberrant splicing. Unfortunately, the specific binding motifs are not precisely defined for even the best characterized and most abundant SRFs and maps of their location, as part of SREs in human disease genes, are missing or sparse. We have used iCLIP (Individual nucleotide-resolution Crosslinking and Immunoprecipitation) to identify the genome-wide binding sites of two SRFs, namely SRSF5 and hnRNPA1. HeLa cells expressing T7-tagged versions of the SRFs are UV-crosslinked to induce covalent bonds between RNA binding proteins and RNA. SRFs are immunoprecipitated with a T7-antibody, followed by SRF removal by proteinase treatment. Eventually this RNA is used as a template for cDNA synthesis and PCR. The amplified products are then submitted for next-generation sequencing and analyzed. SRSF5 and hnRNPA1 binding sites were identified across the genome and consensus binding motifs were generated. The consensus binding site for hnRNPA1 resembles the previously published hnRNPA1 binding matrices generated from other cell types, thus confirming the validity of our approach. Among the possible SRSF5 binding sites identified was exon 2 in the ACADM gene. A disease causing silent mutation, c.87A>G, causes ACADM exon 2 skipping in patient cells, and RNA affinity purification showed that SRSF5 binding to an SRE is disrupted by the c.87A>G mutation. One of the identified binding sites of hnRNPA1 is located across the weak 5'splice site of a disease-causing pseudoexon in the MTRR gene, indicating that hnRNPA1 inhibits pseudoexon inclusion in normal cells. In line with this, mutagenesis of this weak 5'splice site in a minigene results in pseudoexon activation. We have generated global maps of SRSF5 and hnRNPA1 binding sites and identified splicing events, which depend on SRSF5 and hnRNPA1. Moreover, we have illustrated how this can be used to identify and characterize disease causing aberrant splicing for instance caused by apparently neutral sequence variants.

551M

Recursive splicing: a novel regulatory mechanism in the human brain that contributes to the processing of long genes. *W.A. Emmett[3], C.R. Silbey[1,2], N. Haberman[1], M. Briese[2,4], .D Trabzuni[1,5], M. Ryten[1,6], J. Hardy[1], M. Modic[2,7], T. Curk[8], V. Plagnol[3], J. Ule[1,2], UK Brain Expression Consortium.* 1) Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, WC1N 3BG, UK; 2) MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, UK; 3) University College London Genetics Institute, Gower Street, London WC1E 6BT, UK; 4) Institute for Clinical Neurobiology, University of Würzburg, Versbacherstr. 5, 97078, Würzburg, Germany; 5) Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh 11211, Saudi Arabia; 6) King's College London, Department of Medical & Molecular Genetics, Guy's Hospital, London SE1 9RT, UK; 7) Institute of Stem Cell Research, German Research Center for Environmental Health, Helmholtz Center Munich, 85764 Neuherberg, Germany; 8) Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia.

Recent studies have shown that expression of long genes in the mammalian brain is perturbed by specific regulatory factors linked to neurodevelopmental or neurodegenerative disorders, suggesting unique regulatory mechanisms. Here we find that most genes that are long in all vertebrate species are specifically expressed in the brain. Moreover, computational analysis of deep human post-mortem brain RNA-Seq data followed by extensive experimental validation showed that in some of these long genes, the longest intron uses a newly discovered regulatory mechanism in humans called recursive splicing.These recursive splice sites (RSS) are highly conserved across evolution and many of them appear to have common evolutionary origins. They are recognized by the spliceosome, and followed by a poison or alternative exon. Owing to the high level of conservation and their likely functional role, these sites are strong candidates to harbour disease causing mutations and should be considered in future genome sequencing projects, especially of neuro-developmental disorder cases.

552T

Complex alternative splicing patterns in human hematopoietic cell subpopulations revealed by third-generation long reads. *A. Deslattes Mays[1], E. Zheng[2], R. Sebra[3], P. Baybayan[2], M. Schmidt[1], A. Wellstein[1].* 1) Lombardi Cancer Center, Georgetown University, Washington, DC; 2) Pacific Biosciences, Menlo Park, CA; 3) Icahn School of Medicine at Mount Sinai, Institute for Genomics and Multiscale Biology, New York, NY.

Background: Alternative splicing expands the repertoire of gene functions and is a signature for different cell populations. Here we characterize the transcriptome of human bone marrow subpopulations including progenitor cells to understand their contribution to homeostasis and pathological conditions such as atherosclerosis and tumor metastasis. To obtain full-length transcript structures, we utilized long reads in addition to RNA-seq for estimating isoform diversity and abundance. Method: Freshly harvested, viable human bone marrow tissues were extracted from discarded harvesting equipment and separated into total bone marrow (total), lineage-negative (lin-) progenitor cells and differentiated cells (lin+) by magnetic bead sorting with antibodies to surface markers of hematopoietic cell lineages. Sequencing was done with SOLiD, Illumina HiSeq (100bp paired-end reads), and PacBio RS II (full-length cDNA library protocol for 1 - 6 kb libraries). Short reads were assembled using both Trinity for de novo assembly and Cufflinks for genome-guided assembly. Full-length transcript consensus sequences were obtained for the PacBio data using the RS_IsoSeq protocol from PacBio's SMRTAnalysis software. Quantitation for each sample was done independently for each sequencing platform using Sailfish to obtain the TPM (transcripts per million) using k-mer matching. Results: PacBio's long read sequencing technology is capable of sequencing full-length transcripts up to 10 kb and reveals heretofore-unseen isoform diversity and complexity within the hematopoietic cell populations. A comparison of sequencing depth and de novo transcript assembly with short read, second-generation sequencing reveals that, while short reads provide precision in determining portions of isoform structure and supporting larger 5' and 3' UTR regions, it fails in providing a complete structure especially when multiple isoforms are present at the same locus. Increased breadth of isoform complexity is revealed by long reads that permits further elaboration of full isoform diversity and specific isoform abundance within each separate cell population. Sorting the distribution of major and minor isoforms reveals a cell population-specific balance focused on distinct genome loci and shows how tissue specificity and diversity are modulated by alternative splicing.

553S

Analysis of the expression levels of chitinase-like proteins, Ym1, Ym2 and breast regression protein-39, in mouse tissues. *M. Ohno, Y. Kida, M. Sakaguchi, Y. Sugahara, F. Oyama.* Dept Applied Chemistry, Kogakuin Univ, Hachioji, Tokyo, Japan.

Mice produce chitinase-like proteins (CLPs), which are structurally homologous to chitinases but lack the ability to degrade chitin. The CLPs belong to glycosyl hydrolases family 18. Mice express primarily three CLPs, including, Ym1, Ym2 and breast regression protein-39 (BRP-39). Recently, CLPs have attracted considerable attention due to their increased expression in a number of many pathological conditions, including asthma, allergies, rheumatoid arthritis and malignant tumors. Little is known, however, about the significance of CLPs increased expression levels during pathophysiological states. Because CLPs lack chitinolytic activity and detectable functions, their biochemical properties have only been partially defined. The quantification of Ym1, Ym2 and BRP-39 individually is important steps in gaining insight into the *in vivo* regulation of the CLPs. In this study, we established quantitative real-time PCR system to quantify the expression of Ym1, Ym2 and BRP-39 individually and compare their expression levels with those of mammalian chitinases and reference genes in mouse tissues. We found that Ym1 and BRP-39 mRNAs were expressed at a high level in the mouse lung, whereas Ym2 mRNA was the most abundant CLP in the stomach, followed by lung. Furthermore, the expression levels of Ym1 and BRP-39 in the mouse lung were higher than those of two active chitinases.

554M

Cross-species gene expression analysis of chitinase-like proteins with mammalian chitinases using qPCR in mouse and human tissues. *F. Oyama, M. Ohno, Y. Kida, M. Sakaguchi, Y. Sugahara.* Dept Applied Chemistry, Kogakuin Univ, Hachioji, Tokyo, Japan.

Mice and humans express two active chitinases, chitotriosidase and acidic mammalian chitinases. They also produce chitinase-like proteins (CLPs), which lack any detectable functions. Mice express Ym1, Ym2 and breast regression protein-39 (BRP-39), whereas humans produce YKL-40, the human homologue of BRP-39, but do not synthesize homologues of Ym1 and Ym2. The chitinases and CLPs exhibit sequence homology to bacterial chitinases and belong to glycosyl hydrolases family 18. YKL-40 levels have been shown to be elevated in patients with rheumatoid arthritis, atherosclerosis, asthma, allergies, and cancers. Here, we established and validated quantitative real-time PCR systems to quantify the expression of YKL-40 and compared their expression levels with those of mammalian chitinases and reference genes using the same scale in human tissues. We found that YKL-40 mRNA was widely expressed in human tissues. The highest levels of YKL-40 mRNA were detected in human liver, followed by the kidney and lung tissues. Then we quantified and compared the expression levels in mouse and human tissues using the human-mouse hybrid standard DNA. Our results indicate that BRP-39/YKL-40 mRNA was expressed at high levels in the mouse lung but was weakly expressed in normal human lung. There were prominent differences in BRP-39/YKL-40 expression in the lungs of humans and mice. In contrast, YKL-40 mRNA expression in human liver tissue was higher than BRP-39 expression in mouse liver.

555T

Transcriptome Profiling of Gamma Secretase Inhibition in Breast and Ovarian Cancer Cells. *A. Mannermaa[1,2], H. Ahonen[1,2], H. Peltonen[1,2], M. Hiltunen[3], A. Haapasalo[3], M. Heinäniemi[4], M. Nykter[5], V. Kytölä[5], V-M. Kosma[1,2].* 1) School of Medicine, Institute of Clinical Medicine, Pathology & Forensic Med, University of Eastern Finland, Kuopio, Finland; 2) Imaging Center, Clinical Pathology, Kuopio University Hospital, Kuopio, Finland; 3) School of Medicine, Institute of Clinical Medicine, Neurology, University of Eastern Finland, Kuopio, Finland; 4) A. I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland; 5) Institute of Biomedical Technology, University of Tampere, Tampere, Finland.

Gamma secretase (GS) is a transmembrane protease complex that acts on over 90 substrates, many of which are involved in signaling networks dysregulated in cancer and Alzheimer's disease. GS inhibitors are known to inhibit cancer growth, and this is largely attributed to their ability to disrupt Notch signaling. It is, however, likely that the other substrates also play a role in the anticancer effect of GS inhibition. Here we report a whole transcriptome sequencing study assessing the effect of DAPT, a GS inhibitor, on gene expression in MCF7, MDA-MB-453 and OVCAR3 cells. The RNA-sequencing generated 73 Gb of raw sequence data. Our data shows that GS inhibition (GSI) by DAPT causes cell line specific and global alterations in gene expression in MCF-7, MDA-MB-453 and OVCAR-3 cells. The analysis (edgeR) yielded 15, 6 and 36 differentially expressed (DE) genes for MCF-7, MDA-MB-453 and OVCAR-3 cells (FDR = 1%). None of the genes were shared by all cell lines. Next, DE was analyzed between grouped treated and untreated samples. A total of 61 genes were identified to be differentially expressed, 30 of them downregulated and 31 upregulated (FDR = 1%). Notably, NOTCH1 was included in the downregulated group, and the upregulated genes included ribosomal and histone protein coding genes. The most prominent biological effects were downregulation of Notch signaling pathway and upregulation of ribosomal and histone proteins. In order to explore the biological effect of GSI, the list of DE genes was analyzed for KEGG pathway enrichment. Upregulated genes were enriched in Ribosome (q=2.25E-09) and Systemic lupus erythematosus (SLE) (q=0,011) pathways. The SLE pathway was enriched due to its pathway containing histone genes. As expected, the downregulated genes were found enriched in Notch signaling pathway, which included EP300, NOTCH1, CREBBP, HES1 and NCOR2 (q=0,0046). Further, Connectivity Map analysis applied to the GSI signature identified 16 compounds (FDR = 1%) suggesting that the gene expression signature of GSI shares significant similarity with CDK, HDAC and Topoisomerase 1 inhibitors. Overall, our data proposes that the effects of GSI are mediated through multiple mechanisms, implying that the cancer growth inhibiting properties of gamma secretase inhibitors are not exclusively due to inhibition of Notch signaling.

556S

Understanding cellular heterogeneity of pancreatic islets of Langerhans using single-cell RNA-seq. *M. Garieri, C. Borel, F.A. Santoni, C. Howald, P. Ribaux, P.A. Halban, S.E. Antonarakis, E.T. Dermitzakis.* Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland.

Pancreatic islets of Langerhans consist of several endocrine cell types. Previous studies of sorted cell populations have provided limited insight into the cellular diversity within islets. Here, we investigate the heterogeneity of human islet cells by performing single cell (SC) RNA-seq. We studied islet cell transcriptomes using islets from 2 healthy organ donors and 2 individuals with type 2 diabetes (T2D). RNA-seq was performed in the whole islet (bulk) and 149 SCs from the healthy donors (respectively 68 SCs and 81 SCs) and 126 from T2D (respectively 63 SCs and 63 SCs). On average we obtained 36.6 millions reads per SC. Clustering analysis allowed us to identify the two most abundant cell types, one expressing high levels of insulin (INS) and the other glucagon (GCG); two additional but less abundant cell populations were each expressing the other major islet hormones, either somatostatin (SST) or pancreatic polypeptide (PPY). In particular we were able to identify in the samples of the two normal individuals the following gene specific SCs: 92 INS, 12 GCG, 2 SST and 4 PPY SCs. For the two T2D individuals we found respectively 31 INS, 26 GCG, 1 SST and 0 PPY SCs. Interestingly we notice that the previously assumed mutually exclusive expression of INS and GCG in beta and alpha cells respectively is not universally observed since many cells expressing high levels of both INS and GCG suggesting an intermediate alpha-beta cell phenotype: 16 from the two Normal individuals and 37 for the T2D SCs were expressing INS and GCG. This novel cell subpopulation was more abundant in the patients than in the normal individuals (Fisher p-value < 0.0003). Currently we are performing RNA-seq of sorted beta SCs from the same individuals to assess the variability within this specific cell subpopulation. In addition, comparison with known eQTLs and allele-specific expression analysis is ongoing in order to identify T2D specific gene expression regulation. This study of single cells is expected to provide a better understanding of the different cell types of the pancreatic islets in in health and disease.

557M

Analysis of Genome-Wide RNA-Sequencing Data Reveals Age of the CEPH/Utah (CEU) Lymphoblastoid Cell Lines Systematically Biases Gene Expression Profiles. *Y. Yuan, L. Tian, D.S. Lu, S.H. Xu.* Chinese Academy of Sciences and Max Planck Society (CAS-MPG) Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.

In human, Lymphoblastoid cell lines (LCLs) from the CEPH/CEU (Centre d'Etude du Polymorphisme Humain - Utah) family resource have been widely used for examining the genetics of gene expression levels. However, we noted that CEU/CEPH cell lines were collected and transformed approximately thirty years ago, much earlier than the other cell lines from the pertaining individuals, which we suspected could potentially affect gene expression, data analysis and interpretation of results. In this study, by analyzing RNA sequencing data of CEU and the other three European populations, we systematically examined and evaluated the potential confounding effect of LCL age on gene expression levels and patterns. Our results indicated that gene expression level and pattern of CEU cell lines have been biased by the older age of CEU cell lines. Interestingly, most of CEU-specific expressions are associated with functions such as cell proliferation which are more likely due to older age of cell lines rather intrinsic characters of the population. We suggested the results should be carefully explained when CEU LCLs were used for transcriptomic data analysis in future studies.

558T

Transcriptome analyzes in human corneas derived from keratoconus and control individuals from Poland. *M. Gajecka[1,2], J.A. Karolak[1], M. Kabza[1], D.M. Nowak[1], P. Polakowski[3], J.P. Szaflik[3].* 1) Department of Genetics and Pharmaceutical Microbiology, Poznan University of Medical Sciences, Poznan, Poland; 2) Institute of Human Genetics, Polish Academy of Sciences, Poznan; 3) Department of Ophthalmology II, Medical Faculty, Medical University of Warsaw.

Keratoconus (KTCN, OMIM 148300) is a degenerating, usually bilateral disorder of the eye, characterized by progressive stromal thinning resulting in conical shape of the cornea. Here we present a pilot keratoconus study based on transcriptome analyzes in human corneas derived from keratoconus and control individuals from Poland. A total of 96 individuals from Polish population are included into this study: 48 keratoconus patients and 48 individuals without keratoconus phenotype. Detailed clinical evaluation in each individual involved in the study was completed. Protocol for RNA extraction from human cornea was established. Total RNA from corneas for RNAseq and small RNA analyzes was extracted. Illumina HiScan system is applied in transcriptome analyzes in this study. Transcriptome analyzes were initiated. Preliminary results of transcriptome analyzes in human corneas will be presented. Since genetic factors involved in keratoconus etiology are in majority unrecognizable, the newest available technologies are necessary to be involved in further keratoconus research. Support: National Science Centre in Poland, Grant no. 2012/05/E/NZ5/02127.

559S

Defining tissue compartment gene expression dysregulation in asthma by multi-tissue whole transcriptome sequencing. *A. Wesolowska-Andersen[1], R. Davidson[1], C. Urbanek[1], G. Sajol[2], C. Eng[3], J. Rodriguez-Santana[4], M. Castro[2], E.G. Burchard[3,5], M.A. Seibold[1,6].* 1) Integrated Center for Genes, Environment and Health, National Jewish Health, Denver, CO; 2) Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO; 3) Department of Medicine, University of California, San Francisco, CA; 4) Centro de Neumologia Pediatrica, San Juan, PR; 5) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA; 6) Department of Pediatrics, National Jewish Health, Denver, CO.

The extent to which gene expression in asthmatics is dysregulated among systemic immune (blood) tissue relative to respiratory airway tissues is unclear. Additionally, although 22 genome-wide association study (GWAS) asthma risk loci have been identified, expression of 161 genes within these loci (+/–200 kb of the reported markers) in the blood and airway tissues has not been systematically examined nor has their differential expression in asthma disease. To investigate blood vs. airway gene expression patterns in asthma we performed whole transcriptome (WT) RNA-seq on whole blood and brushed nasal airway epithelium from Puerto Rican children with asthma (n=10) and controls (n=10); and whole blood and brushed bronchial airway epithelium from adults with asthma (n=12) and healthy controls (n=8). Whole transcriptome RNA-seq was performed in a 2×100bp paired end fashion using Illumina Truseq methodology. RNA-seq data was analyzed using the Tuxedo analysis pipeline. We identified 440, 347, and 70 genes differentially expressed among asthmatics in the bronchial, nasal, and blood WT data, respectively. There was substantial overlap (65 genes) in upper (nasal) and lower (bronchial) airway differential expression. In contrast, we found less overlap of blood differentially expressed genes with both the lower (14 genes) and upper airways (9 genes). Among the 161 GWAS genes, 72.3% (99 genes) were expressed in all tissues, 85.1% (137 genes) were expressed in at least one tissue, 23.4% (32 genes) were expressed only in airway tissues, and 2.2% (3 genes) were only expressed in the blood. There was a significant enrichment for GWAS genes among genes differentially expressed in the airway (17 genes, p=9.4×10–6), in contrast only 1 GWAS gene was among the genes differentially expressed in blood of asthmatics (p=0.38). Among differentially expressed GWAS genes were *IL1RL1*, which encodes for the IL-33 receptor, signaling of which is known to drive airway Th2 inflammation, and *IKZF3*, a chromatin remodeling factor important for B lymphocyte differentiation. Our results indicate asthma disease is associated with a larger number of airway (both upper and lower) vs. blood gene expression changes. Additionally, we find largely unique groups of differentially expressed genes in the asthmatic airway vs. blood. Our examination of GWAS loci genes suggests that risk conferred by multiple asthma GWAS genes may operate through modulation of gene expression, specifically in the airway.

560M

Identification of Gene Expression Signatures in Alopecia Areata. *J. Cerise[1], A. Jabbari[1], J. Mackay-Wiggan[1], M. Duvic[5], M.K. Hordinsky[6], V.H. Price[7], D. Norris[8], R. Clynes[1,2,4], A.M. Christiano[1,3].* 1) Dept of Dermatology, Columbia University, New York, NY; 2) Dept of Pathology, Columbia University, New York, NY; 3) Dept of Genetics & Development, Columbia University, New York, NY; 4) Dept of Medicine, Columbia University, New York, NY; 5) Dept of Dermatology, MD Anderson Cancer Center, Houston, TX; 6) Dept of Dermatology, University of Minnesota, Minneapolis, MN; 7) Dept of Dermatology, UCSF, San Francisco, CA; 8) Dept of Dermatology, University of Colorado, Denver,CO.

Alopecia areata (AA) is a common autoimmune disease manifesting with hair loss that ranges from spontaneously resolving patches (patchy AA, AAP) to lifelong total scalp (alopecia totalis, AT) or total body (alopecia universalis, AU) involvement. AA is a highly prevalent autoimmune disease in which the hair follicle is attacked by cytotoxic T lymphocytes. Obstacles to identifying treatments for AA include limited understanding of disease pathogenesis, as well as the molecular signatures of affected lesions. In order to identify gene expression signatures that correlate with disease, we conducted a comprehensive expression analysis of affected scalp skin of patients with AA. Scalp biopsies from approximately 20 subjects with AAP, 20 subjects with either AT or AU, and 20 healthy control subjects were obtained from lesional/perilesional scalp skin of AA patients and normal scalp skin of healthy controls. Additionally, non-lesional samples were taken from patients with AAP. RNA was extracted from these samples and microarray analysis was performed on Affymetrix microarray data in order to identify genes differentially expressed among the three sample groups. Using Ingenuity Systems' IPA to interrogate genes upregulated in AT/AU and AAP compared with Normal biopsies, we identified two striking gene expression signatures in total skin from human AA, namely, an IFN-γ response (IFN) signature, including IFN-gamma and IFN-inducible chemokines, a cytotoxic T cell (CTL) signature, including CD8 and granzymes, implicating these effectors as the dominant inflammatory cells in AA pathogenesis. We identified a hair keratin (KER) signature using IPA, and the GO term Epidermis Development using Gene Set Enrichment Analysis (GSEA) as being enriched in genes that were downregulated in AA with respect to Normal human skin. Further, GEDI (Gene Expression Dynamics Inspector) was used as a computational tool that allowed us to correlate several predominant gene expression signatures with disease severity. Our findings uncovered new insights into distinct molecular signatures associated with the different forms of AA, and provide a framework for identifying biomarkers of disease and novel potential therapeutic targets. Our gene expression findings support the notion that AU/AT lesions exhibit an extreme inflammatory profile in the spectrum of disease seen in AAP.

561T

Dynamic Transcriptome Changes in Cell-free Synovial Fluid following Meniscal Injury Suggests the Potential for Early Intervention. *D.D. Vance[1,3], E. Rampersaud[2], A. Mehta[2], S.N. Page[2], B.P. Lesniak[2], J.M. Vance[3], M.A. Pericak-Vance[2], L.D. Kaplan[3], L. Wang[2].* 1) University of Miami, Miller School of Medicine, Miami, FL; 2) John P. Hussman Institute for Human Genomics,University of Miami, Miller School of Medicine, Miami, FL; 3) UHealth Sports Performance and Wellness Institute, University of Miami, Miller School of Medicine, Miami, FL.

Meniscus injury is one of the leading causes of osteoarthritis (OA), often leading to knee joint replacement. Osteoarthritic changes in articular cartilage have been seen as early as 2 years post-meniscal damage. We hypothesize that molecular changes are initiated within the knee capsule at the time of meniscus injury and long before the manifestation of OA. Identification of those early changes via a minimally invasive procedure could facilitate early intervention to prevent OA development. Towards this end, we characterized transcriptome in synovial fluid (SV) at different time points (1 to 7 months) following a meniscal injury. Eight male patients ages 24~48 years with meniscal injury diagnosed by magnetic resonance imaging (MRI) were ascertained. SV was collected from the affected knee joint by needle aspiration at the time of meniscectomy, prior to initial incision. Total RNA was extracted from cell-free SV and sequenced on Illumina HiSeq2000. The RNA-Seq generated 18 million to 59 million reads for each sample. The STAR program was used to align the reads against human reference genome to map RNA species. Aligned reads were normalized by estimating effective library size. The normalized gene expression is expressed as count per million (CPM). The EdgeR program was used to perform clustering and differential gene expression analysis. In total, 764 RNAs (including coding, non-coding RNA and microRNA) were present in the SV (CPM>1 in at least one sample). Clustering analysis demonstrated that samples with longer injury duration (LID) and shorter injury duration (SID) formed distinctive clusters. Differential gene expression analysis revealed that 65 RNAs were down-regulated and 78 RNAs were up-regulated (false discovery rate [FDR]<0.05) in 4 individuals with SID compared to 4 individuals with LID. The most differentially expressed gene is *SLC2A9* ($P=1.2\times10^{-13}$, $FDR=9.6\times10^{-11}$), which is barely detectable in SID but is abundant in LID samples. *SLC2A9* is expressed in normal articular chondrocytes and can be induced by proinflammatory cytokines, e.g. IL-1beta. SLC2A9 is believed to play a role in the development and survival of chondrocytes in cartilage matrices. Our data demonstrate dynamic transcriptome changes in SV following meniscal injury and suggest the utility of using RNA biomarkers in SV to monitor disease progression and guide early intervention.

562S

Defining the transcriptional landscape of microRNAs in human peripheral blood. *G. Hemmrich-Stanisak[1], Z. Du[1], M. Huebenthal[1], M. Paulsen[1], J. Hartwig[1], P. Rosenstiel[1], D. Kabelittz[2], A. Franke[1].* 1) Inst. of Clinical Molecular Biology, University of Kiel, Kiel, Germany; 2) Institute of Immunology, University of Kiel, Kiel, Germany.

Next generation sequencing of small RNAs (sRNAseq) is a widely used application to establish microRNA-based diagnostic or prognostic marker profiles for conditions like diseases. However, it remains a challenging task as PCR-based sRNAseq protocols suffer from the limitation that measurement of a particular miRNA is not independent from other miRNAs. The introduction of biases towards certain miRNAs during sequencing library construction leads to discrepancies in the overall abundance of sequenced miRNAs. Particularly in complex tissues such as peripheral blood these effects can be dramatic. Due to the differential miRNA expression in different blood cell types and the uneven distribution of cell types based on different conditions e.g. diseases, age, gender, etc. it is difficult to distinguish between true positive, false positive but also false negative signals. To enable correction for the above-described confounding effects and to understand in general how miRNAs are distributed in different cell populations in whole blood samples as drawn in clinical practice, we sought to investigate the transcriptional profiles of miRNAs in seven distinct peripheral blood cell populations of 40 healthy individuals using sRNAseq. To complete the picture we also examined circulating miRNAs from serum and investigated the miRNA content of isolated blood-borne exosomes. We used magnetic activated cell sorting (MACS) to purify CD56+ (NK cells), CD19+ (B cells), CD8+ (T cells), CD4+ (Th cells), CD14+ (Monocytes), CD15+ (Neutrophils) and CD235a+ (Erythrocytes) cell populations to gain comprehensive insights into the miRNA transcriptome of most blood cell types. The results indicate distinct cell type specific miRNA profiles that accurately reflect the developmental relationships between the different cell lineages. A substantial fraction of miRNAs is expressed in subsets or exclusively in one single cell type. Interestingly, Erythrocytes have a distinct miRNA signature containing more than 100 different miRNA species - a fact to be taken into account when investigating miRNA transcriptomes of whole blood samples. Clustering analysis revealed distinct groups of co-regulated miRNAs that in turn show close relationships in their respective validated target genes and the corresponding biological pathways. The here presented data will help to interpret miRNA-based marker profiles more carefully, taking into account technically and naturally occurring confounding effects.

563M

Systematic Analysis of Age and Sex Effect Identified Different Behaviors between Coding and Non-coding Genes and Two Age-dependent Patterns of Gene Expression. *M. Narahara[1], Y. Tabara[2], T. Kwaguchi[2], F. Matsuda[2], R. Yamada[1].* 1) Statistical Genetics, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Human Disease Genomics, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.

Sexual dimorphism and aging lead to difference in a wide range of phenotypes (*e.g.* disease prevalence and courses). Elucidating the molecular mechanisms underlying the differences would provide new insights for complex diseases. We analyzed gene expression microarray data of human peripheral blood samples obtained from 298 individuals to characterize sexually dimorphic genes and aging genes. This study provides new insights particularly into differences between the autosomes and X chromosome, and between coding and non-coding RNAs.

We identified 2,570 (2,071) sex-biased transcripts (genes), 1,672 male-biased and 898 female-biased transcripts, by FDR<5%. Autosomal transcripts were more often male-biased, whereas X-specific transcripts were more often female-biased (P=8.9E-07). We estimated that ~7.05% of genes at X-specific regions were incompletely inactivated. Female-biased transcripts on the X chromosome were significantly enriched at the short arm compared to the long arm (P=0.0003). Such a difference was not evident for male-biased genes and ncRNAs.

We assessed aging effects by a LOESS regression, and, by FDR<5%, identified 5,254 (3,891) transcripts (genes) for females but no transcripts for males. Subsequently, we analyzed the 5,254 age-dependent transcripts with hierarchical clustering, which identified two clusters. 38% of aging transcripts were classified into the first cluster that showed increasing tendencies with age, and enriched by ncRNAs. 62% were classified into the second cluster that showed decreasing tendencies with age, and enriched by mRNAs.

Lastly, we analyzed age-sex interaction. We chose the most significant transcripts by FDR<30%, for which we conducted hierarchical clustering, and identified two clusters. Notably, in both clusters, female-male differences were largest at the youngest age and became smaller as age.

In conclusion, we identified large numbers of sexually dimorphic genes and aging genes for human whole blood samples. We analyzed these genes in details such as comparison among chromosomal regions and RNA types, and clustering analysis.

564T

Whole Transcriptome Sequencing in Alzheimer's Disease Reveals Gene Expression Differences Related To *TGFβ1* Pathway. *K. Nho[1,2], V. Ramanan[1,3], S. Risacher[1], X. Xuei[4], H. Gao[2], Y. Liu[2,3], T. Foroud[2,3], H. Edenberg[3,4], A. Saykin[1,2,3,5].* 1) Radiology and Imaging Sciences, Indiana University, Indianapolis, IN; 2) Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, USA; 3) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA; 4) Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA; 5) Department of Neurology, Indiana University School of Medicine, Indianapolis, IN, USA.

Background: Human brain function depends on the precise regulation of gene expression. Gene expression analysis (transcriptome profiling) has become a major focus of neurodegenerative disease research. Prior studies suggest that abnormal gene expression patterns may contribute to the onset and progression of Alzheimer's disease (AD). Our aim was to perform whole transcriptome analysis of blood-derived RNA to identify blood-based biomarkers in AD. **Methods:** We performed RNA-Seq on 37 participants (7 healthy control (HC), 8 subjective cognitive decline (SCD), 5 early mild cognitive impairment (EMCI), 10 late MCI (LMCI), 7 AD) from the extensively studied IMAS (Indiana Memory and Aging Study) cohort. The libraries were sequenced on SOLID 5500XL and the resulting 75 nt reads were mapped to human reference genome (hg19) and a splice-junction library, respectively, using BFAST. Read counts were calculated using bamutils from NGSUtils [Breese MR, 2013]. After normalization, we identified genes that are differentially expressed between any of the diagnosis groups (stages of AD development) using the edgeR package. Benjamini and Hochberg's algorithm was used to control the false discovery rate (FDR). We analyzed for pathways with enrichment of association in order to identify biological pathways overrepresented, with threshold for input defined as nominal gene-level significance (uncorrected $p<0.05$) using MetaCore. **Results:** We identified 31 differentially expressed genes between any of the diagnosis groups within FDR-corrected $p < 0.05$. These included *TGFβ1* (FDR-corrected $p = 3.93 \times 10^{-4}$) and *SERPING1* (FDR-corrected $p = 3.58 \times 10^{-3}$). The deficiency of *TGF*β1 signaling has been shown to increase both Aβ accumulation and Aβ-induced neurodegeneration in AD models. *SERPING1* encodes a protein that inhibits the inflammatory process. These significantly differentially expressed genes suggest possible early stage up-regulation. A pathway analysis of all differentially expressed 1,022 genes with uncorrected $p<0.05$ revealed enrichment in 10 functional pathways within FDR-corrected $p < 0.05$ including multiple pathways related to cell cycle, immune activation, and apoptosis. **Conclusions:** Whole transcriptome sequencing analysis of blood in AD reveals differentially expressed genes related to immune activation including *TGF*β*1*. These data add to the growing understanding of the genetics and pathobiology of AD that may lead to novel therapeutic targets for AD.

565S

Predictive Gene Markers of Multipotent Stromal Cell Proliferation. *I.H. Bellayr, J.L. Lo Surdo, S.R. Bauer, R.K. Puri.* Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, MD.

Multipotent stromal cells (MSCs) are known for their distinctive ability to differentiate into multiple cell lineages such as adipocytes, chondrocytes and osteocytes. They can be isolated from numerous tissue sources including bone marrow. Because of their differential potential and secretion of many growth factor(s), MSCs are thought to exhibit inherent qualities of regeneration and immune suppression. Based on these characteristics, MSCs are seen as advantageous for the field of regenerative medicine in treating a variety of injuries and disorders, in addition to graft versus host disease. Since the percentage of MSCs derived from bone marrow is low, MSCs must be cultured for several cell passages to obtain sufficient cell numbers for a desired cell-based therapy. However, after several rounds of passaging, we have shown previously that the quality of these cells declines as demonstrated by decreases in cell proliferation, increases in cell size, reduced multipotent differentiation potential and differences in gene expression. In this study, we wish to identify molecular markers of proliferation of MSCs that can predict the cell population quality. Human MSCs derived from the bone marrow of 8 different donors were grown under identical conditions and total RNA was harvested at cell passages 3, 5, and 7. The proliferative potential was measured for each donor/passage using two different assays, percent confluency at 96 hours and percentage of EdU positive cells after 6 hours in culture. Total RNA was hybridized on a two color microarray for each donor/passage and gene expression data was correlated with both cell proliferation assays. Using a regression model and a multiplicity adjustment, 28 genes were identified as statistically significant and highly correlated ($0.72 \leq r \leq -0.73$) with cell proliferation. When the significant gene lists were analyzed by Ingenuity Pathway Analysis software, these genes were involved in the top scoring networks ($p<0.05$) of cellular growth and proliferation, cellular development, and cell cycle. Thus, we have identified novel gene markers that are indicative of MSC proliferative quality and may be used to rapidly assess a population of MSCs.

566M

TRANSCRIPTOME APPLIED TO A BLOOM'S SYNDROME: IMMUNOLOGICAL INSIGTHS. *M.M. Montenegro[1,2], G.M. Novo-Filho[1,2], E.A. Zanardo[1,2], R.L. Dutra[1,2], A.T. Dias[1,2], T.V.M.M. Costa[1,2], A.M. Nascimento[1,2], F.B. Piazzon[1,2], C. Milani[1,2], S.C.S. Andrade[3], G. Gasparin[3], C.A. Kim[2], L.D. Kulikowski[1,2].* 1) Department of Pathology, Cytogenomic Laboratory / LIM 03, HC-FMUSP, São Paulo, Brasil; 2) Genetic Unity, Department of Pediatrics, Children Institute, HC-FMUSP, São Paulo, Brasil; 3) Department of Zootechny, ESALQ-USP, Piracicaba, Brasil.

Transcriptome profiling of patterns of RNA expression could be a powerful approach to identify networks of genes that play a role in disease. We investigate the relevant genetic markers that could be associated to chromosomal instability in patients affected by a Bloom's syndrome. Methodology: We performed gene expression profiling using high throughput sequencing (deep-sequencing RNAseq) using the Illumina's HiSeq 2500 platform followed by differential gene expression analysis of samples derived from two patients affected by Bloom's syndrome and three unaffected controls. The data analysis was generated using specialized softwares for preprocessing (CASAVA 1.8.2 e Seqyclean), mapping and alignment (Bowtie2 v2.1.0 e Samtools v.0.1.18) and differential expression analysis (HTSeq-count v.05.4.p2). Results: The RNAseq assay revealed the precise location of transcription limits, with resolution of a single nucleotide and high level of efficiency showing high genetic complexity. It was possible the mapping of 22.334 genes, which 11.938 had minimum coverage required for analysis of differential expression. Of these, 399 were presented differentially expressed, being 216 up regulated in the group with Bloom syndrome and 183 up regulated in the control group. Unexpected all 216 genes of Bloom syndrome are connected to immunological systems. Conclusions: Our results suggested that gene expression network in Bloom's syndrome could interfere in the regulation of the pathways associated with the immunological systems regulation probably caused by disturbance of DNA repair mechanisms. The study of the transcriptome using next generation sequencing is a competent approach to the broad coverage of the sequence with presenting a good cost-coverage. Thus, enable construction of a genic interation network leading to new possibilities to investigate the pathogenesis of Bloom's syndrome just as well in another chromosomal instability syndromes.

567T

Significant transcriptional changes in 15q duplication but not Angelman syndrome deletion dental pulp stem cell derived neurons. *L. Reiter[1,2], R. Memon[3], S. Goortha[1], C. Valdez[1], Q. Tran[4], M. Donaldson[3], D. Bridges[2,5], N. Urraca[1].* 1) Dept. Neurology, UTHSC, Memphis, TN; 2) Dept. Pediatrics, UTHSC, Memphis, TN; 3) Dept. Pediatric Dentistry, UTHSC, Memphis, TN; 4) Dept. Preventive Medicine, UTHSC, Memphis, TN; 5) Dept. Physiology, UTHSC, Memphis, TN.

A major problem in studying neurogenetic syndromes is obtaining live neurons from people with these disorders. We have taken a novel approach to studying gene expression in neurons from Angelman or Dup15q syndrome subjects using dental pulp stem cells (DPSC). Shed teeth were collected to generate DPSC lines from 3 AS deletion, 3 idic(15) and three neurotypical subjects. RNAseq was performed on DPSC and DPSC derived neurons. We identified 20 genes in AS, 120 genes in idic(15) and 3 genes in both groups (*DPT*, *MED12L* and *AKRIC1*) that were significantly different from control DPSC-neurons. Copy number correlated to gene expression for most genes across the 15q11.2-q13.1 critical region. Two thirds of the genes differentially regulated in idic(15) were down-regulated compared to controls including the transcription factors *FOXO1* and *HAND2*, while in AS the genes did not show a clear trend except in the 15q region. Pathway analysis revealed increased cytokine activity related genes including four genes that increased in idic(15) samples: *CCL7*, *CCL2*, *MMP1* and *MMP3*; while steroid hormone biosynthesis genes were slightly enriched in both idic(15) and AS neurons. Overall there were more dramatic changes in gene expression in the idic(15) duplication than AS deletion cell lines, perhaps because the mechanism of AS may be through protein targeting by UBE3A. Nonetheless, the finding of a significant increase in both *HERC2* and *UBE3A* expression in idic(15) neurons and significant decrease in these two genes in AS deletion neurons may impact the AS phenotype, at least in deletion cases.

568S

Viral RNA detection with RNA-Seq using capture technology. *S.M. Gross, I. Khrebtukova, L.C. Watson, F. Schlesinger, S. Pathak, R. Kelley, T. Hill, T. Singer, G.P. Schroth.* Illumina Inc, San Diego, CA.

Identification of viral RNA in human samples is a useful tool for human pathology research and clinical diagnostics. However, most methods for detection of viral RNA, such as qPCR and microarray hybridization, are highly targeted, provide no sequence information, or fail to capture information regarding the broad range of viral RNA abundance reflective of viral load and/or transcriptional activity. Previously, we have demonstrated the utility of using biotinylated capture probes to enable generation of quality RNA sequencing (RNA-Seq) libraries from highly degraded formalin-fixed, paraffin-embedded (FFPE) samples. This capture approach provides a quantitative view of the human coding transcriptome by enriching for next-generation sequencing (NGS) RNA-Seq library fragments derived from exonic sequence over those derived from intergenic or ribosomal RNA, and is able to capture coding RNA that constitutes much less than 1% of the total transcriptome. Using this same concept of sequence enrichment, we are designing an assay to capture NGS library fragments derived from viral sequences of interest from the background of human host sequences. Initial test panels target some important human pathogens, such as human papilloma virus (HPV), enterovirus, hepatitis, and West Nile virus. This method aims to facilitate RNA-Seq for both quantitation of viral RNA in relation to the human host and the ability to collect sequence information useful for the discovery of novel viral variants.

569M

Identifying conserved genomic responses to inflammation in vascular endothelial cells. *L. Antounians[1,2], A. Medina-Rivera[2], J. Dennis[3], L. Rapkin[4], F. Gagnon[3], M.D. Wilson[1,2].* 1) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 2) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) CIHR-STAGE Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 4) Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada.

Vascular endothelial cells are involved in the inflammatory response associated with many diseases. In response to cytokines like tumor necrosis factor alpha (TNFα), a potent pro-inflammatory signaling molecule, NFKB and JNK signaling pathways become activated and regulate the genomic response to inflammation. Transcription factors (TF) and epigenetic modifications control the dynamic regulatory landscape of cells. It is not known if the epigenetic regulation of the inflammatory response is conserved in mammalian species, many of which serve as models for human disease. To understand the epigenetic regulation of vascular endothelial cells in the context of inflammation, we are studying regulatory regions of the genome. We are using chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) to identify epigenetic modifications and TF binding sites: H3K27ac, a mark of active enhancers; H3K36me3, a mark of active gene bodies; H3K4me2, a mark of active and latent enhancers and promoters; and two TFs cJun and RelA that are activated by the signaling cascades JNK and NFKB, respectively. To perform a controlled comparison of endothelial cell inflammatory gene regulation, we are culturing primary endothelial cells isolated from the aorta of four mammalian species - human, mouse, rat, and cow - with and without treatment of TNFα. Using multiple sequence alignments, we are comparing the location of epigenetic modifications and TF binding between species. Since evolutionary conservation of epigenetic modifications or TF binding sites has been shown to enrich for tissue-specific and process-specific functions, we aim to identify conserved and species-specific regulatory elements. Preliminary results show that approximately 16% of enhancers are conserved between human and rat, while 5% of cJun binding sites are conserved for aortic endothelial cells grown under basal conditions. We find that 74% of conserved cJun sites (N=995) and 84% of conserved H3K27ac sites (N=14,188) are shared with human umbilical vein endothelial cells. This suggests that conservation of regulatory regions is important for vascular endothelial cell identity. Evolutionarily conserved regulatory regions will be used to prioritize single nucleotide polymorphisms in regulatory regions related to disease.

570T

Transcriptional and Epigenetic Landscape of Megakaryocytes derived from Human Induced Pluripotent Stem Cells. *L.J. Vasquez[1], TM. Moreau[2], C. Ghevaert[2], N. Soranzo[1,2].* 1) Human Genetics, Wellcome Trust Sanger Institute, Cambridgeshire, United Kingdom; 2) Department of Haematology, University of Cambridge, Cambridge, United Kingdom.

Every year 250 thousands platelet concentrates are issued by the UK Blood and Transfusion Service at a cost of 55 million GBP. One of the main challenge is to meet the demand for highly-matched products required by multi-transfused patients who have developed immunity against common products donated by the general donor population, and critically so for patients with chronic diseases who have long-term need of transfusion support. Platelets are enucleated circulating corpuscules released from bone marrow megakaryocytes (MKs) and can be irradiated to obtain safe end products. They are therefore ideal targets for in vitro production of blood products that are ABO- and HLA-typed from banks of human induced pluripotent stem cells (hiPSCs). Protocols for the derivation of MKs from hiPSCs are becoming available [1]. For instance, a novel transcription factor-driven forward programming protocol has been developed in the Ghevaert laboratory [2]. However, the genomic and functional equivalence of hiPSC-derived to primary blood cells remains to be verified before they are introduced in the clinic. We present here a genome wide characterization of the transcriptomics and epigenetic landscape of hiPSC-derived MKs exhibiting mature platelet markers and functionality. We present results of the global assay of histone markers of enhancer and heterochromatin (H3K4me1, H3K27Ac and H3K27me3). DNA methylation was probed genome wide using Illumina 450K arrays. Next we interrogated whether hiPSC-MKs were equivalent to MKs derived from neonatal cord blood (CB) or adult peripheral blood (PB). Finally, we associated down regulation of a number of platelet specific genes in the former such as CD9 and VWF with incomplete epigenetic remodelling of the hiPSC genome. Indeed, our study provides mechanistic insights in the reported phenotypic and functional differences between fetal, neonatal and adult MKs, offering new developments for in vitro derived platelets for transfuion. References: 1.Takayama, N. & Eto, K. Pluripotent stem cells reveal the developmental biology of human megakaryocytes and provide a source of platelets for clinical application. Cell Mol Life Sci (2012). 2.Moreau, T. et al. Generation of Mature Megakaryocytes from Pluripotent Stem Cells by a Chemically-Defined Transcription Factor-Based Forward Programming Approach. To be submitted.

571S

BRF1 mutations alter RNA polymerase III-dependent transcription and cause neurodevelopmental anomalies. *P.L. Tan[1], F. Hög[2], N. Sowada[3], M. Kousi[1], A. Medeira[4], H. Thiele[5], F. Lepri[6], L. Wenzeck[2], A. Radicioni[7], T.L. Schwarzenberg[8], D.J. Morris-Rosenahl[9], J. Altmüller[5], P. Nürnberg[5,10,11], B. Dallapiccola[6], M.L. Dentici[6], P. Cramer[2,12], C. Kubisch[3], N. Katsanis[1], G. Borck[3].* 1) Cntr of Human Disease Modeling, Duke University Med Cntr, Durham, NC; 2) Gene Center Munich and Department of Biochemistry, Center for Integrated Protein Science CIPSM, Ludwig-Maximilians-Universität München, 81377 Munich, Germany; 3) Institute of Human Genetics, University of Ulm, 89081 Ulm, Germany; 4) Serviço de Genética, Hospital S. Maria, 1649-035 Lisboa, Portugal; 5) Cologne Center for Genomics (CCG), University of Cologne, 50931 Cologne, Germany; 6) Bambino Gesù Children's Hospital, IRCCS, 00165 Rome, Italy; 7) Department of Medical Pathophysiology, Sapienza University, Rome, Italy; 8) Department of Neonatology, Sapienza University, Rome, Italy; 9) National Heart and Lung Institute, Imperial College, London SW3 6LY, United Kingdom; 10) Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, 50674 Cologne, Germany; 11) Center for Molecular Medicine Cologne (CMMC), University of Cologne, 50931 Cologne, Germany; 12) Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany.

RNA polymerase III (Pol III) synthesizes tRNAs and other small non-coding RNAs to regulate protein synthesis. Dysregulation of Pol III transcription has been linked to cancer, and germline mutations in genes encoding Pol III subunits or tRNA processing factors cause neurogenetic disorders in humans, such as hypomyelinating leukodystrophies and pontocerebellar hypoplasia. Here we describe an autosomal recessive disorder characterized by intellectual disability and cerebellar hypoplasia as well as facial dysmorphic features, short stature, microcephaly, dental anomalies, and scoliosis. Whole exome sequencing revealed biallelic missense alterations of BRF1 as the cause of this syndrome, a finding supported by the recapitulation of key neurodevelopmental phenotypes upon suppression of brf1 in zebrafish embryos. BRF1 associates with BDP1 and TBP to form the transcription factor IIIB (TFIIIB), which recruits Pol III to target genes. Disease-causing mutations reduce BRF1 occupancy at tRNA target genes in S. cerevisiae and impair cell growth. Moreover, BRF1 mutations reduce Pol III-related transcription activity in vitro. The alleles were corroborated to be loss of function using the zebrafish in vivo model when tested in the context of the canonical isoform but not in an alternative transcript. These results show that BRF1 mutations that reduce protein activity cause a neurodevelopmental syndrome and that BRF1-mediated Pol III transcription is required for normal cerebellar and cognitive development and also highlight the biological importance of different transcript variants on human pathogenesis.

572M

Association study of the Human Leukocyte Antigen-G and gallstone disease in Han Chinese. *H. Yang[1,2], S. Shih[2,3,4], Y. Lee[1,4,5,6,7].* 1) Medical Research, Mackay Memorial Hospital, New Taipei City, Taiwan; 2) Mackay Junior College of Medicine, Nursing and Management, Taipei, Taiwan; 3) Division of Gastroenterology, Department of Internal Medicine, Mackay Memorial Hospital, Taipei, Taiwan; 4) Department of Medicine, Mackay Medical College, New Taipei City, Taiwan; 5) Department of Pediatrics, Mackay Memorial Hospital, Taipei, Taiwan; 6) Department of Pediatrics, Taipei Medical University, Taipei, Taiwan; 7) Institute of Biomedical Sciences, Mackay Medical College, New Taipei City, Taiwan.

Background and Goal: Gallbladder stone induce inflammatory responses and affect extra-hepatic bile ducts called gallstone disease (GSD). The pathology and environmental risk factors of GSD are well documented. However, genetic polymorphisms of immune or inflammatory regulation in GSD remain unclear. Human leukocyte antigen (HLA)-G was defined as non-classical Major histocompatibility complex class I based on relative low polymorphisms. HLA-G proteins have seven different isoforms from alternative splicing and have immunosuppressive properties. HLA-G expression is able to confer resistance to Nature killer cell- or cytotoxic T lymphocyte-mediated destruction. HLA-G exerts its inhibitory effect broadly as this molecule has been found to interact with numerous inhibitory receptors on a variety of immune cells. Although only a handful of studies have been published that evaluate a possible role for HLA-G in inflammatory conditions of the bowel, we believe it prudent to include this material to increase risk for the development of malignancy in the setting of inflammatory conditions of the chronic gastrointestinal and inflammatory diseases, such as GSD. In this study, we investigated whether polymorphisms of the HLA-G gene were associated with susceptibility of GSD. Methods: Genomic DNA was isolated from whole blood samples of 234 female patients with GSD and 399 gallstone-free female controls. HLA-G +1537A/C, 14-bp insertion/deletion, and +3142G/C polymorphisms were genotyped using PCR-RFLP or Pre-Developed TaqMan Allelic Discrimination Assay. Genotype, allele, and carrier frequencies of the three HLA-G SNPs were determined by direct counting. Hardy-Weinberg equilibrium was assessed for each SNP in both control and study groups using Haploview 4.2. Results: HLA-G +3142G/G genotype shows susceptibility to GSD in female patients (OR = 0.66, Pc = 0.03). Allele +3142G also shows the same result (OR = 0.75, Pc = 0.02). The other two polymorphisms site show the same frequencies with controls. Conclusions: Based on our findings, HLA-G +3142G/G confer as protective in female patient with GSD. Our results need to be replicated and verified in another large-sized cohorts or other ethnics.

573T

Structural Variation Analysis Using Nanochannel Genome Mapping to Evaluate Genome Integrity after Induction of Pluripotency of Human Fibroblasts. *E.H. Cho[1], H. Dai[1], A. Pang[1], R. Williams[2], K. Nazor[2], K. Bhutani[3], N. Schork[3], J. Loring[2].* 1) BioNano Genomics, Inc., San Diego, CA; 2) Center for Regenerative Medicine, The Scripps Research Institute, La Jolla, CA; 3) Human Biology, J. Craig Venter Institute, La Jolla, CA.

Four key genes, POU5F1, SOX2, KLF4, and MYC, are commonly used for reprogramming human fibroblasts into stem cells. Induced pluripotency of human fibroblasts into stem cells currently relies on three different transfection methods of these key genes: retroviral, Sendai virus, and mRNA. The latter two methods are non-integrating methods, while the former method integrates into the genome. Retroviral transfection to induce pluripotency is currently the most efficient method, but due to the potential damaging effects of viral integration into the chromosomes, it is thought to be of high risk for future clinical applications. To investigate whether structural changes result during the process of induction of pluripotency in stem cell genomes we used nanochannel genome mapping technology from BioNano Genomics (BNG) to compare structural differences among cell lines derived by the three reprogramming methods and the parental genome. By producing de novo genome map assemblies of each genome, comparing to a reference genome, and cross comparing all four assemblies, we determined sample-specific SV calls for each of the three methods. After automated assembly and comparison, followed by manual verification, 14 sample-specific SVs were found in the retrovirus-induced sample, 12 in the Sendai virus induced sample, and 8 in the mRNA induced sample, all with respect to the Hg19 human reference genome. Examination of this dataset provided no evidence that any of the methods used to induce pluripotency produced malignant SVs into the genome. Subsequent studies confirming these results using orthogonal methods, such as next-generation sequencing, could provide strong evidence for continued pursuit of clinical applications using any of these methods. Furthermore, confirmation of no significant SV differences using the more efficient retroviral transfection method may give more confidence in the use of this method in clinical therapeutic iPSC applications.

574S

An exercise filled lifestyle may alter the gut metagenome exposed to polychlorinated biphenyl. *E. Rampersaud[1], M. Veerapen[1], M. Toborek[2].* 1) John P. Hussman Inst Human Genomic, University of Miami Miller School of Medicine, Miami, FL; 2) University of Miami.

The gut microbiome comprises of multiple phyla of microbes which responds to environmental stressors. Moreover, the gut microbiome composition has been implied in the catabolism of drugs and other xenobiotic compounds such as polychlorinated biphenyl (PCB). PCBs are persistent organic pollutants which are carcinogenic. Our group has previously shown that exercise can attenuate changes conferred by oral PCB in the murine gut microbiome. However, the role of the gut microbiome in exercise-based PCB degradation remains unclear. We propose to study the composition and functional characteristics of the gut metagenome upon the exposure to PCB and exercise. 32 mice were used in 4 study groups comprising of sedentary and exercised mice which were either exposed to PCB or not (PCB delivery vehicle only). All mice were voluntarily exercised or remained sedentary for 5 weeks and were treated with oral gavage of relevant PCB congeners or vehicle. Mouse feces were collected before gavage and 2 days after gavage. Collected fecal matter was subjected to total DNA extraction and subsequently sequenced using the Illumina HiSeq 2000. All raw reads were assembled, mapped and analyzed using MG-RAST v3.3.6s and MatR v1.0.0. All samples were uploaded and analyzed successfully with MG-RAST. Our analyses revealed significant differences between sedentary and exercised mice gut microbiome -- observably, an increase in bacteroidetes in exercised mice ($p < 0.001$). The bacteroidetes increase is more apparent upon treatment with PCB in exercised mice but a notable decrease in microbiome diversity based on the α-diversity value. No observable microbiome clustering was found between groups. Based on preliminary functional analyses, PCB exposure increases presence of microbes involved in information storage and metabolic proteins ($p < 0.05$). Our results indicate and replicate that exercise attenuates the effects oral PCB exposure phenotypically and alters the gut metagenome. The gut microbiome is known to alter immune systems and the observed alteration likely indicates a possible change to immune responses responsible for PCB degradation. Further studies are required to elucidate specific pathways and mechanisms involved in PCB degradation with the gut microbiome. Potentially, this study can be used to remediate populations exposed to oral PCBs.

575M

Characterizing gut microbiota variation across diverse rural African populations. *M.E.B. Hansen[1], M. Rubel[1], A. Bailey[2], K. Bittinger[2], A. Laughlin[2], A. Ranciaro[1], W. Beggs[1], S. Thompson[2], F.D. Bushman[2], S.A. Tishkoff[1].* 1) Genetics Dept., University of Pennsylvania, Philadelphia, PA; 2) Microbiology Dept., University of Pennsylvania, Philadelphia, PA.

Africans harbor the greatest levels of human genetic variation, and differences in diet are likely to have produced distinct selection pressures resulting in genetic adaptations. An additional important source of variation that can influence both health and disease is the gut microbiome. To better understand how genetic adaptations, culture, and diet interact, we analyzed the gut microbiomes from ethnically diverse rural Africans and urban European and African Americans in Philadelphia using ribosomal marker classification (16s RNA V1/V2) from fecal samples. The African cohort is composed of populations with different diets and subsistence practices, including hunter-gatherers, nomadic pastoralists, and traditional agricultural groups. Samples from 120 rural Africans were extracted, sequenced, and compared with a dataset of 102 European and African American samples, using identical methods and the Roche 454 platform. The impact of diet and ethnicity was characterized by changes in relative abundance of bacterial taxa. Comparison of the rural African to the urban Philadelphia samples confirms previous studies that show marked differences between African and non-African gut flora. In contrast, within Africa there are relatively subtle differences between populations despite highly distinct diets and high levels of genetic differentiation. Among African populations we find that differences in the gut microbiome are not entirely diet dependent, but rather, are correlated also with geography and ethnicity. This represents the largest microbiome study to date of ethnically diverse Africans living indigenous lifestyles and provides novel data from previously uncharacterized African groups.

576T

Interactions between host genetics and diet affect gut microbiota and influence metabolic traits in mouse and human. *E. Org[1], J. Wha[2], Y. Blum[1], B. Emert[1], E.Y. Kang[2], B. Parks[1], R. Knight[3,4], T. Drake[5], M. Laakso[6], E. Eskin[2,7], A.J. Lusis[1,7].* 1) Department of Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA; 2) Computer Science Department, UCLA, Los Angeles, CA; 3) Department of Chemistry and Biochemistry, University of Colorado Boulder, CO; 4) Howard Hughes Medical Institute, Boulder, CO; 5) Department of Pathophysiology and Laboratory Medicine, UCLA, CA; 6) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 7) Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA.

An increasing body of evidence suggest that host genetics as well as the diet affects the composition of gut microbiota. Moreover, it is clear that shifts in microbial communities can lead to cardiovascular and metabolic diseases. In our study we examined natural variation in gut microbiota in a population of about 100 inbreed strains of mice from the Hybrid Mouse Diversity Panel (HMDP) and a population of a 200 men from the Metabolic Syndrome in Men (METSIM) study. Profiling of the gut microbiota in the HMDP (n= 599 mice) revealed that the genetic background strongly influences gut microbiota composition and that the gut microbiome composition explains a considerable portion of the variation in obesity traits. There was insufficient power to estimate heritability in the human sample, nevertheless associations with traits were observed. We also observed evidence of strain-specific shifts in gut microbiota after feeding mice with obesifying, high fat/high sucrose (HF/HS) diet, suggesting a strong effect of host genetics on the plasticity of gut microbiota. Genome-wide association of the panel identified 7 significant loci in mouse genome associated with relative abundances of the families Clostridiaceae, Ruminococcaceae, Lachnospiraceae, Turicibacteracea and Lactobacillaceae. Integrating datasets from multiple intermediate phenotypes (clinical parameters, metabolite levels and gene expression) provided additional support for relationships between specific microbes and metabolic traits. Consistent with our findings in mice, we detected significant correlations between Roseburia intestinalis and metabolic phenotypes in the human study. In addition, we showed similar microbial affinities in human and mice by inferring microbial co-occurrence networks. Our findings suggest that gut microbiome may be one of the key factors in the missing heritability of complex diseases and the integration of microbiota data with different intermediate phenotypes may allow us to obtain valuable insights into diseases-associated metabolic pathways.

577S

Drosophila fragile X mental retardation protein is associated with chromatin and regulates replication stress-induced DNA damage response. *Y. Cheng[1, 2], W. Zhang[2], Y. Li[1], Z. Chen[2], D. Chen[2], P. Jin[1].* 1) Human Genetics, Emory University, Atlanta, GA., USA; 2) State Key Laboratory of Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, 100101, P.R. China.

Fragile X syndrome, a common form of inherited mental retardation, is caused by the loss of the fragile X mental retardation protein (FMRP). As a selective RNA-binding protein, FMRP has been shown predominately localized in cytoplasm regulating the translational control. However, it is known that a small portion of FMRP is present in nucleus and its nuclear function has been elusive. We have found that Drosophila dFmr1 is required for replication stress-induced H2Av phosphorylation in DNA damage response (DDR). Replication stress could induce the expression of dFmr1 and promote the nuclear accumulation of dFMR1. We show that dFMR1 is associated with chromatin in a domain-specific manner upon the stimulation of replication stress, which is essential for its ability to induce the phosphorylation of H2Av. Furthermore, we have performed ChIP-seq analyses using dFMR1 antibody and identified specific dFMR1 binding sites in S2 cells. By comparing the distributions of dFMR1 between control and UV exposure conditions, we found that the sites bound by dFMR1 upon UV treatments significantly overlap with replication origins identified previously. These results together reveal an unexpected nuclear role of FMRP in DDR, and uncover a feed-forward mechanism by which dFmr1 and early DDR induced by replication stress reciprocally regulate each other, thereby synergistically triggers the activity of DDR signaling cascade.

578M

Impact of Structural Variants on the three-dimensional multi-scale chromatin conformation of the human genome. *D. Plewczynski[1,2,3], P. Szalaj[2,3], J. Kim[2], C. Zhang[2], A. Malhotra[2].* 1) ICM, University of Warsaw, Warsaw, mazovia, Poland; 2) The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; 3) The Centre for Innovative Research, Medical University of Bialystok, Bialystok, podlasie, Poland.

In this study we investigate the causative relation between copy number variants (CNV) and three dimensional chromatin structure of various cell types. Spatial arrangement of chromosomes can play a very important role in many biological processes. By rendering distant genomic regions to a close spatial proximity it can precipitate long-range interactions between genes and transcription factors. Recently developed Chromosome Conformation Capture (3/4/5C) techniques allow us to quantify interaction frequency between genomic regions and infer their spatial arrangement. On the other hand recent studies such as the 1000 Genomes Project suggest that Human genomes are very variable, not only in terms of single mutations, but also large scale structural variants. Our three-dimensional simulation method uses homo-polymer modeling techniques, yet is done separately at different scales: whole nucleus, chromosomes, 10mb regions modeled as random walk, 1mb giant loops structures, 100kb rosettes and random loops, 10kb local neighborhoods. For all spatial scales we use the same evolutionary algorithm that constructs an ensemble of chromatin conformations based on a given experimental contact map containing interaction frequencies. Moreover, in order to overcome limitations of homo-polymer models, we map various local genomic features such as active genes, open chromatin marks, GC ratio and others onto 1D chromatin chain. We also consider structural constraints: resistance to stretching and bending, which accommodate for biophysical chromatin fiber properties. As evolutionary operators we employ three mutation operators (single bead displacement, fragment displacement, fragment rotation) as well as a crossover operator (splitting two structures at a random position and gluing them cross-wise), which are chosen with predefined probabilities. In our study we combine the data being generated by various studies including the 1000 Genomes Project regarding structural variants type, size and location, with their spatial localization in three-dimensional model of the nucleus. We use public HiC maps of contacts and the resources from ENCODE project in order to further characterize local 3D neighborhoods. We hypothesize that the shape of chromatin fiber around deletion, duplication, inversion and mobile element insertions is strongly correlated with the type of CNV, suggesting a causative link.

579T

Higher order chromatin structure and *CFTR* gene regulation: roles of *cis*-regulatory elements and CTCF/cohesin complex. *R. Yang, N. Gosalia, J. Kerschner, S.H. Leir, A. Harris.* Lurie Children's Research Center and Department of Pediatrics, Northwestern University, Feinberg School of Medicine, Chicago, IL.

The cystic fibrosis transmembrane conductance regulator (*CFTR*) gene encodes a cAMP-activated chloride ion channel, which when mutated, causes the genetic disease cystic fibrosis. *CFTR* expression is highly tissue-specific and it has been shown that *cis*-elements play an important role in the regulation of *CFTR*. Some of these elements are located far (>100kb) from the gene promoter and regulate gene expression through chromosome looping which establishes physical interactions. In order to better understand the molecular mechanism of *CFTR* gene regulation and look for additional distal regulatory elements we used 4C-seq (chromosome conformation capture combined with high-throughput sequencing). This technique can identify and quantify interactions between selected genomic sites (viewpoints) and rest of the genome. 4C-seq experiments were performed in human intestinal and airway cell lines that use different *cis*-regulatory elements, and also in primary human airway and epididymis cells. Viewpoints were used at the *CFTR* promoter, at intronic and distal enhancers, and at CTCF-binding insulators flanking the gene. The data correlate well with previous 3C data we generated for the *CFTR* locus and also reveal potential novel *cis*-regulatory elements. Moreover, results confirmed the cell type-specific properties of these chromatin interactions. We subsequently performed studies in which we 1) manipulated the levels of the architectural proteins CTCF and cohesin, or 2) removed specific cis-regulatory elements from the endogenous *CFTR* gene by CRISPR/Cas9-mediated targeting and examined the impact on 3D locus structure and *CFTR* expression. These data provide substantial insight into higher order chromatin structure mediated by both *cis*-regulatory elements and the CTCF/cohesin complex and their critical role in *CFTR* gene regulation.

580S

The Chromatin Architecture of a Haploid Human Cell Line. *D.A. Cusanovich, J. Shendure.* Department of Genome Sciences, University of Washington, Seattle, WA.

In contrast with model organisms, human cell lines have not historically been readily compatible with mutagenesis screens to identify genes involved in specific cellular phenotypes. Recent advances in genome editing technologies - specifically, with the CRISPR/Cas9 system - have enabled efficient genome-wide recessive genetic screening in haploid human cell lines (Wang et al. Science 2014). However, taking advantage of haploid genetics limits the choice of a cellular model primarily to KBM-7, a near-haploid chronic myelogenous leukemia cell line, and its haploid derivative, HAP1. While both have been used successfully in previous studies, HAP1 has not been well characterized. HAP1 was derived in a failed attempt to induce pluripotency and has several notable distinctions from KBM-7, including the fact that it is adherent and that it no longer maintains a second copy of chromosome 8 (Carette et al. Nature 2011), but little else has been reported regarding its cellular state. To facilitate further efforts to dissect the genetic architecture of human phenotypes in these haploid cellular models, we have generated RNA-seq and ATAC-seq data from the HAP1 cell line. For comparison, we are also generating equivalent datasets from other myeloid cell lines, including the parental line (KBM-7), and a lymphoblastoid cell line (GM12878) that has been extensively characterized in previous studies. Beyond serving as a reference dataset for further experimentation, characterization of the nuclear architecture of a haploid cell line may offer insights into fundamental human biology. Both the RNA-seq and ATAC-seq datasets allow for hierarchical clustering of the samples to better understand how related their cellular phenotypes are. Furthermore, genes differentially expressed between HAP1 and KBM-7 may help to both define the cellular identity of HAP1 and explain how this cell line can stably maintain a haploid genome. While the primary goal of this work is to aid further interpretation of ongoing functional genomic experiments in this haploid model cell line, the results may also yield a more nuanced understanding of gene regulation in human cells.

581M

A Computational Pipeline for Characterization, Identification, and Significance Analysis of the Somatic Copy Number Variations from Genome Sequencing Datasets. *A. Harmanci[1], A. Serin Harmanci[2], M. Gunel[2], M.B. Gerstein[1].* 1) Molecular Biophysics and Biochemistry, Yale University, Hew Haven, CT 06527, USA; 2) Departments of Neurosurgery and Genetics, Yale Program in Brain Tumor Research, Yale School of Medicine, New Haven, CT 06510, USA.

Large scale genomic aberrations, namely the structural variations, are important driving factors for tumorigenesis and tumor evolution. Among these variants, the unbalanced variants, the copy number variants, are significant because they may change the dosage of the tumor suppressor or oncogenes and drive the formation of tumor. We are presenting a complete pipeline for identification and significance analysis for somatic CNVs using genome sequencing (exome and whole genome) datasets. Our pipeline takes multiple datasets as input and uses a signal processing based segmentation method to identify CNV calls for each sample. Following CNV identification, pipeline utilizes a genome wide test to assign a significance of recurrence for each copy number variant segment. In addition, our pipeline estimates the ploidy, tumor purity and clonal fraction for each segment, generates the closest subclonal events and computes significance for these segments. We compare our pipeline with several other CNV identification tools and present benchmarking results. We also applied our pipeline on several exome sequencing datasets of patients with brain tumor and present significantly altered clonal and subclonal events including tumor suppressor and oncogenes. In addition to coding elements, we also generate significantly altered non-coding elements.

582T

Method for classifying candidate structural variants into true positives and false positives. *H. Parikh[1], J. Zook[1], M. Pratt[2], G. Bartha[2], M. Eberle[3], M. Salit[1].* 1) Genome Scale Measurements Group, Biosystems and Biomaterials Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA; 2) Personalis Inc., 1350 Willow Road, Suite 202, Menlo Park, CA 94025, USA; 3) Illumina Inc., San Diego, CA 92122, USA.

The human genome contains genomic variants ranging in size from small single nucleotide polymorphisms to large structural variants (SVs). SVs include variations such as novel sequence insertions, deletions, inversions, mobile-element insertions, tandem duplications, interspersed duplications and translocations. SVs have long been implicated in phenotypic diversity and human diseases; however, identifying SVs in a whole genome has proven elusive. Recent advances in next-generation sequencing technologies have facilitated the analysis of SVs in unprecedented detail. However, clinical adoption of human genome sequencing requires methods with known accuracy of SV calls across genome. Hence, we developed methods to make highly confident SV and non-SV calls for NA12878, the pilot genome for the Genome in a Bottle Consortium, for which we recently published high-confidence small variants and homozygous reference sites. Because numerous methods have been developed to find candidate SVs, we decided to develop methods to look for evidence of these SVs in mapped sequencing reads from multiple sequencing technologies. We collected 9 different Gold Standard SV call sets totaling 12,738 deletions that were derived from pedigree sequencing, microarrays, and other validation methods. We annotated these candidate SV sites with parameters such as mean coverage, mean insert-size, numbers of discordant pairs, numbers of soft clipped reads, mean mapping quality score, and numbers of heterozygous and homozygous SNP genotype calls from 3 sequencing technologies (Illumina, Moleculo, and PacBio). In addition, we annotated randomly generated regions to understand characteristics of non-SV regions. Graphical visualization of the annotation parameters has shown clear distinction between true positive and false positive SVs. A key advantage of the proposed method is its simplicity and flexibility to generate various annotation parameters from aligned sequence data based on different sequencing datasets from the same genome. This allows integration of multiple sequencing datasets to identify high-confidence SV and non-SV calls that can be used as a benchmark to assess false positive and false negative rates. We are currently creating a classification method based on the annotation parameters to generate both high-confidence SV calls and high-confidence non-SV calls for genomes selected as reference materials by the Genome in a Bottle Consortium.

583S

Genomic CNVs can cause sudden infant death syndrome (SIDS). *A. Pfeufer[1,2], M. Arnold[1], M. Cohen[3], S. Herms[4], T. Dörk[5], T.A. Plötz[1], I. Sinicina[6], E.A. Mitchell[7], M. Donner[8], D.T. Mage[9], E.R. Behr[10], T.W. Mühleisen[11], S. Cichon[4], T. Meitinger[1,2], A. Peters[1,6], H.W. Mewes[1,2], M. Klintschar[5], T. Bajanowski[12], M. Vennemann[13].* 1) Helmholtz Zentrum München, Neuherberg, Germany; 2) TU München, München, Germany; 3) University of Sheffield, Sheffield, UK; 4) University of Basel, Basel, Switzerland; 5) MHH Hannover, Hannover, Germany; 6) LMU München, München, Germany; 7) University of Auckland, Auckland, New Zealand; 8) Dupont Haskell Global Centers for Health & Environmental Sciences, Newark, USA; 9) Biomolecular Core Laboratory, AI duPont Hospital for Children, Wilmington, USA; 10) St. Georges University of London, London, UK; 11) Forschungszentrum Jülich, Jülich, Germany; 12) Universität Essen, Essen, Germany; 13) Universität Münster, Münster, Germany.

The contribution of monogenic disorders to in sudden infant death syndrome (SIDS) is established (most commonly LQTs and MCAD deficiency). In contrast the importance of complex genetic predispositions is less clear. Also copy number variations (CNV) involving critical genes may predispose to SIDS. Patients: We performed genome-wide SNP and CNV genotyping of 368 SIDS cases using the Illumina HumanHap 660v3 quad array. 320 cases originated from the multi-center German study on sudden infant death (GeSID) and 48 cases were recruited from the Sheffield SIDS study in the U.K. As controls we used 823 population based individuals from the KORA cohort. Methods: In a GWAS design we compared SIDS cases with population controls for genome-wide SNP association. As age matching is not feasible in SIDS we performed sex matching of controls and geographical adjustment by multidimensional scaling. We included Chr.X markers as SIDS exhibits 2:1 male:female sex bias indicating possible genetic risk factors on the sex-chromosomes. Markers were imputed to 4.8 Mio SNPs using the haplotype backbone provided by the 1000 Genomes project. Autosomal SNP-markers were analyzed using an additive model adjusted for sex. Markers on Chr. X were analyzed stratified by sex and then meta-analyzed. In another line of analysis we called genome-wide CNVs in the SIDS cases using the CRLMM algorithm. Results: The GWAS analysis showed no significant association beyond genome-wide significance level ($p<5\times10–8$) neither with autosomal markers nor with markers on Chr.X. CNV analysis revealed deletions in five cases: One case had a 5.2 Mb deletion in 5q3, another case had a 2.2 Mb deletion in 18p1 and a third case had a 370 kb deletion in 5q2. Two further unrelated cases had nearly identical 400 kb deletions in 4q1. None of these CNVs involved genes previously identified to be involved with SIDS or Mendelian diseases involved with premature death. Conclusion: Our GWAS was powered to detect common SNP association signals with OR≥1.8 and MAF≥0.2 with ≥95% power. The failure to detect such associations in the human genome is well in accordance with the expectation of strong evolutionary selection against any common genetic SIDS risk factors. The detection of CNVs in 1,4% of cases (5/368) indicates a small but significant contribution of structural genomic mutations to SIDS etiology.

584M

In vivo dissection, causality identification, and network analysis of copy number variants associated with autism spectrum disorders. *C. Golzio[1,2], S. Moon[1], J.R. Willer[1], M.A. Savidan[1], S.C. Brodar[1], I. Blumenthal[3], A. Ragavendran[3], M.E. Talkowski[3], N. Katsanis[1].* 1) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC, USA; 2) Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA; 3) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; Harvard Medical School, Boston, MA, USA.

Copy number variants (CNVs) are frequent lesions involved in both rare and complex human traits. This has raised the challenge of identifying which genes within a CNV drive clinical traits. We have shown previously how the combinatorial use of surrogate phenotypes in zebrafish embryos and genomic studies can help dissect CNVs on 1q21.1, 8q24.3, and 16p11.2. However, although successful in identifying primary drivers, these data also indicated that each of the primary dosage-sensitive transcripts within each CNV were not sufficient to drive the CNV-associated pathology. We therefore extended our study to model *cis*-epistasis within CNVs and to determine global transcriptional changes and dissect likely contributory transcripts. We focused on the 1q21.1 CNV, encompassing a minimal nine-gene region, deletion of which is associated with microcephaly, while the reciprocal duplication is associated with macrocephaly. Systematic overexpression and suppression of all genes in the CNV showed that dosage perturbation of the chromodomain-helicase-DNA-binding protein *CHD1L*, gave significant head size changes: overexpression of human *CHD1L* mRNA led to macrocephaly, while suppression of *chd1l* lead to a significant decrease in head size. However, given that a) human genetic data suggested that *CHD1L* might not be sufficient to explain causality of this CNV, we asked whether *cis*-CNV and *trans*-CNV genetic interactions might contribute to pathogenesis. To answer the first question, we systematically tested the pairwise interaction of *CHD1L* with each of the other 8 genes within the CNV, though which we identified several epistatic partners that modulate the expressivity of micro- and macrocephaly. For the second question, we generated RNA-seq data from heads of control, *CHD1L* RNA-, and *chd1l* MO-injected embryos; we identified ~700 differentially expressed transcripts; strikingly, the top 20 genes in this cohort have been implicated in the genetic causality of autism. Epistatic analysis of candidates whose dosage is perturbed reciprocally by the CNV showed strong complementation for a subset of genes, indicating that their transcriptional misregulation is likely relevant to the observed neurodevelopmental phenotypes. The combination of CNV dissection with *in vivo* genetic interaction studies highlights the genetic complexity of genomic lesions and informs functional networks relevant to ASD pathomechanism.

585T

Increased relative mitochondrial DNA content in Keratoconus patients. *A. Kondkar[1], K. Abu-Amero[1,3], TA. Azad[1], T. Sultan[1], H. Kalantan[2], A. Al-Muammar[2].* 1) Department of Ophthalmology, College of Medicine, King Saud University, Riyadh, Saudi Arabia; 2) Anterior Segment Unit, Department of Ophthalmology, College of Medicine, King Saud University, Riyadh, Saudi Arabia; 3) Department of Ophthalmology, College of Medicine, University of Florida, Jacksonville, Florida, USA.

Mitochondrial DNA (mtDNA) is extremely prone to oxidative stress. To investigate the possible association of oxidative stress with Keratoconus we estimated the changes in relative mitochondrial DNA (mtDNA) content. The study included119 patients with Keratoconus and 208 controls matched for gender, ethnicity and systemic disease status. We selected controls with higher age group than those of the patients as the mtDNA copy number tend to increase with age. The age mean (SD) was 26.4 (7.6) and 54.5 (14.4) y for patients and controls respectively. Relative mtDNA copy number was estimated by real-time quantitative polymerase chain reaction (qPCR) method using ND1 as a mtDNA gene and human globulin (HGB; also known as cytoglobin gene - CYGB) as the reference single-copy nuclear gene. The mean relative mtDNA content was found to be significantly higher in patients with Keratoconus (1.20 copies) than the normal control subjects (1.04 copies) [mean difference = 0.164 copies, 95% confidence interval (CI) = 0.07 to 0.25; P = 0.0004]. Subjects with high mtDNA content (>1.259 copies i.e, greater than 75th percentile) were found to be at an increased risk of the disease (odds ratio = 2.62, 95% CI = 1.40 to 4.89; P = 0.0025). High mtDNA content appears to be associated with increased risk of Keratoconus. Increased mtDNA content in Keratoconus patients may imply a response to oxidative stress, possibly in part because of mitochondrial respiratory chain defects.

586S

Structural haplotypes of the human amylase locus and their relationship to obesity. *C.L. Usher[1], R.E. Handsaker[1,2], T. Esko[1,2,3], J.E. Moon[1,2], A. Metspalu[3], J.N. Hirschhorn[1,2], S.A. McCarroll[1,2].* 1) Genetics, Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Estonian Genome Center, University of Tartu, Tartu, Estonia.

The three human amylase genes (*AMY2B*, *AMY2A*, and *AMY1*) encode enzymes that digest starch into sugar. These genes reside in a complex genomic locus characterized by inversions, deletions, and tandem duplications. Only a handful of the structural haplotypes are known, yet there is evidence of other, unknown haplotypes segregating in populations.

We first identified these structural haplotypes, as well as their population frequencies and evolutionary history, by tracking their transmission in a hundred father-mother-offspring trios using droplet digital PCR (ddPCR) and sequencing read-depth analysis. We found 10 structural haplotypes, each with different combinations and copy numbers of *AMY2B*, *AMY2A*, and *AMY1*. These structures help define a long history of recurrent mutation and provide clues for more accurate copy number genotyping at this locus.

The presence of higher *AMY1* copy numbers in populations with high starch diets has led to speculation that *AMY1* copy number contributes to obesity. Supporting this idea, a recent study reported that *AMY1* copy number has the largest effect yet observed for common variation on obesity. To investigate the amylase locus's effect on BMI, we used the advanced methods that we had already developed for amylase to genotype copy number for all amylase genes in 500 Estonians from each of the BMI extremes (BMI < 22 and BMI > 33). This design had >99% power to detect effects one-tenth as strong as those reported in the recent study. As expected, we observed strong associations between obesity and SNPs within previously implicated loci, such as *FTO*, which confirms that this sample had sufficient power to detect genetic associations. However, we found no association between obesity and the copy number of any amylase gene.

587M

Constitutional chromothripsis: A novel phenomenon in congenital disorders. *A. Alharini[1], M. Pronold[2], A. Slavotinek[1,3], J. Shieh[1,3].* 1) Department of Pediatrics, Division of Medical Genetics, UCSF, San Francisco, CA; 2) Department of Pathology and Laboratory Medicine, UCSF, San Francisco, CA; 3) Institute of Human Genetics, UCSF, San Francisco, CA.

Chromothripsis (chromosomal shattering) is characterized by complex multiple genomic rearrangements recently observed in cancer genomes (Stephens et al., 2011). It is rarely seen in constitutional studies, and the clinical consequences and implications of these genomic alterations are poorly defined. We investigated whether chromothripsis was observed in patients without known malignancies. We analyzed their clinical course. Out of more than 1800 arrays at our institution run from 2009 until present, two patients with chromothripsis were identified. Patient 1 is an 18 month old term male, a product of an in vitro fertilization (IVF) via intracytoplasmic sperm injection (ICSI) due to low sperm count and motility in the father. He presented with hypotonia and global developmental delay and dysmorphic features including microcephaly, hypertelorism, epicanthal folds, downturned mouth, inverted nipples, bilateral congenital ptosis. Oligonucleotide-SNP array demonstrated eight chromosomal changes, one copy loss and seven copy gains on chromosomes 1, 2, 7, 12, 13 and 14. Six out of those eight genomic changes were de novo. Patient 2 is a 12 month old term male with hypotonia, developmental delay and dysmorphic features including microcephaly, anomalous right ear, unilateral radioulnar synostosis, limb length discrepancy, and digital differences. Microarray revealed two deletions and ten duplications of the chromosome 4. We conclude that chromothripsis can involve multiple chromosomes simultaneously or it can have a predilection to certain chromosomes like chromosome 4 (Kloosterman et al., 2011). With these above cases of chromothripsis present in patients with congenital disorders, it will be important to recognize this phenomenon and its implications for the patients. Further studies are needed to provide the mechanism of chromothripsis and the subsequent clinical approach and management.

588T

End-point zygosity and CNV determination from crude samples. *C. Liu, S. Dong.* MBS R & D, Life Technologies, south san francisco, CA.

End-point zygosity and CNV determination based on post PCR Rn/ΔRn reading is expected to reduce the cost and increase throughput as compared to current real time Cq-based measurement. There is an immediate need by AgBio customers to screen seed zygosity and copy number of transgenes through end-point reading. Moreover, Agbio customers expect to determine zygosity and CNV using crude samples in order to stream-line their workflow. For end-point quantification, the major challenge is to control the saturation of PCR to maintain the segregation of end-point intensity according to the input copy number of the target transgenes in reference to a control gene. We have tried a few approaches such as Asymmetric PCR, ARCS (Amplification Ratio Control System) PCR, Dual Tailing PCR, pyrophosphate removal and controlling plateau of PCR by running a third PCR in the background. Among the tested approaches, controlled plateau of PCR (CoP'ed PCR) showed best separation of copy number through end-point PCR reading. We further optimized CoP'ed PCR conditions using purified DNA and crude plant and blood samples. By CoP'ed PCR, we are able to do zygosity with crude plant samples to satisfy the immediate needs for Agbio customers. Furthermore, since this approach improves the sensitivity for the copy number separation for not only end-point but also real-time PCR for crude blood samples, in the future, it can be used for CNV analysis directly from human blood, such as prenatal chromosome copy number variation detection since the percentage of fetal DNA in the maternal blood sample is only 10%-15%. It is impossible to detect the copy number abnormality through traditional real time or end-point PCR.

589S

Sequencing the genomes of single cells. *P. Ribaux, C. Borel, F. Santoni, E. Falconnet, S.E. Antonarakis.* Dept Genetic Medicine, Univ Geneva Medical School, Geneva, Switzerland.

Whole-genome amplification and next-generation sequencing advances enable investigation of somatic structural and nucleotide variation to single-cell resolution. The ultimate goals of our study are (i) to identify disease-associated somatic mutations and (ii) to uncover the extent of low-abundance DNA variations in individual cancer cells in order to underlie mechanisms of tumor evolution. Because of the technical challenge of detecting and analyzing genomic heterogeneity among single cells, we first analyzed individual cells in culture and tested the robustness of our experimental workflow. We choose the K562 cells, a human immortalized myelogenous leukemia line and F-T21, a human primary Trisomy 21 fibroblast cell line. We used the C1 Single Cell Auto Prep System (Fluidigm) to capture hundreds of individual cells and to generate high quality of individual amplified DNA. So far, 96 barcoded whole-exome libraires were sequenced at deep coverage (PE, 100bp). Variant calls (CNVs and SNVs) were generated with an in-house analysis pipeline. Here, we will discuss the amplification uniformity, the detectable fraction of the exome and the level of DNA contamination. By comparing single cells and bulk of cells datasets, we will assess the percentage of allelic drop out for each each single-cell exome based on the heterozygous SNVs. High quality single-cell genome sequence will greatly enhance the genetic analysis of somatic genomic disorders. C.B. and P.R. contributed equally.

590M

Deletions of Regulatory Boundaries are Associated with Congenital Disease. *M. Spielmann[1,2], J. Ibn Salem[1,2,3], S. Kohler[1], M.I. Love[2,4], H.R. Chung[4,5], N. Huang[6], M.E. Hurles[6], M. Haendel[7], N.L. Washington[8], D. Smedley[6], C.J. Mungall[8], S.E. Lewis[8], C.E. Ott[1], S. Bauer[1], P. Schoeld[9,10], S. Mundlos[1,2], P.N. Robinson[1,2,3,4].* 1) Institute for Medical and Human Genetics, Charite-Universitätsmedizin Berlin, Berlin, Germany; 2) Max Planck Institute for Molecular Genetics, Ihnestr. 63, 14195 Berlin, Germany; 3) Department of Mathematics and Computer Science, Free University Berlin, Takustr. 9, 14195 Berlin, Germany; 4) International Max Planck Research School for Computational Biology and Scientic Computing, Ihnestr. 6373, 14195 Berlin, Germany; 5) Otto Warburg Laboratory, Ihnestr. 63{73, 14195 Berlin, Germany; 6) Wellcome Trust Sanger Institute, CB10 1SA Hinxton, UK; 7) Oregon Health & Science University, Department of Medical Informatics & Clinical Epidemiology, 7239 Portland, OR, USA; 8) Lawrence Berkeley National Laboratory, Mail Stop 84R0171, 94720 Berkeley, CA, USA; 9) The Jackson Laboratory, 04609 Bar Harbor, ME, USA; 10) University at Cambridge, Department of Physiology, Development and Neuroscience, Downing Street, CB2 3EG Cambridge, UK.

Recent data from genome-wide chromosome confirmation capture analysis (Hi-C) indicate that the human genome is divided into conserved megabase-sized local chromatin self-interacting regions called topological domains. These topological domains form the regulatory backbone of the genome and are separated by regulatory boundary elements or barriers. Copy-number variations (CNVs) can potentially alter the topological domain architecture by deleting or duplicating the barriers and thereby allowing enhancers from neighboring domains to ectopically activate genes causing misexpression and disease, a mutational mechanism that has recently been termed “enhancer adoption". In this study the Human Phenotype Ontology (HPO) database was used to relate the phenotypes of 922 deletion cases recorded in the DECIPHER database to monogenic diseases associated with genes in or adjacent to the deletions. We sought combinations of tissue-specific enhancers and genes adjacent to the deletion and associated with phenotypes in corresponding tissue, whereby the phenotype matched that observed in the deletion. We compared this computationally with a gene-dosage pathomechanism that attempts to explain the deletion phenotype based on haploinsufficiency of genes located within the deletions. Up to 11.8% of the deletions could be best explained by enhancer adoption or a combination of enhancer adoption and gene-dosage effects. Our results suggest that enhancer adoption caused by deletions of regulatory boundaries may contribute to a substantial minority of CNV phenotypes and should thus be taken into account for their medical interpretation.

591T

Sensitive and efficient analysis of somatic mosaicism using genomewide SNP arrays and haplotypes. *S. Vattathil[1,2], L. Huang[2], P. Scheet[2,1].* 1) Human and Molecular Genetics Program, University of Texas at Houston Graduate School of Biomedical Sciences, Houston, TX; 2) Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX.

More comprehensive characterization of somatic mosaicism allows better inference of background mutation rates and mechanisms and will help define the relevance of somatic mutation to human phenotype. Among the challenges to practical survey of somatic mutations is that detection methods must be at once sensitive (since mutations may exist in only a small fraction of sampled cells), efficient (since the rate of mosaicism is low and many samples must be analyzed), and agnostic (since mutation locations are not known *a priori*). hapLOH, a computational algorithm we developed, provides sensitive, efficient, and agnostic profiling of somatic chromosomal imbalance using best-guess haplotype estimates and whole-genome SNP array data. We used hapLOH to characterize mosaicism in 31,223 individuals from 10 studies conducted as part of the GENEVA Consortium. We called 1,134 mosaic mutations (ranging in size from 42 Kb to 146 Mb [all of chromosome 8]) in 895 individuals, which is a substantially higher rate than the original published analysis of these data (Laurie *et al.*, Nat. Gen., 2012). The results validate recent predictions that the extent of somatic mosaicism has been underestimated due to lack of sensitivity for low-frequency mutations, and confirm the value of using haplotype information for sensitive detection of allelic imbalance. We describe the locations of recurrent somatic mutations, which are likely to contain genes important for proliferation, and copy number classifications for events with sufficient log R ratio evidence. Since hapLOH is sensitive to very low-frequency events, more than half of the observed calls cannot be classified even though the signal of imbalance is strong. Sample-specific empirical null simulations and comparisons of observed B allele frequency and log R ratio shifts to predicted values suggest the rate of false positives is low. Incidentally, we found that depressed phase concordance can distinguish inherited from somatic mutations.

592S

Monozygotic Twin Pairs: CNV and sequence concordance. *A. Abdellaoui[1], E. Ehli[2], J.J. Hottenga[1], Z. Weber[2], H. Mbarek[1], G. Willemsen[1], T. van Beijsterveldt[1], A. Brooks[3], J.J. Hudziak[4], P.F. Sullivan[5], E.C.J. de Geus[1], K. Ye[6], P.E. Slagboom[7], G.E. Davies[2], D.I. Boomsma[1].* 1) Biological Psychology, VU University Amsterdam, Amsterdam, Noord Holland, Netherlands; 2) Avera Institute for Human Genetics, Avera McKennan Hospital & University Health Center, Sioux Falls, SD, USA; 3) Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA; 4) University of Vermont, College of Medicine, Burlington, VT, USA; 5) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA; 6) The Genome Institute, Washington University School of Medicine, St. Louis, Missouri, USA; 7) Molecular Epidemiology, Leiden University Medical Center, Leiden, Netherlands.

Monozygotic (MZ) twins are genetically identical at conception, making them informative subjects for studies on somatic mutations. Copy number variants (CNV) are responsible for a substantial part of genetic variation, have relatively high mutation rates, and have been associated with susceptibility to disease, such as autism and schizophrenia. We conducted a genome-wide survey for post-twinning de novo CNVs (i.e., not shared by co-twins) in ~1,100 MZ twin pairs who had been repeatedly phenotyped across a wide range of traits, and of which a large proportion has gene-expression and methylation data available. CNVs from 1,097 MZ twin pairs were measured in DNA from peripheral blood samples (mostly in adults) or buccal epithelium (mostly in children) with the Affymetrix 6.0 microarray. Whole-genome sequencing was performed in DNA from blood samples from 13 MZ twin pairs and their parents (12x coverage - Illumina, and 2 twin pairs additionally sequenced with Complete Genomics). We found a total of 153 putative post-twinning de novo CNVs >100 kb, of which the majority resided in the same unstable genomic region (15q11.2). Based on how well the raw intensity signals visually agreed with CNV calls made by the two algorithms, a first selection was made of eleven de novo CNVs from 15q11.2 for a first series of qPCR validation experiments. Two out of eleven post-twinning de novo CNVs were validated with qPCR in the same twin pair. This 13-year old twin pair did not show large phenotypic differences. The remaining putative de novo CNVs from 15q11.2 were found significantly more often in older twins, suggesting that we are capturing real signals. The large putative de novo CNVs detected with microarray data were not present in the subsample that had whole-genome sequence data available. We do expect the whole-genome sequence data to allow us to search for smaller de novo CNVs that cannot be detected with micro-array data.

593M

Even and odd: defining the structure of variation at the human amylase CNVs. *J.A.L. Armour[1], D. Carpenter[1], S. Dhar[1], L.M. Mitchell[1], B. Fu[2], J. Tyson[1], N. Shwan[1], F. Yang[2], M.G. Thomas[3].* 1) School of Life Sciences, University of Nottingham, Nottingham, UK; 2) Wellcome Trust Sanger Institute, Hinxton, UK; 3) Research Department of Genetics, Evolution and Environment, University College London, UK.

The human amylase genes are a long-established example of a human gene CNV, and may have undergone adaptive changes in human evolution. Genotyping amylase variation accurately has been problematic because of both the wide range of salivary (*AMY1*) CNV (with copy numbers commonly in the range 2–14), and the high sequence similarity between *AMY1* and neighbouring pancreatic amylase (*AMY2A/AMY2B*) genes. Most published work has measured *AMY1* gene copy number using qPCR, and presented this CNV as a quasi-continuous variation independent of *AMY2A/2B*. Using PRT and other methods capable of resolving individual repeat unit differences, we have been able to characterise the amylase CNVs in unprecedented detail, demonstrating that in all populations studied most amylase haplotypes contain an odd number of copies of *AMY1*, and therefore that most people have an even diploid total. These conclusions match those defined by Groot *et al.* 25 years ago using segregation analysis of Southern blot hybridization patterns, and are supported by new fibre-FISH analysis. We have now also defined less common haplotypes containing even numbers of *AMY1*, which are associated with CNVs of pancreatic amylase (*AMY2A/2B*) genes. We have defined four different series of variant haplotypes containing even numbers of *AMY1* in combination with *AMY2A/2B* deletions or duplications. These haplotypes show significant differences between populations studied, with about 20% of Europeans carrying an odd number of copies of *AMY1* and a deletion or duplication of *AMY2A*. Our measurements, including fibre-FISH data, also suggest error in the qPCR calibration of data presented by Perry et al, resulting in the systematic underestimation of *AMY1* copy number in their study claiming natural selection for high copy numbers in populations with traditionally high-starch diets. Calibration derived from Perry *et al.* was adopted by the recent BMI association study of Falchi *et al.*, suggesting that the copy number determination in that study may also have been subject to a similar systematic error. Overall, our conclusions would not be accessible to those measurement methods (such as qPCR) incapable of accurately resolving individual repeat unit differences at high copy number.

594T

Genome-wide analysis of copy number variants and their association with kidney transplantation outcomes. *L. Bassaganyas[1,2], PY. Kwok[1,2].* 1) Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA.

Kidney transplantation is the treatment of choice for end-stage renal failure in terms of patient survival and quality of life. However, acute rejection (AR) and chronic allograft nephropathy (CAN) remain obstacles to post-transplant health in most of patients. The success/failure of kidney transplantation mainly depends on inflammatory and immune factors that might be influenced by genetics, but the specific association between gene variants and graft survival/rejection is still poorly understood. Copy number variants (CNVs) affect both normal phenotypic variation and disease and are mostly enriched by genes involved in inflammation and immune response. We examined the potential influence that immune-related genes within CNV regions might have in different transplant outcomes. Using DNA microarray data and the PennCNV-Affy algorithm, we obtained the CNV profiles for 536 (270 donors and 266 recipients) patients previously grouped in three different outcomes: normal working transplant (TX), AR and CAN. The comparison between CNVs detected for each group revealed a higher number of gains in CFHR3 and CFHR1 genes among TX samples and a higher frequency of losses of these genes in AR subset, suggesting a role of the complement system regulation in the graft survival. Moreover, we found more CNVs affecting KIR genes in AR recipients and CAN donors in comparison with TX samples, implying that specific genomic architecture of the polymorphic KIR locus might have an influence to the renal transplant outcome. This study constitutes the first report on the risk association between CNVs and the kidney transplantation outcomes.

595S

Inverted repeats mediate complex genomic rearrangements including quadruplication. *C.R. Beck[1], C.M.B. Carvalho[1,2], L. Banser[3], T. Gambin[1], D. Stubbolo[3], B. Yuan[1], K. Sperle[3], S.M. McCahan[3,4], M. Henneke[5], P. Seeman[6], G.M. Hobson[3,4,7], J.R. Lupski[1,8,9].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Centro de Pesquisas Rene Rachou- FIOCRUZ, Belo Horizonte, MG, Brazil; 3) Nemours Biomedical Research, Alfred I. duPont Hospital for Children, Wilmington, DE; 4) Jefferson Medical College, Thomas Jefferson University, Philadelphia, PA; 5) University Medical Center Göttingen, Georg August University, Department of Pediatrics and Adolescent Medicine, Division of Pediatric Neurology, Germany; 6) Department of Pediatric Neurology, 2nd Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic; 7) University of Delaware, Department of Biological Sciences, Newark, DE; 8) Department of Pediatrics and Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 9) Texas Children's Hospital, Houston, TX.

Inverted repeats (IRs) predispose loci to inversions and can lead to copy number alterations by mediating complex genomic rearrangements (CGRs). Clustering of breakpoints in cases of duplication and triplication of the dosage sensitive *PLP1* gene occurs within three sets of distal (telomeric), nested inverted repeats. These CGRs result in Pelizaeus-Merzbacher disease (PMD, MIM #312080), and approximately 80% of PMD individuals have amplification of *PLP1*. We systematically investigated the region encompassing *PLP1* in both unaffected and affected individuals to identify the IRs responsible for the CGRs and the mechanism of rearrangement. In phenotypically normal individuals, we identified a copy number neutral polymorphism between a set of IRs in the cluster. Southern blotting determined that the inversion is prevalent in the population, with a frequency of approximately 40%. Furthermore, integrating these data with sequence information for the locus showed that inversion alleles were present on two haplotypes in the region. Therefore, there is a common polymorphic inversion at this locus that appears to be recurrent. We have assembled a cohort of 17 Individuals affected with PMD who have breakpoints clustering within the distal IRs. The 17 patients all have CGRs, and duplication-inverted triplication-duplication rearrangements were observed in 16 individuals. Through array comparative genomic hybridization, Southern blotting, quantitative PCR, and cloning methodologies, we determined that the IRs responsible for the inversion are implicated in these rearrangements. Breakpoint junction analysis has underscored the presence of point mutations close to CGRs that are replicative in origin and the presence of additional template switches at 5/17 or ~30% of these events. Finally, in one patient, a CGR resulting in duplication of *PLP1* and quadruplication of a proximal genomic region was identified. In this patient, junction sequencing revealed that the mechanism of formation is consistent with rolling-circle amplification; a mechanism predicted by the microhomology-mediated break-induced replication (MMBIR) model. Thus, detailed analysis of the *PLP1* locus has provided experimental evidence for another mechanism by which inverted repeats affect genomic change.

596M

Deciphering the complex effect of the 16p11.2 duplication using large family-based cohorts. *W.K. Chung[1], D. D'Angelo[2], S. Lebon[3], Q. Chen[2], L. Hippolyte[4], E. Hanson[5], A. Maillard[4], E.H. Sherr[6], R. Bernier[7], W.A. Faucett[8], R.P. Goin-Kochel[9], A. Mace[10,11], A. Wallace[7], E. Grant[12], K. Mannik[13], S. Martin[10], I. Caldeira[4], L.A. Green Snyder[14], D.H. Ledbetter[8], C.L. Martin[8], P. Mukherjee[15], M.B. Ramocki[16], S.J. Spence[17], K.J. Steinman[18], J. Tjernagel[19], J.S. Beckmann[4,10], A. Reymond[13], J.E. Spiro[19], S. Jacquemont[4], Simons VIP Consortium.* 1) Division of Molecular Genetics, Columbia University, New York, NY; 2) Mailman School of Public Health, Columbia University, New York, NY; 3) Dept. of Pediatrics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 4) Service de Genetique Medicale, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 5) Dept. of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, MA; 6) Dept. of Neurology, University of California, San Francisco, San Francisco, CA; 7) Dept. of Psychiatry and Behavioral Science, University of Washington, Seattle, WA; 8) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 9) Dept. of Pediatrics, Psychology Section, Baylor College of Medicine, Houston, TX; 10) Dept. of Medical Genetics, University of Lausanne, Lausanne, Switzerland; 11) Swiss Institute of Bioinformatics, University of Lausanne, Lausanne, Switzerland; 12) Dept. of Radiology, Boston Children's Hospital, Harvard Medical School, Boston, MA; 13) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 14) Boston Children's Hospital, Boston, MA; 15) Dept. of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA; 16) Dept. of Pediatrics, Section of Pediatric Neurology and Developmental Neuroscience, Baylor College of Medicine, Houston, TX; 17) Dept. of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA; 18) Dept. of Neurology, Seattle Children's Research Institute & University of Washington, Seattle, WA; 19) Simons Foundation, New York, NY.

The 600kb 16p11.2 duplication and deletion (BP4-BP5) are among the most frequent copy number variants, and the duplication is associated with neurodevelopmental disorders, including autism spectrum disorder, schizophrenia, and a decrease in head circumference and BMI. The 16p11.2 locus, like duplications of other genomic regions, shows highly variable expressivity and incomplete penetrance. To decipher the complex effects of the 16p11.2 duplication on cognitive, behavioral, medical, and anthropometric traits, we studied a cohort of 270 duplication carriers as well as a group of 184 deletion carriers from the 16p11.2 European and Simons Variation in Individuals Project (Simons VIP) consortia. We used linear mixed models to estimate the effect of the duplication, compared to non-carrier family members, on cognitive and anthropometric traits, while accounting for age, gender, head circumference (HC), consortium, and family effects. Global cognition is significantly impacted by the duplication. The largest effect is observed in probands with an average decrease in full scale IQ (FSIQ) of 21 points relative to non-carrier family members after adjusting for covariates. In carriers ascertained through cascade family testing, this decrease is 12 points in pediatric non-proband carriers and 9 points in adult carriers. The large FSIQ variance is a striking feature of the duplication with increased proportions at both extremes of the IQ distribution when compared to the 16p11.2 deletion group. Paradoxically, while the mean FSIQ in duplication carriers is higher than that of the deletion carriers (79.1 vs. 77.4), the frequency of very low FSIQ (<40) is increased 15-fold in the duplication compared to the deletion group (7.7% vs 0.5%). ASD was diagnosed in 13.7% of the duplication carriers. Additional DSM-IV diagnoses were present in 41% of the carriers. The duplication was associated with a 0.5 and 1 standard deviation decrease in HC and BMI, respectively, compared to their non-carrier family members. The duplication is associated with a greater range in neurocognitive phenotype when compared to the reciprocal deletion. Although carriers of second deleterious CNVs were excluded from this analysis, additional genetic variants may be contributing to the phenotype of the cognitively severely impaired duplication carriers.

597T

Population genetics and mutation analysis of an exceptionally copy number variable sperm gene. *A. Davis[1,2], S.A. McCarroll[1,2].* 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) The Broad Institute of Harvard and MIT, Cambridge, MA.

Though multiallelic copy number variants (mCNVs) cover thousands of genic and intergenic regions of the human genome, little is known about their population distributions and even less about their mutability. It is commonly assumed that mCNVs mutate with high frequency, but their mutational events and other population genetic properties have been hard to observe, as mCNVs have been difficult to analyze molecularly. We have investigated these properties for *SPANXB* (sperm protein associated with the nucleus on the X chromosome family member B), a sperm protein-encoding gene that is extremely copy number variable in humans. We developed molecular assays that yield crisp integer measurements of *SPANXB* copy number. We observe from one to ten copies of the 12kb *SPANXB*-containing duplication per chromosome, and see striking differences in copy number distribution between African, Asian, and European populations (HapMap/1000 Genomes YRI, CHS, CEU, and IBS). This copy number variation is not well captured by SNPs, as the highest correlation (r^2) between a SNP and SPANXB copy number is 0.37. In analysis of phased haplotypes containing SNP and CNV alleles, we find that individual copy number alleles are captured with r^2 ranging from 0.006 to 0.61. We also determined that copy number of *SPANXB* is correlated with its RNA expression in sperm by examining DNA and RNA from sperm samples from 30 individuals. Few new copy number mutations have been observed for highly multiallelic CNVs such as *SPANXB*, though the presence of many copy number alleles in human populations suggests that they have mutated many times among human ancestors. Due to the presence of many segregating copy number alleles, new mutations are difficult to observe using the canonical method of examining offspring for differences in genotype from their parents. We have leveraged genotype and sequence data to identify pairs of men with recent shared ancestry in the genomic region surrounding *SPANXB* and determined *SPANXB* copy number in these individuals. Among 50 pairs with recent shared ancestry drawn from a cohort of 7079 males, we observe three copy number mutation events at *SPANXB*, all in pairs of men whose shared ancestry at the locus is at least seven generations old. As we expand our approach to larger and different cohorts, we anticipate building more complete models of the mutational properties of this and other mCNV loci.

598S

Functional Effects of Copy Number Variant Junctions. *K. Dumas[1], J.T. Shieh[1,2].* 1) Division of Medical Genetics, Department of Pediatrics, University of California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA.

Copy number variation can cause genic interruptions and novel transcripts. Investigating the functional impact of these events may yield insight into an important source of human genetic variation. The goal of this investigation is to study the impact of rare, disruptive germline CNVs on protein behavior. Rare copy number variants in 13 patients of the UCSF study of congenital anomalies were identified using Affymetrics SNP 6.0 genotyping arrays. After stringent CNV calling, predicted breakpoints within genes were identified and validated by sequencing CNV boundaries. Endogenous and truncated transcripts, representing wild type and disrupted alleles respectively, were cloned and eGFP fusion constructs generated for protein studies. Expression studies in 293T cells, imaging, and western blot analysis were performed to determine protein production, stability and secretion in culture. We found that duplications have the potential to create novel, altered transcripts. Sequencing and RT-PCR analysis revealed that one such CNV breakpoint led to transcript truncation but successful RNA expression in primary patient tissue. As full length and truncated (69% of full length amino acid content) products were possible, both products were investigated. Protein made from a full length construct in transfected cells was expressed normally and secreted, whereas protein made from the truncated construct was produced but altered in expression and behavior. These data demonstrate rare CNVs can create novel transcripts that alter the protein repertoire. Further investigation should be directed at understanding the role of disruptive CNVs in protein function.

599M

Identifying Population-Specific Structural Variation in Human Blood Group Genes. *K. Fox[1], T. Shaffer[1], D. Crosslin[1], M. Delaney[2,3], J. Johnsen[2,4], D. Nickerson[1].* 1) School of Medicine, Genome Sciences, University of Washington, Seattle, WA, USA; 2) Puget Sound Blood Center, Research Institute, Seattle WA, USA; 3) University of Washington, Department of Laboratory Medicine, Seattle, WA, USA; 4) University of Washington, Department of Medicine, Seattle, WA, USA.

Blood type determination *via* hemagglutination is the clinical gold standard. Assigning units of blood based on serological typing can sometimes fail to identify less common blood type variants, risking exposure of the recipient to clinically significant allo-antigens. Blood types reflect genetic variation at blood group genes, which harbor low frequency variants encoding hundreds of blood types not detected by conventional blood typing. High throughput sequencing approaches, such as exome sequencing, have the potential to assign high resolution blood types based on DNA sequence data. Rh (*RHD, RHCE*) and MNS (*GYPA, GYPB, GYPE*) blood group genes contain known, common structural variants (SVs) that result from unequal crossing-over or gene conversion events. However, the influence of SV characterization on blood typing for cross-matching has yet to be fully explored.

Allosensitization is most common in chronically transfused patients, particularly in sickle cell disease (SCD) which affects 1 in 500 African-Americans. SCD patients are exposed to multiple blood donors over their lifetimes, and are at high risk (>20%) of developing significant alloantibodies. The National Heart, Lung, and Blood Institute (NHLBI) funded the Exome Sequencing Project (NHLBI-ESP), and, through this effort, subjects of African-American (n=1,714) and European ancestry (n=3,405) with heart-, lung-, and blood-related phenotypes were exome sequenced. Using these data, we present read-depth-based SV calls, including 70 unique duplications and 66 unique deletions in six human blood group genes (*ABO, RHD, RHCE, GYPA, GYPB, GYPE*) from 5,119 exomes with >20X coverage. We confirmed known deletions (Rh D-, MNS GPB null) and have identified novel SV in the ABO blood group gene (deletion of *ABO* exons 5–7, predicting blood type O). These analyses highlight the need for further studies not only to characterize the diversity of SV in these genes, but also to explore the phenotypic effects of such variation in the context of transfusion medicine.

600T

Characterization of Copy Number Variation and Loss of Heterozygosity Using high resolution SNP MicroArray -The Miami Experience. *G. Ghaffari[1], B.J. Ilagan[1], C. Hung[1], S.A. Hosseini[1], P. Benke[2], R. Yusupov[2], L. Brenton[1], M. Rodriguez[1], B. Johnson[1], O.A. Bodamer[1].* 1) Division of Clinical and Translational Genetics, Dr. John T. Macdonald Foundation, Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, FL 33136, USA; 2) Memorial Regional Hospital, 1150 N 35th AVE # 490, Hollywood, FL., 33021, USA.

Background: MicroArray analysis is considered the first-tier clinical diagnostic test for individuals with unexplained developmental delay/intellectual disability (DD/ID), autism spectrum disorders (ASD), and/or multiple congenital anomalies (MCA). In addition to clarification of chromosome deletions and duplications; Microarray provides information regarding long contiguous stretches of homozygosity (LCSH). Here we report the results of MicroArray analysis from patients with DD, ASD, and MCA referred to the Clinical Molecular Genetic Laboratory (CMGDL), at the University of Miami from June 2012 to May 2014. Methods: The Illumina HumanOmni1-Quad or CytoSNP-850K BeadChip (Illumina,CA) array platforms were used to determine the presence of Copy Number Variations (CNVs) and or Loss of Heterozygosity (LOH) from 150 . CNV was reported based on gene content and phenotype previously mapped to the called regions. LOH was reported based on the size and gene content related to the reported clinical phenotype. ACMG-recommended terminology was used to report the results. Results: Pathogenic copy number variations were detected in 19/150 (12.6%) patients ranging from 77Kb to whole chromosome gain or loss. Pathogenic CNVs were detected in the following chromosomes: chr. 16 in 5 patients (16p11.2;16p12.2; 16q23.1q23.2); X chr.in 5 patients, including a Mosaic Turner syndrome; chr. 1 in 3 patients (1q31.3;1q43q44;1q41q44); chr. 22 in 2 patients (22q11.22 q11.23; 22q11.21); chr. 5 (5q13.2); chr. 6 (6q27); chr. 18 (18p11.32q23). LOH was detected in 18/150 (12.0%) patients ranging from 6 Mb to 600 Mb. 83 (55.3%) patients had no CNVs or LOH spanning any genes or genes known to be associated with disease. Their results were reported as normal. In 30 (20%) patients we detected CNVs of uncertain significance. Conclusion: These results demonstrate the utility of SNP genotyping data for the detection of clinically significant abnormalities including CNVs, mosaism or LOH which may harbor recessive mutations. Interpretation of LCSH and CNVs remains challenging particularly without detailed clinical information.

601S

The role of the genomic architecture of transposable elements in the formation of copy number variants: evidence from one schizophrenia family. *G. Guffanti[1], S. Gaudi[2], P. DeCrescenzo[1], H. Mangalam[3], A. Rodriguez[4], R. Madduri[4], C. Pato[5], F. Macciardi[6].* 1) Department of Psychiatry Division of Epidemiology & Division of Child and Adolescent Psychiatry Columbia University/NYSPI New York, NY; 2) Istituto Superiore di Sanita', Rome, Italy; 3) OIT, University of California, Irvine, Irvine, CA; 4) MCS Argonne National Laboratory Computation Institute, University of Chicago; 5) Department of Psychiatry and the Behavioral Sciences Keck School of Medicine, USC, Los Angeles, CA; 6) Dept of Psychiatry & Human Behavior, University of California, Irvine, Irvine, CA.

Many CNVs are flanked by transposable elements (TE). It is emerging that the genomic architecture of TE might predispose certain regions to the mutational mechanisms that result in CNVs. The breakpoints of CNVs are frequently observed within the repetitive sequences of TEs, including SINE, LINE, and LTR, which suggested that regions of the genome that possess repeat sequences are more prone to rearrangements arising independently in different individuals. We sought to explore the role of TEs flanking de novo and inherited CNVs in a family of four siblings, three schizophrenia probands and one unaffected sibling, descending from a schizophrenia affected mother and an unaffected father. This family is part of the Genomic Psychiatric Cohort (GPC) sample (Pato et al, 2013). We used ForestSV to assess CNVs from whole genome sequences (~40x). CNV deletions were included in the analysis if the prediction confidence score was >0.7, which keep the error rate below 5%. First, we determined pattern of transmission of CNV deletions from the affected mother and the unaffected father to three probands and sibling; second, we estimated the prevalence of reference TEs (hg19) in 1,000 bp regions upstream and downstream the breakpoints of de novo and inherited CNVs. The number of CNVs per individual ranged from 2,020 to 2,131 unique deletions. We discovered a total of 3,780 inherited autosomal deletions, including 657 transmitted to probands and 101 to the unaffected sibling, respectively. 575 deletions were found in one or more siblings but not in the parents and were classified as de novo. Of these, 327 de novo deletions were unique to the probands, 61 to the sibling and 191 were shared by siblings, both affected and unaffected. Structurally, more than 60% of de novo deletions were flanked with Alus, 50% with LINEs and 65% with LTR at either the 5' or 3'. Our survey of regions at the breakpoints of CNVs deletions confirm the pattern of genomic instability introduced by the repeat sequences of TE and support their role in the mechanisms that lead to copy number variants formation.

602M

De novo Germline Variants from WGS of Autism Spectrum Disorder Trios. *M. Gujral[1], W. Brandler[1], D. Malhotra[1], J. Estabillo[1], T. Gadomski[1], A. Watts[1], T. Chapman[1], D. Antaki[1], T. Solomon[1], A. Moyzis[1], A. Bhandari[1], L. Wong[1], C. Corsello[1], N. Akshoomoff[1], L. Iakoucheva[1], E. Courchesne[2], J. Sebat[1].* 1) Psychiatry, University of California, San Diego, La Jolla, CA., USA; 2) Neuroscience, University of California, San Diego, La Jolla, CA., USA.

Autism spectrum disorder is a genetic anomaly with diverse footprints. Based on the recent clinical estimates, it impacts more than 1% of the children around the globe. In last few years, there has been considerable effort to delineate the genetic basis of the disorder through genome sequencing projects. Despite finding several rare de novo genetic variants from exome and whole genome sequence studies, many rare variants remain to be discovered. Here, we report rare de novo genetic variants through the whole genome sequencing of 161 samples comprising of 47 probands, 19 healthy siblings, their 46 parent-pairs, and a healthy family trio contributed by total of 47 families. All de novo single nucleotide variants (SNVs) and copy number variations (CNVs) were detected through the machine learning approaches applied to the processed data for 67 trios. Our findings include three novel de novo nonsense mutations, five de novo deletions and one duplication, all six CNVs are > 10 kb. These CNVs, five de novo deletions and a single duplication, have been validated through Affy6 and Illumina array platforms. The three nonsense mutations observed in probands are likely pathogenic. All the four de novo deletions observed in probands are rare and encompass exons. Largest de novo SV in our samples is ~500 kb duplication observed in a proband spanning through several genes. Taken together, it is clear that the burden of de novo nonsense SNVs and CNVs is heavily skewed towards the probands. We have been able to determine the parent of origin for about 30% of total de novo SNVs and one forth of the SNVs are contributed by mother.

603T

Multi-allelic copy number variation humans. *R. Handsaker[1,2,3], V. Van Doren[2], L. Boettger[2], S. McCarroll[1,2,3].* 1) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 2) Department of Genetics, Harvard Medical School, Boston MA; 3) Stanley Center for Psychiatric Research, Cambridge, MA.

Copy number variation (CNV) is widespread in human populations, affecting thousands of genes and functional elements. Progress has been made in recent years in understanding simple, di-allelic CNVs that arise by one-step deletion or duplication. A substantial fraction of all inherited CNVs in humans, however, is contributed by variants that appear in different genomes at high frequency and in widely different numbers of copies (from 2 up to 12 or more copies). The inability to molecularly type multi-allelic CNV (mCNV), due to the wide range of copy numbers present in different genomes and the difficulty of molecularly discriminating high copy numbers, have excluded mCNV from most studies of human genetic and phenotypic variation. The structural alleles and haplotypes that segregate at such loci and their relationship to phenotypes remain largely unknown.

To address these questions, we first developed novel ways to use increasingly abundant whole genome sequencing data to identify mCNV loci and the copy number alleles that segregate at these loci. We applied these methods to 849 genomes sequenced in Phase 1 of the 1000 Genomes Project at low coverage and to hundreds of other deeply sequenced genomes, creating a locus, allele and haplotype map for human duplication CNV. We typed 3,926 duplication CNVs, of which 1,432 appear to exhibit three or more alleles. We also developed new molecular-biological assays to evaluate genotype accuracy at these loci. The discrete integer copy number determinations from these assays are highly (99%) concordant with our computational inferences from sequencing data, even at high copy numbers (4–9), suggesting that both methods are highly accurate.

We evaluated the impact of CNV on gene dosage and find that 90% of the impact on gene-dosage variation arises from about 100–200 highly polymorphic mCNVs. Using published RNAseq data, we observe that differences in gene dosage almost always lead to corresponding changes in gene expression. We describe a phenomenon of “runaway” duplication haplotypes, on which specific genes have mutated to high copy number on specific haplotypes in specific human populations during the past 50,000 years. We developed methods and data resources to statistically impute mCNV alleles from SNP haplotypes, and describe both successes and limitations in imputing these genetic variants, as a means to relate them to phenotypes in large association studies.

604S

Comparative performance analysis of high-resolution, genome-wide array platforms for copy number variation detection. *R.R. Haraksingh[1], A. Abyzov[2], A.E. Urban[1].* 1) Psychiatry, Stanford University, Palo Alto, CA; 2) Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN.

Copy Number Variation (CNV) is a major class of genomic variation in humans. Many CNVs are benign and constitute normal variation while others have pathogenic effects. Detecting CNVs genome-wide at high resolution and determining their nature (i.e. deletions or duplications) is important for both clinical diagnostics and biomedical research. At present, sequencing-based CNV detection cannot affordably compete with array-based methods. Clinical cytogenetics relies heavily on arrays for detecting pathogenic CNVs. Array-based genome-wide CNV detection platforms are also used to efficiently and cost effectively study genome structure in large cohorts and to track genome structural stability for quality control in cell cultures, particularly in stem cell model systems. Array-based CNV calls also provide orthogonal validation of sequencing-based calls. In this study, we characterize the relative performances of all currently available commercial high-density oligonucleotide arrays for genome-wide CNV detection. This follows up on our earlier study [PMID: 22140474] using a previous generation of arrays. Here, we collected CNV data from 15 different array designs: two Affymetrix, four Agilent, and nine Illumina designs. The extensively characterized HapMap/1000 Genomes sample, NA12878, was hybridized to each array in two technical replicates. Two sets of CNVs were then called from each array. One set was called by the company-recommended software. The other was called using Nexus Copy Number software by Biodiscovery as a uniform process for all the platforms. Each call set was then compared to a gold standard derived from the 1000 Genomes Project sequencing data. The gold standard contains only very high confidence CNVs found by sequencing in NA12878, and supported by multiple analytic principles (e.g. paired-end, split-read, or read-depth analysis), experimental validation (e.g. PCR), or both. We determined the overall sensitivities of the platforms for genome-wide CNV calling, and whether any of the platforms perform better at discovering particular subtypes of CNVs such as those in a specific size range or genomic context, or of a specific nature. Generally, arrays with more probes and those targeted to known CNV regions while maintaining sufficient genome-wide coverage show the highest number of CNV calls and sensitivities, most accurate breakpoint resolution, and largest size range of CNV calls. A breakdown of results for each platform will be presented.

605M

Discovering Copy Number Variations in ClinSeq®: A Large-Scale Whole Exome Sequencing Study. *C.S Hong[1], D. Ng[1], L.N. Singh[1], N.F. Hansen[2,3], J.C. Mullikin[2,3], L.G. Biesecker[1,2].* 1) Medical Genomics and Metabolics Genetics Branch, National Human genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 2) NIH Intramural Sequencing Center, National Human genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 3) Comparative Genomics and Cancer Genetics Branch, National Human genome Research Institute, National Institutes of Health, Bethesda, MD.

Copy number variations (CNVs) can alter gene expressions and cause pathological conditions. Using CNV prediction tools, we comprehensively examined exome sequences from the ClinSeq® project for CNVS using publicly available software: the XHMM software suite. ClinSeq® is a clinical sequencing project composed of 971 individuals from the metro D.C. area who are enrolled for whole exome sequencing. For each individual, more than 100 clinically relevant phenotypes and family history are measured and recorded. The exome sequences were analyzed for copy number variations (CNVs) using publicly available software: XHMM software suite. A total of 12,606 CNVs were (1695 deletions and 7910 duplications) called in 971 samples. The median size of all calls was 12,960 bp. Real-time PCR validations of putative CNVs in CYP2D6 of 26 samples were all verified at 100% confirmation rate, suggesting highly sensitive predictions. Functional enrichment analysis by DAVID showed that the list of non-singleton genes in putative CNV regions was enriched for Ca2+ dependent cell-cell adhesion, keratin/intermediate filament, and olfaction. The most prevalent CNV occurred in TYRO3, with 188 individuals (19% of ClinSeq® population) predicted to have a CNV in TYRO3. This result is in agreement with the high prevalence of a polymorphic retrocopy insertion of TYRO3 into an intron of the ENOX1 gene on chr13. We also observed evidence of a fusion gene, TFG-GPR128. Our findings demonstrate the value of CNV and structural variant analysis from exome sequence in clinical settings.

606T

Optimization of a High-Throughput, Cost-effective, Reliable Method for Large-Scale Screening of CNVs at Candidate Loci using qPCR. *C. Kao, R. Pellegrino, J. Garifallou, T. Guettouche, H. Hakonarson.* Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

Copy number variations (CNVs) are known to contribute to phenotypic diversity and can be causative factors for many types of human disorders and diseases. Array-based methods are the most common and reliable means for detecting CNVs and providing estimates of CNV length and boundaries. However, these array-based technologies are not necessarily cost-effective for assessing small numbers of candidate loci and require turnaround times of weeks to months for custom-designed arrays to be produced, which becomes impractical for particular applications. We were interested in screening for rare CNVs across five known pathogenic loci (22q11.21, 16p11.2, 15q13.3, 2p16.3, and 2p25.3) across the biobank collection at the Center for Applied Genomics, which contains samples approaching 200,000 in total. Herein, we describe the development of a cost-effective workflow to screen candidate loci for CNVs via quantitative, real-time PCR in a high-throughput format. We find that particular "master mix" reagents perform more robustly than others in terms of stability/shelf-life, ability to withstand variations in DNA quality, and uniformity across different primer/probe sets. We also establish quality control (QC) metrics based on amplification curve parameters to "weed out" false-positives arising from poor sample quality and developed a robust clean-up protocol for problematic samples that fail QC. The sensitivity and selectivity of this qPCR workflow is compared in ~1,000 samples where CNV calling based on an orthogonal approach (i.e. SNP-based whole-genome arrays) was also done, and the performance of both approaches showed remarkable concordance. Thus, our approach provides a cost-effective means for screening for candidate CNVs across large (multi-thousand) sample sets in a rapid, high-throughput manner with an unprecedented level of rigor and consistency that previously had not been achieved using standard qPCR-based workflows.

607S

SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints. *H. Kehrer-Sawatzki[1], J. Vogt[1], K. Bengesser[1], S. Bammert[1], K.B.E. Claes[2], K. Wimmer[3], V.F. Maunter[4], R. van Minkelen[5], E. Legius[6], H. Brems[6], M. Upadhyaya[7], C. Lazaro[8], J. Högel[1], T. Rosenbaum[9], L. Messiaen[10], D.N. Cooper[7].* 1) Human Genetics, University of Ulm, Ulm, Germany; 2) Centre for Medical Genetics, Ghent University Hospital, Belgium; 3) Division of Human Genetics, Medical University Innsbruck, Austria; 4) Department of Neurology, University Hospital Hamburg Eppendorf, Germany; 5) 5 Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 6) Department of Human Genetics, KU Leuven, Belgium; 7) Institute of Medical Genetics, School of Medicine, Cardiff University, UK; 8) Molecular Diagnostics Unit, Hereditary Cancer Program, Catalan Institute of Oncology (ICO-IDIBELL), L'Hospitalet de Llobregat, E-08908 Barcelona, Spain; 9) Department of Pediatrics, Duisburg General Hospital, Germany; 10) Medical Genomics Laboratory, Department of Genetics, University of Alabama at Birmingham, Alabama, USA.

Background: Genomic disorders are caused by copy number changes that may exhibit recurrent breakpoints processed by nonallelic homologous recombination. However, region-specific disease-associated copy number changes have also been observed which exhibit non-recurrent breakpoints. The mechanisms underlying these non-recurrent copy number changes have not yet been fully elucidated. Results: We analyze large NF1 deletions with non-recurrent breakpoints as a model to investigate the full spectrum of causative mechanisms, and observe that they are mediated by various DNA double strand break repair mechanisms, as well as aberrant replication. Further, two of the 17 NF1 deletions with non-recurrent breakpoints, identified in unrelated patients, occur in association with the concomitant insertion of SINE/variable number of tandem repeats/Alu (SVA) retrotransposons at the deletion breakpoints. The respective breakpoints are refractory to analysis by standard breakpoint-spanning PCRs and are only identified by means of optimized PCR protocols designed to amplify across GC-rich sequences. The SVA elements are integrated within SUZ12P intron 8 in both patients, and were mediated by Target-Primed Reverse Transcription of SVA mRNA intermediates derived from retrotranspositionally-active source elements. Both SVA insertions occurred during early postzygotic development and are uniquely associated with large deletions of 1 Mb and 867 kb, respectively, at the insertion sites. Conclusions: Since active SVA elements are abundant in the human genome and the retrotranspositional activity of many SVA source elements is high, SVA insertion-associated large genomic deletions encompassing many hundreds of kilobases could constitute a novel and as yet under-appreciated mechanism underlying large-scale copy number changes in the human genome.

608M

Genome-wide association study of copy number variation with hematological traits using family-based Samples. *B. Kim[1], S. Moon[1], M. Liu[2], Y. Kim[1], M. Hwang[1], Y. Kim[1], R. Elston[3], B. Han[1], S. Won[2].* 1) Korea National Institute of Health, Chungwon-gun, South Korea; 2) Chung-Ang University, Seoul, South Korea; 3) Western Reserve University, Cleveland, OH.

Copy number variants (CNVs) are widely distributed throughout the human genome (Feuk et al., 2006; Sharp et al., 2006) and have been considered as important genetic factors for human diseases (Lupski, 2008; McCarroll and Altshuler, 2007). It is known to play an important role in the genetics of complex diseases. In this study, we performed genome-wide CNV association analyses of hematological traits on 522 Korean family samples to infer a genetic architecture underlying quantitative traits. Statistical methods for CNV association analysis can be categorized into two different strategies. First, the observed probe intensity measurement can be directly used to detect association of CNV with the phenotypes of interest. Second, the most probable copy number is estimated by maximum likelihood and association of the most probable copy number with the phenotype is tested. Genome-wide CNV association study identified a region significantly associated with mean corpuscular hemoglobin concentration (MCHC) (P = 2.5710-5). CNV genotypes at this region were validated with a PCR experiment and the same association result was replicated in 4694 additional independent samples (P = 7.0010-5). Conclusions: This finding may contribute to alleviate the issue of copy number uncertainty in the CNV association analysis with hematological traits.

609T

Comprehensive analysis of large structural variants in well-characterized human genomes. *E. Lam[1], A. Hastie[1], A. Mak[2], Y. Lai[2], H. Cao[3], D. Cao[3], W. Andrews[1], H. Dai[1], M. Austin[1], F. Trintchouk[1], M. Saghbini[1], T. Anantharaman[1], K. Haden[1], X. Xu[3], P.-Y. Kwok[2], H. Cao[1].* 1) BioNano Genomics, San Diego, CA; 2) University of California, San Francisco, San Francisco, CA; 3) BGI-Shenzhen, Shenzhen, China.

Genome mapping in nanochannel arrays (BioNano Genomics) represents a new single-molecule platform complementary to short-read sequencing for genome assembly and structural variation analysis. Extremely long molecules of hundreds of kilobases fluorescently labeled at sequence motifs and elongated in nanochannels enable direct interrogation of genome structure at a high resolution.

The high throughput of the BioNano Irys system has made possible, for the first time, rapid analysis of multiple genomes and cross-sample comparison to identify genome structural variation at high resolution. To date, we have de novo assembled more than 20 normal and diseased human genomes and analyzed their structural variation content. Our genome map assemblies cover at least 90% of non-N-base portions of the genome and also extend into subcentromeric and subtelomeric regions of the genome.

Here, we present results from extensive analysis of an Asian genome and a CEPH trio. We detected hundreds of large structural variants per genome and haplotype differences in these genomes. In the YH genome, we found 708 insertions/deletions and 17 inversions larger than 1 kb. Without considering 59 SVs that overlap with N-base gaps in hg19, 609 out of 666 (90%) are supported by orthogonal experimental methods (resequencing- and/or fosmid assembly-based) or historical evidence in public databases. For the CEPH trio, we identified novel and previously reported structural variants consistent with Mendelian inheritance. We also used publically available sequence read data to confirm and refine our SV calls.

Overall, our genome map assemblies provide valuable structural information otherwise difficult or impossible to decipher with short-read sequencing data alone.

610S

Estimating the parental haplotype source of germline-transmitted de novo duplications. *Y. Liu[1,2], S. Vattathil[2,3], L. Huang[2], X. Xiao[4], G.E. Davies[5], E.A. Ehli[5], J.J. Hottenga[6], A. Abdellaoui[6], I. Ruczinski[7], S. Arur[9], D. Boomsma[6], T.H. Beaty[8], P. Scheet[1,2,3].* 1) Program in Biostatistics Bioinformatics and Systems Biology, The University of Texas Graduate School of Biomedical Science, Houston, TX, USA; 2) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; 3) Human & Molecular Genetics Program, The University of Texas Graduate School of Biomedical Sciences, Houston, TX, USA;; 4) Center for Genomic Medicine and Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Hanover, NH, USA; 5) Avera Institute for Human Genetics. Sioux Falls, SD, USA; 6) Biological Psychology, VU University, Amsterdam, The Netherlands; 7) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; 8) Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; 9) Department of Genetics, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA.

De novo germline-transmitted duplications arise during gamete formation (possibly driven by meiotic recombination) and represent the genesis of a genetic polymorphism that has been associated with both complex and Mendelian forms of human disease. Details of their formation, however, are not well understood. For example, there are two possible sources for the extra copy of DNA in the offspring (de novo duplication); the parent in which the duplication arose (1) transmitted one copy from each homologous chromosome, or (2) transmitted two identical copies from a single chromosome. Here we present a new computational approach to interpret the source of DNA in parental chromosomes contributing to a de novo duplication using population genetic data from family-based genome-wide association (GWA) studies. Specifically, we first use trios to identify de novo duplications in offspring. We then check for consistency among plausible combinations of transmitted haplotypes and observed SNP array data (genotypes implied by the "B allele" frequencies) for the offspring (duplication carrier), under the two scenarios described above. To examine this phenomenon we leverage two large family-based studies. In a preliminary analysis of one of these, we identified 61 de novo duplications in 2,078 independent trios. Application of our method to these events indicates a tendency for duplications to derive from the same parental chromosome (situation 2 above). We are currently attempting to strengthen our conclusions with additional analyses of monozygotic twins from the other study and by ruling out sources of artifact, such as somatic mutation.

611M

Copy number variants identified in Japanese women. *O. Migita[1,2], K. Maehara[2], K. Nakabayashi[2], K. Okamura[3], K. Hata[2].* 1) Department of Pediatrics, St. Marianna University School of Medicine, Kanagawa, Japan; 2) Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan; 3) Department of Systems BioMedicine, National Research Institute for Child Health and Development, Tokyo, Japan.

Using a high-resolution SNP array, we examined copy number variations (CNVs) found within 411 healthy Japanese women. With increasing public concern about infertility and the frequent involvement of chromosomal anomalies in miscarriage, analyses of CNVs have been used to identify the genomic regions responsible for each process of childbearing. Although associations between CNVs and phenotype have been reported, previous studies accumulated unsusceptible CNVs with insufficient phenotypic information on pregnancies. It will be necessary to collect large number of control data that focused on normal parity. We collected samples from Japanese women who have experienced normal delivery without significant complications and have compiled 1043 copy number variable regions. The copy number differences in these regions may be irrelevant not only to infertility but also to a wide range of diseases. The utility of this resource are assumed to be useful for reducing the number of candidate pathogenetic variants, especially in Japanese subjects. Because our identification strategy was based on a microarray technique, it is inevitable that cross-hybridization would have occurred. We carefully curated output data for each CNVR and validated 9 regions out of 12 homozygous deleted regions by conventional PCRs and 10 copy variable regions by the NanoString analysis system. We found that many implausible calls were situated in regions with high GC contents; for example, in subtelomeric regions. Those were mostly gain-type CNVs rather than loss-type CNVs. Development of more sophisticated methods and algorithm to determine CNV regions are awaited.

612T

Rare large CNVs are associated with intellectual disability, education level, and female fertility in general population. *K. Männik[1,2], R. Mägi[3], A. Mace[4,5], A. Maillard[5], H. Alavere[3], A. Kolk[3,6], L. Leitsalu[3], A.M. Ferreira[1], M. Noukas[2,3], J.S. Beckmann[4], S. Jacquemont[5], Z. Kutalik[4,7], A. Metspalu[3], A. Reymond[1].* 1) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 2) Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 3) Estonian Genome Centre, University of Tartu, Tartu, Estonia; 4) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 5) Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; 6) Department of Neurology and Neurorehabilitation, Children's Clinic, Tartu University Hospital, Tartu, Estonia; 7) Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland.

The impact of large rare CNVs has almost exclusively been evaluated using clinical cohorts and it is unclear how these variants affect health in general populations. To investigate the burden of rare large CNVs in the general population, we analyzed the Estonian Genome Centre cohort. It is a longitudinal, prospective, population biobank encompassing 5% of the Estonian adult population and linked to comprehensive personal, educational, medical and daily life data. Within a subset of 7877 individuals, we identified 65 carriers of known genomic disorder lesions, equivalent to a prevalence of 0.8% in the general adult population. Their phenotypes are reminiscent of those described for carriers of identical rearrangements identified in disease cohorts. Importantly some of the associated traits appear to have been previously overlooked due to age-dependent penetrance. We then generated the genome-wide map of rare autosomal CNVs and identified ~20% of the screened population (n=1567) as carriers of CNVs ≥125kb with frequency ≤0.05%. When compared to the population, carriers of deletions ≥250kb and duplications ≥1Mb show a decrease in education achievement (a proxy for intelligence), an increase in the prevalence of intellectual disability and an alteration of females' fecundity. In addition to CNV size, these effects are associated with the number and "quality" of contained genes. They are pronounced in carriers of deletions and duplications that encompass ≥2 and ≥11 genes, respectively. Concordantly, we found that the cumulative set of protein-coding genes encompassed by rare deletions is significantly enriched for GO processes involved in neurodevelopment, behavior, learning, memory and cognition (13 out of top-20 most significant processes). Stratification of deletions by presence of at least a neurodevelopmental gene within the rearrangement embedded genes or high haploinsufficiency score, but not presence of a ohnolog, correlated with ID prevalence. Our results suggest that rare CNVs account for a substantial portion of the population variance of educational attainment. Carriers of syndromic genomic lesions or of non-recurrent rare CNVs identified in reference cohorts should not readily be considered as healthy carriers. Rare large CNVs represent a significant public health issue as they impact life quality on multiple levels including medical and socioeconomic metrics.

613S

The impact of smaller CNVs and inherited gene-disruptive SNVs in sporadic autism. *T. Turner[1], N. Krumm[1], S.S.C. Sequencing Consortium[2], B.P. Coe[1], A.N. Raja[1], E.E. Eichler[1,3].* 1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA; 2) Simons Foundation Autism Research Initiative, New York, NY; 3) Howard Hughes Medical Institute, Seattle, WA.

The Simons Simplex Collection (SSC) consists of ~2,500 sporadic autism families with one affected child. We analyzed 9,483 exomes (2,377 families) for a complete spectrum of genetic variation, including *de novo* and rare inherited single nucleotide variants (SNVs) and copy number variants (CNVs). To maximize sensitivity, we re-mapped all data using BWA-mem and applied multiple methods for each class of mutation, including CoNIFER and XHMM for CNVs and GATK HaplotypeCaller and FreeBayes for SNVs. We generated targeted genotypes across all exome-derived CNV calls using SNP-probe copy number estimates (via CRLMM) in order to validate CNV events as small as 1 kb (median size = 30kb) and increasing our yield by 31%. Our comprehensive scan enables the assessment of small CNV events (both *de novo* and rare) and rare inherited SNVs, neither of which have been previously assayed in autism spectrum disorder (ASD) families. To assess the relative burden of small CNVs, we focused on 1,200 families with data available for parents, affected proband, and unaffected sibling. *De novo* events showed an overall 2.1 proband/sibling burden for rare *de novo* events ($p = 1.03\times10^{-3}$), and CNVs were larger in probands (median = 93 kb) than in siblings (median=51 kb) ($p = 9.6\times10^{-4}$). For rare, inherited genic CNVs (median size = 30kb), we detected a 1.13 proband/sibling burden with ~50% of samples containing such an event, and found this burden was specific to proband-sibling pairs with strongly discordant phenotypes (via SRS score). We analyzed inherited SNVs and found a similar significant excess (OR = 1.14, $p < 0.0002$) of private inherited truncating mutations in highly conserved genes (RVIS score < 50); furthermore, this burden increased in the most conserved genes (reaching OR = 1.4 for the 1% most conserved). Finally, we combined all types of mutations into a comprehensive map and identified a set of genes with multiple *de novo* or inherited CNVs and SNVs, including *TNRC6B*, *CHD2*, *CSMD1*, and *DSCAM*. These data strongly support an inherited risk component for gene-disruptive events even among cases of sporadic autism.

614M

Somatic Copy Number Variation at Birth. *A. Valind[1], C. Haikal[1], M.EK. Klasson[1], M.C. Johansson[2], M. Soller[1], B. Baldetorp[2], D. Gisselsson[1].* 1) Department of Clinical Genetics, Lund University, Lund, Sweden; 2) Department of Oncology, Lund University, Lund, Sweden.

A dogma in genetics for the last century has been that all cells in an individual's body are genetically identical, with the exception of germ cells and some cells in the immune system. Recent technological advances have made it possible to measure variation in a genome-wide fashion, giving scientists the possibility to measure genomic variation between different organs in the same individual. This development, while still in its infancy, has forced scientists to reconsider the dogma of genetic identicalness of somatic cells within an individual. We have used high resolution SNP-arrays to quantify somatic variation at the segmental level, and fluorescence in situ hybridization (FISH) to quantify variation at the chromosome level in a cohort of stillborn fetuses and neonates. This revealed two organ-specific types of somatic variation. In the thymus, a recurrent genomic signature of T-cell receptor rearrangement was detected, reinforcing the view that most T-cells undergo TCR gamma/delta rearrangements early in development. In the liver, we found an intercellular variation in chromosome copy number that was higher than in other organs, indicating that the high-grade aneuploidy seen in livers of adults is manifest already in utero. In contrast to reports from adult liver, this occurred in the absence of polyploidy. We also found that that the overall level of organ specific copy number variation was considerably lower in fetuses and neonates compared to adults, providing additional evidence for the accumulation of copy number variants over time also in a non-neoplastic context.

615T

Complex structural variant in a family with Autism Spectrum Disorder discovered by whole genome sequencing. *S. Walker, R.K.C. Yuen, K. Tammimies, M.J Gazzellone, B. Thiruvahindrapuram, T. Nalpathamkalam, D. Merico, S.W. Scherer.* Genetics and Genome Biology, Hospital for Sick Children, Toronto, M5G 0A4, Ontario, Canada.

Autism Spectrum Disorder (ASD) is a highly heterogeneous neurological condition with many contributing risk genes and loci. Genome wide microarray and sequencing studies have implicated numerous genes, for example members of the NRXN and SHANK families, *PTCHD1*, *CHD8*, and *SCN2A* as risk candidates in ASD. Mutations in voltage gated sodium channel *SCN2A*, many of which *de novo* in origin, have previously been associated with several forms of epilepsy. In recent whole exome and genome sequencing data in ASD, *SCN2A* has emerged as one of only a small number of genes with multiple independent *de novo* loss of function mutations, making it one of the best new candidate genes. Through our own ASD whole genome sequencing project, we have uncovered a complex structural variant at *SCN2A* in a family quartet comprised of parents and two affected female offspring. Initially, a 1.7kb single exon (exon 18) deletion was discovered as a *de novo* variant in both children thought to have arisen by germline mosaicism. However, further analysis revealed that the deletion breakpoint can also be detected in a blood DNA sample from the father, despite his having two copies of exon 18. Using detailed molecular analyses, we have shown that in the father there is a complex rearrangement such that he carries one intact allele and in addition to the transmitted deletion, he also possesses a third structure carrying a neighbouring intronic deletion. We are pursuing a number of different technical approaches to resolve the nature of this variant and the mechanism by which it has arisen. We are also assessing mRNA transcripts from all family members to investigate the functional impact of the variant structures. Resolving this complex variant will aid in the interpretation of the exonic deletion in the affected individuals in this family and also demonstrates the importance of detailed genomic data in understanding disease risk variants. Our data have further implications for studying models of genetic disease as other similar complex events could give rise to cryptically inherited variants misinterpreted as *de novo* events. Furthermore, this example highlights the power of whole genome sequencing to reveal variants not detectable by other technologies.

616S

Genome-wide copy number variants analysis identifies deletion variants associated with ankylosing spondylitis. *S. Yim[1], S. Jung[1], T. Kim[2], S. Shim[3], Y. Chung[1].* 1) The Catholic University of Korea, College of medicine, Seoul, South Korea; 2) Hanyang University Hospital for Rheumatic Diseases, Seoul, Korea; 3) Chungnam National University Hospital, Daejeon, South Korea.

We aimed to discover ankylosing spondylitis (AS)-associated copy number variants (CNVs) in Korean subjects and their synergistic roles in the development of AS. We performed genome-wide association study (GWAS) in 309 AS patients and 309 controls using Agilent CNV microarray. AS-associated CNV regions (CNVRs) were replicated in two independent sets (625 cases and 891 controls in total) by qPCR and deletion-typing PCR. In the CNV-GWAS, we identified 227 CNVRs significantly associated with AS. Of the candidate CNVRs, nine were successfully replicated in the first replication; 1q32.2 (HHAT), 1p34.2 (BMP8A), 2q31.2 (PRKRA), 6p21.32 (HLA-DPB1), 11q22.1 (CNTN5), 13q13.1 (EEF1DP3), 14q24.2 (RGS6), 16p13.3 and 22q11.1 (IL17RA). Five deletions of them in 1q32.2, 2q31.2, 6p21.32, 13q13.1 and 16p13.3 were risk-increasing CNVRs and the other four were protective. In the second replication, four CNVRs in 1q32.2, 2q31.2, 6p21.32, and 16p13.3 were replicated. Subjects with CNVRs in four or more risk-increasing loci had 18.0 times higher risk than those without any deletions (OR=17.98, P=2.3×10−7). Subjects with CNVRs in two or more protective loci had 5.2 times lower risk than those without any deletions (OR=0.19, P=4.0×10−10). The additive effects of simultaneous events showed a CNVR frequency-dependent manner. Through deletion-typing PCR and sequencing, the exact sizes and breakpoint sequences were defined in three CNVRs. The mechanism of all three deletions were found to be microhomology-based nonhomologous end joining. Our results can help to identify the pathogenic mechanisms of AS and can easily be applied for developing the prediction algorithm of AS development.

617M

Correction of artefacts in estimation of rare copy number variants for analyses of burden and association in type 1 diabetes. *N.J. Cooper, C. Shtir, J.A. Todd.* Diabetes and Inflammation Laboratory, University of Cambridge, United Kingdom.

Copy number variations (CNVs) have been proposed as a possible source of 'missing heritability' in complex human diseases. Several studies of type 1 diabetes (T1D) have found null associations with common copy number polymorphisms (CNPs), but CNVs of low frequency and high penetrance could still play a role. We used the Log-R-Ratio (LRR) intensity data from a dense single nucleotide polymorphism (SNP) array, ImmunoChip, to detect rare CNV deletions (rDELs) and rare CNV duplications (rDUPs) in 6,808 T1D cases, 9,954 controls, and 2,206 families with T1D-affected offspring. Our initial analyses using a standard approach with PennCNV software were affected by biases that led to detection of CNVs subsequently confirmed by quantitative polymerase chain reaction (qPCR) to be false positives. We developed a series of quality control (QC) tests that were demonstrated to identify and correct probe intensity biases. Systematic testing and validation showed that the ImmunoChip-wide case-control odds ratio (OR) for rDELs and rDUPs was highly sensitive to QC thresholds and procedures. Analysis using an optimal QC pipeline, with the highest validated accuracy, suggested no overall CNV burden for T1D, but some association for rDELs longer than 500 kb (OR=1.67, p=.003), noting that ImmunoChip genome coverage is mostly concentrated in known autoimmune regions. No specific CNV regions (CNVRs) showed association at a corrected p-value threshold, although CNVR frequencies were lower than expected (most less than 0.1%), significantly reducing statistical power. Transmission disequilibrium tests (TDT) run on the family dataset provided no support for CNVRs contributing to familial clustering of T1D. We present an R-package, plumbCNV, that provides an automated approach for QC and detection of rare CNVs, which can facilitate equivalent analyses of large-scale SNP array datasets in any disease. A power analysis indicating sample sizes required to obtain convincing evidence for single locus tests is also provided.

618T

The structural architecture of SNP effects on complex traits. *L. Davis, E. Gamazon, N. Cox.* Section Genetic Medicine, The University of Chicago, Chicago, IL.

Despite this year marking the tenth anniversary of the discovery of copy number variation (CNV), current SNP based analysis methodologies continue to collapse the homozygous (A/A), hemizygous (A/0), and duplicative (A/A/A) genotype states treating the genotype variable as irreducible or unaltered by other co-occurring forms of genetic (e.g., structural) variation. Our understanding of common, genome-wide CNV suggests the canonical genotype construct may belie the enormous complexity of the genome. Here we present multiple analyses of several phenotypes and provide novel methods supporting a conceptual shift that embraces the structural dimension of genotype. We developed a method we call “copy number integrated GWAS” (cni-GWAS), which properly accounts for allelic dosage, and applied it to the detection of expression quantitative trait loci, in LCLs, finding that 20% of expression-associated SNPs (eSNPs) increased in significance by at least nearly 3.3 orders of magnitude. We investigated their functional consequences and found enrichment for enhancer elements and DNase hypersensitivity sites at transcription factor binding sites with greater binding affinity than expected (p=0.03). We then developed a method to calculate the heritability attributable to these “unmasked” eSNPs in multiple phenotypes. Second, to determine the impact of CNV and copy number stability on rare coding variants likely to increase disease risk, we assessed the overlap between CNV regions and loss of function (LOF) variants from the 1000 Genomes project, and from published exome sequencing in autism. We found that regardless of phenotype or de novo status, approximately 80% of all LOF variants fall within structurally variable regions of the genome; furthermore, a much larger proportion of LOF variants in copy number stable (CNS) regions trigger nonsense-mediated decay (p=0.016). Lastly, we determined the aggregate phenotypic variance accounted for by SNPs distributed across CNV or CNS regions of the genome in Tourette Syndrome and obsessive-compulsive disorder. We find very different results across these two neuropsychiatric phenotypes and argue that dosage sensitivity is a genomic property, which may elucidate genetic architecture. Taken together, these results argue for the inclusion of a structural dimension and suggest that some portion of the “missing” heritability may be recovered through integration of the structural dimension of SNP effects on complex traits.

619S

Genetic variation in introns that flank alternatively spliced exons: A new way to look for disease-related variants. *A. Neininger[1], W.C.L Stewart[1,2,3], D.A. Greenberg[1,2].* 1) Battelle Center for Mathematical Medicine, Nationwide Children's Hospital, Columbus, OH., 700 Children's Drive; 2) The Ohio State Department of Pediatrics, Columbus, OH., 700 Children's Drive; 3) The Ohio State Department of Statistics, Columbus, OH., 1958 Neil Ave.

Simple exonic mutations are not the only kinds of genetic variants that can lead to inherited disease. Intronic variation is also important, but the relatively high degree of variability in introns makes the identification of specific disease-related alleles difficult. For example, the gene BRD2 is related to epilepsy susceptibility and it contains no exonic disease-related mutations, but certain sequences of an intron are associated with disease. This intron is highly variable and it also contains an alternatively spliced (AS) exon. This led us to ask: Are introns surrounding AS exons more variable than other introns? AS exons are found in approximately 75%-90% of genes, and 1/3 of these exons can introduce a premature termination codon (PTC). Therefore, understanding the variability that regulates alternative splicing could allow us to identify intronic sequences that are related to disease. Using sequence data from 1000 Genomes, we first compared the frequency of common variants in the introns flanking AS and regularly spliced exons. We found statistically significant differences between introns that flank AS exons compared to those that flank regularly spliced exons. This finding suggests that this variation influences regulation of splicing. We also subdivided the introns flanking AS exons into two groups: those that introduce a PTC leading to nonsense-mediated decay, and those that do not. We then explored a method that evaluates different intronic sequences on the basis of physical similarity and free energy. Our preliminary results from the analysis of 46 epilepsy patient haplotypes and 48 control haplotypes suggest that the free energy of the hypervariable BRD2 intron that flanks the AS exon appeared able to predict which sequences exist in the population and which do not. Moreover, sequence similarity alone does not appear to be associated with disease. We did not observe any statistical differences between these two groups. In conclusion, approaches that account for differences in the physical gene structure may give us a better way to understand intronic sequence variation, and consequently, a better way to identify disease related “alleles”.

620M

Expression of human acidic mammalian chitinase in *Escherichia coli* and analysis of its properties. *K. Okawa, A. Kashimura, Y. Kobayashi, M. Ohno, M. Sakaguchi, Y. Sugahara, F. Oyama.* Applied Chemistr, Faculty of Engineering, Kogakuin University, Hachioji, tokyo, Japan.

Acidic mammalian chitinases (AMCase) have been shown to be closely associated with asthma in mouse model, allergic inflammation and food processing. AMCase is expressed in mice and humans. The amino acid sequence of human AMCase shares 81% identity with that of mouse counterpart. There are several variants with the amino acid substitutions in human AMCase. Recent genetic association analyses of AMCase haplotypes for asthma revealed significant associations between the variant haplotype in several asthma cohorts. However it still remains to elucidate the pathophysiological significance of human AMCase *in vivo*. To study the roles of AMCase, we expressed human AMCase as a fusion protein A and V5-His in *E. coli* and compared its chitinolytic activity with that of mouse counterpart. The chitinolytic activity of human AMCase was about 1/50 of that in the mouse counterpart. These results indicate that the specific activity of human AMCase is notably low when compared with mouse AMCase.

621T

Beyond the 1000 Genomes Project. *L. Clarke, H. Zheng-Bradley, A. Datta, I. Streeter, D. Richardson, P. Flicek, The 1000 Genomes Consortium.* Vertebrate Genomics, European Molecular Biology Laboratory - European BioInformatics Institute (EMBL-EBI), The Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom.

The 1000 Genomes Project provides an essential reference catalog of human variation with more than 60 million variant sites ranging from single nucleotide polymorphisms to structural variant events including inversions and duplications. Also provided are global allele frequencies and genotypes for 2535 individuals from 26 different populations across Europe, Africa, East and South Asia and the Americas, which enable many other projects to better interpret their results. Primary uses for the 1000 Genomes data sets include imputation panels to create whole genome variant sets from exome or array-based genotypes; as filters of “normal” or shared variation in rare disease or cancer sequencing projects; and to explore demography and selection in human populations. The 1000 Genomes Project is now drawing to a close. Here we describe plans to maintain the resource in order to ensure it remains the valuable data set it is today by providing long-term support for the 1000 Genomes Project resource. For example, we will continue to host both the FTP site (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp) and the project website (http://www.1000genomes.org) to ensure the community can access both the raw data and the documentation about the project. We will also create a stable version of the 1000 Genomes Browser (http://browser.1000genomes.org) based on the project's final date release. This project specific Ensembl-based browser displays all of the 1000 Genomes variants as soon as possible and will use the GRCh37 assembly of the human reference genome. We will also maintain the existing tools and incorporate new ones as appropriate to enable users to easily access the data they desire. Our most popular tools are the Data Slicer—that allows users to select genomic subsections of our alignment (BAM) and variant (VCF) files and thus download just the piece of the file they need—and the Variation Pattern Finder, which allows users to discover patterns of shared variation in a specific region of the genome. Other tools include the VCF to PED converter, which allows users to generated PLINK format files from remotely hosted VCF files and the recently introduced the Allele Frequency Calculator that will calculate allele frequencies in bulk for specific sub populations from our VCF files.

622S

Empirical and computational survey of functional regulatory genetic variants. *G.A. Moyerbrailean[1], C.T. Harvey[1], C.A. Kalita[1], X. Wen[2], F. Luca[1], R. Pique-Regi[1].* 1) Wayne State University, Detroit, MI; 2) University of Michigan, Ann Arbor, MI.

Genome-wide association studies (GWAS) have been instrumental to identify genetic variants associated with complex traits. However, risk variants tend to fall in non-coding regions and likely affect gene regulatory mechanisms. Even with increasing regulatory annotations provided by ENCODE and others, predicting the impact of a sequence change on gene regulation remains a challenge. Identification and functional characterization of sequence variants in regulatory elements is crucial for understanding the molecular determinants of complex traits. Here, we integrate DNA sequence with DNaseI footprinting data to predict the impact of a sequence change on transcription factor binding. Applying this approach to 653 DNase-seq samples, we identified 3,831,862 regulatory variants predicted to affect active regulatory elements for a panel of 1,891 transcription factor motifs. We observe that these variants are enriched for having a MAF < 1%, are more likely to be in enhancer regions, and tend to affect factors active in < 5 cell-types. Using our newly developed method, QuASAR (Quantitative Allele-Specific Analysis of Reads), we examined the data for variants exhibiting allele-specific binding (ASB). We identified 3,217 binding variants within footprints that are significantly imbalanced (20% FDR). Overall, we estimate that 56% of our predicted regulatory sites in footprints show an ASB signal. To assess the effect these variants may have on complex phenotypes, we examined their association with complex traits and observed that ASB-SNPs are enriched 1.22-fold for complex traits variants (from GWAS). Using results from meta-studies for lipid levels and height, we identified factors whose binding sites were enriched for associated SNPs. Factors for lipid levels include GR and HNF4a, as well as regulators of immune function such as CREB, PU.1, IRF-1 and IRF-2. Factors for height include stem cell and developmental regulation factors such as Oct-4 and Nkx2-5. Additionally, these annotations can be very useful for fine-mapping. For example, because rs532436 (a genetic loci associated with an increase in LDL) is in a footprint for the factor USF, the annotation increases the likelihood from 40% to 70% that this is the functional SNP in the loci. These results show that integration of footprint annotations into GWAS meta-study results improves the identification of likely causal SNPs and provides a putative mechanism by which the phenotype is affected.

623M

The success of whole exome sequencing diagnosis in a large cohort of patients with Mendelian disorders. *T. Roscioli[1,2,3], L. Ewans[1,2], M.J. Cowley[3], K. Ying[3], Y. Zhu[4], C. Walsh[5], E. Lee[5], M. Field[4], A.M. Turner[1], D. Mowat[1], A. Hackett[4], E.P. Kirk[1], R. Sachdev[1], M.L Freckmann[1], M. Lipke[1], M. Buckley[5], M. Dinger[3].* 1) Department of Medical Genetics, Sydney Children's Hospital, Sydney, NSW, Australia; 2) School of Women's and Children's Health, University of New South Wales, Randwick, NSW, Australia; 3) Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, NSW,; 4) Newcastle GOLD Service, Hunter Genetics, Waratah, NSW, Australia; 5) SEALS laboratory, Prince of Wales Hospital, Randwick, NSW, Australia.

Molecular diagnosis in patients with Mendelian disorders has been improved significantly by advances in next generation sequencing (NGS). NGS has resulted in enhanced mutation identification, genetic management and in some cases the creation of novel therapies. Mendelian disorders are most amenable to diagnosis though whole exome sequencing (WES) as the coding 2% of the genome is highly enriched for disease-causing mutations. WES was applied to a cohort of 52 patients from 37 families who were selected from clinical genetics clinics in New South Wales and for whom no molecular diagnosis was known. This cohort is phenotypically heterogeneous, however the majority of patients presented with intellectual disability (62%) consistent with population frequency and diagnostic importance. Other diagnoses included skeletal dysplasias (14%), retinitis pigmentosa (8%), haematological disorders (8%), seizures (3%), metabolic conditions (3%), and dysmorphic syndromes (3%). DNA from annotated exons and splice sites, untranslated regions, and the mitochondrial genome were captured with a Nextera extended exome kit on Illumina HiSeq 2500 sequencers. Genomic data was annotated with the Variant Effect Predictor (VEP) and common variants were excluded based on population polymorphism data and impact on protein function using the GEMINI (Genome MINIng) software. The most likely inheritance model was applied to each family based on pedigree analysis and indication for referral and in consanguineous families, regions of homozygosity were identified and applied as an additional filter. Mutation pathogenicity scoring systems were applied including PROVEAN and CADD scores for missense and other mutations. We present an overview of the testing outcomes including a preliminary cost-effectiveness analysis. Initial results from 13 families show a diagnostic success rate of approximately 25% for mutations in known disease-related genes, and a likely novel disease gene for intellectual disability involved in neurite growth in a consanguineous family.

624T

Identifying tagging SNPs for African specific variation from the African Diaspora Genome. *H.R. Johnston[1], Y. Hu[1], J. Gao[1], G. Abecasis[2], M. Hansen[3], R. Genuario[3], D. Bullis[3], R.A. Mathias[4], K.C. Barnes[4], Z.S. Qin[1], CAAPA Consortium.* 1) Department of Biostatistics and Bioinformatics, Emory University Rollins School of Public Health, Atlanta, GA; 2) University of Michigan School of Public Health, Ann Arbor, MI; 3) Illumina, Inc. San Diego, CA; 4) Department of Medicine, Johns Hopkins University, Baltimore, MD.

The Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) includes high coverage whole genome sequence data (~30x depth) on ~1,000 subjects of African ancestry and extends the patterns of variation catalogued in the Thousand Genomes Project and Exome Sequencing Project to a spectrum of populations representing a wide range of African ancestry in the Americas. An interim data freeze (N=643) of CAAPA includes: 329 African Americans; 125 African Caribbeans; 164 African ancestry samples with a notable Latino component; and 25 samples from Nigeria. One of the primary goals of CAAPA is to develop an 'African Diaspora Power Chip' to address the concern that current commercially available GWAS chips have made a limited effort to tag African specific variation. The goal of the African Diaspora Power Chip project is to utilize sequence variants within CAAPA to build a genotyping chip that tags as much African specific variation as possible. Since many GWAS studies have already been conducted using the Illumina OmniExpress chip, it was chosen to be the primary partner chip for the African Diaspora Power Chip.

47.9 million SNPs were observed in the CAAPA samples. 15.6 million variants with a minor allele frequency ≥1% were included in the set to be tagged. Variants that could be well imputed based on the OmniExpress chip using the 1000 Genomes Phase I African Reference Panel were removed from the selection pool. Fugue, a haplotype construction software package, was used to determine the pairwise LD between each pair of SNPs in the remaining set. These LD estimates were used by FESTA, a software package designed to select the optimal set of tagSNPs, to select SNPs using an r^2 threshold of .8. All tag SNPs with an MAF >= 1.6% were retained. Additional content including SNPs targeting African specific diseases and the HLA region, plus a "GWAS backbone" was added.

Analysis of our coverage of SNPs >= 1% MAF in the CAAPA population was conducted to ensure tagSNP selection success. On Chromosome 22, the Power Chip alone tags 11% of variants at .9 r^2. This increases to 14% at .8 and 28% at .5. In conjunction with the OmniExpress array, the pair of chips tags 26% of the variants at .9 r^2. This increases to 32% at .8 and 50% at .5.

625S

The role of the host genetic variability in the influenza-A virus susceptibility. *A.C. Arcanjo Silva[1,7], G. Mazzocco[2,6], S.F. Oliveira[1,3], D. Plewczynski[2,3,5], J.P. Randomski[2,4].* 1) Programa de Ps Graduao em Biologia Animal, Universidade de Braslia, Braslia, DF 70910-900, Brazil; 2) Interdisciplinary Center for Mathematical and Computational Modeling, Warsaw University, Pawińskiego 5A, Bldg. D, PL-02106 Warsaw, Poland; 3) The Jackson Laboratory for Genomic Medicine, c/o University of Connecticut Health Center; 263 Farmington Avenue, Farmington, CT 06030, USA; 4) Institute of Biotechnology and Antibiotics, Starościńska 5, PL-02516 Warsaw, Poland; 5) Yale University, New Haven, CT, USA; 6) Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland; 7) Batzer Laboratory of Comparative Genomics, 202 Life Sciences Building, Louisiana State University, Baton Rouge, LA 70803, USA.

We have explored genomic variability of the genes responsible for host-pathogen interactions and the host-specific immunological response to influenza-A infection, which represents a serious health threat for a significant part of the world's population. The aftermath of influenza infection is determined by a complex set of host-pathogen interactions, where genomic variability on both sides influences the final outcome. Although a large body of literature describes influenza virus variability, only a small fraction covers the issue of host variance. The goal of this work was to explore the variability of host genes responsible for host-pathogen interactions in three different phases of infection: binding of the virus to a target cells, viral detection mechanisms inside the host cell and immune response mechanisms aimed at virus clearance. Seven to ten genes were selected in each of the phases of infection, and then their variability was assessed in the 1000 Genomes Project database. Some variants that were described in the literature as associated to influenza infection were accessed to determine their frequency in the populations of the database. The majority of the variants found in the selected genes were present in intronic regions, except for the ST6GAL1 and TRIM25 genes, which showed higher number of variations in the 3'UTR region. Of the known variants related to influenza infection, Japanese and Chinese samples show the bigger frequencies of protective alleles (as in TLR3 rs5743313C and IL1A rs17561C alleles). The IL1B rs1143627C allele, on the other hand, increases susceptibility to influenza infection, and it is present in high frequencies in all populations of the 1000 Genomes Project, corroborating the rapid dispersion of the virus in the human population. Moreover frequencies of the rs2564978T/T genotype (promoter of CD55 gene that is highly associated with a severe form of influenza infection) in the populations of the database were concordant with mortality and morbidity rates described for 2009 pandemic H1N1. These results suggest that the susceptibility to influenza has an important heritable component. Although the clinical outcome of an influenza virus infection can't be entirely attributed to the effects of host genomic variability, these factors have been shown to play a crucial role within this context.

626M

The European Genome-phenome Archive (EGA) as a federated effort to provide secure global access to genome and phenotype data for hundreds of thousands of samples. *I. Lappalainen[1], J. Almeida-King[1], C. Gonzalez[1], J. Kandasamy[1], V. Kumanduri[1], I. Medina[1], G. Saunders[1], A. Senf[1], J.D. Spalding[1], S. ur-Rehman[1], M. Alberich[2], A. Cerreno Torres[2], J. Rambla de Argila[2], O. Martinex Llobet[2], A. Navarro[2,3,4], P. Flicek[1], J. Paschall[1].* 1) EMBL-EBI, Hinxton, Cambridge, United Kingdom; 2) Centre for Genomic Regulation, Barcelona, Spain; 3) Insitute of Evolutionary Biology, Barcelona, Spain; 4) 4Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

The European Genome-phenome Archive (EGA) is now a joint service from the European Bioinformatics Institute (EBI), UK and the Centre for Genomic Regulation (CRG), Spain. This service provides a permanent archive for all types of genetic, -omics and phenotypic data that has been consented for use in biomedical research, but where the data access is managed through an application and secure encrypted data delivery process to limit risks to the privacy of research participants. Accepted submissions include raw data from genome sequence, transcriptome, epigenome or proteomics experiments. The EGA also stores called variants, genotypes, study summary statistics and associated sample phenotypes. Access to the data is managed by appropriate data access committees (DACs) that will approve access based on requests that meet the data-use and patient consent rules governing their studies. Here we describe how EGA service will scale for future submission, data archiving and dissemination including key improvements in phenotypic query of studies as well as the provision of new models of data access including cloud-based and real-time data streams in addition to traditional file download. EGA service is accessible at https://www.ebi.ac.uk/ega/home and at https://ega.crg.eu/.

627T

Integrative Japanese Genome Variation Database from the cohort study of Tohoku Medical Megabank Organization (ToMMo). *Y. Yamaguchi-Kabata, Y. Kawai, T. Mimori, F. Katsuoka, N. Nariai, K. Kojima, I. Danjoh, S. Saito, X. Pan, J. Yokozawa, R. Saito, Y. Sato, K. Tsuda, Y. Kuroki, K. Kinoshita, J. Yasuda, M. Yamamoto, M. Nagasaki.* Dept. Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku UniversityTohoku University, Sendai, Miyagi, Japan.

Tohoku University Tohoku Medical Megabank Organization (ToMMo) was founded to establish an advanced medical system to foster the reconstruction from the Great East Japan Earthquake. The organization will develop a biobank that combines medical and genome information during the process of rebuilding the community medical system and supporting health and welfare in the Tohoku area. We have started the prospective cohort study in the region, with genome analyses, to establish the advanced, personalized medicine based on the individuals genomic data. Although reports of common variants with their frequencies are accumulating for each population, many of low-frequency variants remain undetected or lack of their frequencies. Therefore, cataloging genomic variants from whole-genome sequencing and estimation of variant frequency are necessary for foundation of genomic medicine for each population. To make a reference panel of genomic variation of Japanese, we have sequenced whole genomes of about 1,000 cohort participants, and collected single nucleotide variants (SNVs). Integrative Japanese Genome Variation Database provides data of genomic variations obtained by whole-genome sequencing of Japanese individuals who participate the genome cohort study of ToMMo. The current release provides SNV data obtained from the individuals from the cohort. The first release contains data of SNVs (about 5 millions on autosomes) that exist at least 5.0 % frequency in the samples. This browser can be used to search those SNVs and get their information such as frequency. The database and variants data would be useful for genome analysis and strategy of genomic populations for local populations.

628S

Complete haplotypes of the human light chain immunoglobulin loci from a hydatidiform mole BAC library provide insights into locus-specific signatures of genetic diversity. *K. Meltz Steinberg[1,2], C.T. Watson[3,4], T.A. Graves-Lindsay[1], R. Warren[5], M. Malig[2], J. Schein[5], R.A. Holt[5], R.K. Wilson[1], E.E. Eichler[2,6], F. Breden[4].* 1) The Genome Institute, Washington University, St. Louis, MO; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada; 5) Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada; 6) Howard Hughes Medical Institute, Seattle, WA.

Immunoglobulin gene (IG) loci and their associated germline variation are critical for pathogen-mediated immune response, but establishing complete reference sequences is problematic because of large-scale duplications and the fact that many genomic resources are established from somatically rearranged material. Here, we present a complete haplotype from the essentially haploid hydatidiform mole, CHM1. We sequenced a set of tiling BAC clones from CHM1 across the light chain IG loci, kappa (IGK) and lambda (IGL), to create single haplotype representations of these regions. The IGL haplotype is 1.25Mb of contiguous sequence from nine clones with four novel V gene and one novel C gene alleles. There is also an 11.9kbp insertion that does not contain any functional genes. The IGK haplotype is comprised of a 644kbp proximal and a 466kbp distal contig separated by a gap that is also present in the GRCh37 reference genome sequence. Our effort added an additional 49kbp of unique sequence that extended into the gap. The IGK haplotype contains six novel V gene and one novel J gene alleles. In addition, we identified a 16.7kbp region with increased sequence identity between the proximal and distal IGK contigs with signatures of interlocus gene conversion. In two instances, we observed the presence of functional alleles that had been previously classified as either “distal” or “proximal” alleles residing at loci in the alternate location. Due to the large inverted duplications, over 80% of the IGK variable (IGKV) locus consists of segmental duplications with >95% sequence identity suggesting that, unlike in the heavy chain locus (IGH), the duplications may be responsible for sequence homogenization rather than diversity. When we compared nucleotide and structural diversity among the light chain and heavy chain haplotypes, we find a three to six fold enrichment of diversity in the IGH locus compared to the light chain loci, that supports the theory that the heavy chain is more important in determining antigenic specificity.

629M

Defining Variation Sensitive Regions in Genes Associated with Hearing Loss. *A.N. Abou Tayoun[1,4], C.A. Cassa[2,4], D.M. Jordan[2], A.L. Muirhead[1], K.A. Lafferty[1], A.L. Hernandez[1], J. Shen[1], M.S. Lebo[1,3], S.R. Sunyaev[2], H.L. Rehm[1,3], S.S. Amr[1].* 1) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA; 2) Division of Genetics, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA; 3) Division of Pathology, Harvard Medical School, Cambridge, MA; 4) Authors with equal contribution.

The classification of novel variants is a major challenge facing clinical genetics laboratories. Because most novel variants do not have functional, genetic or population data to support clinical classification, in silico approaches are commonly used to prioritize candidate disease-causing variants. We have previously shown that a systematic evaluation of gene-disease associations can largely eliminate unnecessary interpretation of variants in genes with weak disease association. Applying this approach to 145 genes included in hearing loss panels, we found 54 genes (37%) with insufficient association to hearing loss that may be excluded from testing panels. Here, we extend this approach to the domain level in the remaining 91 genes with sufficient evidence for disease association, to test the hypothesis that systematic evaluation of domain-disease association can improve variant interpretation. Using Pfam predicted domain boundaries (http://pfam.xfam.org), the frequency of variation in the general population (Exome Sequencing Project, N=6,503), and clinically classified variants from ClinVar (http://www.ncbi.nlm.nih.gov/clinvar), we determine the domain level tolerance to variation within each gene. We statistically identify regions that are most sensitive to functional variation in the general population and also most commonly impacted in symptomatic individuals. This approach leverages domain functional annotation and associated disease in each gene. Our data show that this approach has utility in finding regions of greatest interest in predicting variant pathogenicity in known disease genes. For example, almost all pathogenic missense variation is confined to the ion channel domain of the KCNQ4 gene and the DNA binding domains of the PAX3 and SOX10 genes. These domains represent 15–46% of the total coding region in their respective genes, and are almost devoid of functional variation in the general population. In contrast, regions outside these domains had a high level of variation in the general population, but almost no pathogenic missense variants in symptomatic individuals. This knowledge helps define variation sensitive regions in hearing loss genes that can be used to prioritize candidate disease variants, increasing the sensitivity and specificity of novel variant assessment within these genes.

630T

Deep targeted sequencing of SLE associated LD blocks reveals multiple putatively functional variations in strong LD with SLE GWA SNPs: A haplotype based assessment of disease risk. *P. Raj[1], QZ. Li[1], D. Karp[1], E. Rai[1], C. Liang[1], B. Wakeland[1], K. Viswanathan[1], I. dozmorov[1], N. Olsen[2], JA. James[3], JA. Kelly[3], B. Lauwerys[4], PK. Gregersen[5], BP. Tsao[6], P. Gaffney[3], EK. Wakeland[1].* 1) Immunology, UT South western Med Center, Dallas, TX. 5323 Harry Hines Blvd. Dallas, Texas, 75390; 2) 2Division of Rheumatology, Penn State Hershey Medical Center, 500 University Drive, Hershey PA 17033; 3) Oklahoma Medical Research Foundation, 825 North East 13th Street, Oklahoma City, Oklahoma 73104, USA; 4) 4Service de Rhumatologie, SSS/IREC/RUMA, Cliniques Universitaires Saint-Luc & Université catholique de Louvain Avenue Hippocrate 10, bte B2.5390, 1200 BRUXELLES, Belgium; 5) Feinstein Institute for Medical Research in Manhasset, New York; 6) Division of Rheumatology, University of California Los Angeles, Los Angeles, California, United States of America.

A strong genetic predisposition has been observed in susceptibility to SLE. Genome-wide association analyses (GWAS) have identified more than 30 SLE risk loci in humans thus far, and ongoing analyses by the SLE immunochip consortium will probably define even more. We performed deep targeted sequencing of the SLE associated LD blocks with aim to develop a genomic strategy for the functional characterization of all the genetic risk alleles in SLE. For this, we sequenced ~4.4 Mb of the human genome in 773 SLE patients and 576 controls using a custom target enrichment sequencing strategy. We called 146K pass quality variations with an average of 107-fold coverage of the entire linkage disequilibrium (LD) blocks at more than 70 potential disease risk loci, thus providing detailed sequence information for all of the identified SLE risk loci. We obtained 98 percent genotype concordance between sequencing and immunochip genotypes on subset of sample. The common (MAF>0.05) genetic variants identified in this dataset were used to assemble a list of potentially functional variations in both coding and non-coding (from ENCODE) regions within these LD blocks, thus allowing the disease associations within SLE risk loci to be integrated with functional properties. Then, haplotype were formed including known GWAS and potentially functional variations in strong LD with it. We found multiple functional variations in strong LD with SLE GWA SNP in many SLE genes. For example, in case of STAT4, SLE risk haplotype formed based on STAT4 SLE GWA SNP and functional variations (ENCODE SNPs and eQTLs) pose increased risk (increased odds ratio of 1.7) than SLE GWA SNP alone (OR=1.4). Also, this risk haplotype associate with the up regulation of STAT1 and STAT4 gene expression. Thus, similar haplotype analysis was done in other SLE associated loci and phylogenetic relationships were determined between various risk and protection haplotypes by constructing Median-Joining (MJ) networks. On the other hand, low frequency and rare variations (MAF<0.05) were analyzed using KBAC and CMC methods. In conclusion, study demonstrate that how multiple functional variation can contribute to disease risk in an additive fashion and also identify loci in which both common and rare variations may contribute to SLE susceptibility.

631S

Mutation characteristics in families with two or more siblings with Autism Spectrum Disorder. *R.K.C. Yuen[1], B. Thiruvahindrapuram[1], D. Mercio[1], S. Walker[1], K. Tammimies[1], T. Nalpathamkalam[1], S. Hamilton[1], Y. Liu[1,2], M. Gazzellone[1], L. D'Abate[1], E. Deneault[1], J.L. Howe[1], R. Liu[1], A. Thompson[3], M. Uddin[1], C.R. Marshall[4], R.H. Ring[5], B.A. Fernandez[6,7], W. Roberts[8], P. Szatmari[9], S.W. Scherer[1,10].* 1) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 2) Jinan Pediatric Research Institute, Qilu Children's Hospital of Shandong University, Jinan, Shandong, China; 3) Department of Psychiatry and Behavioural Neurosciences, Offord Centre for Child Studies, McMaster University, Hamilton, ON, Canada; 4) Molecular Genetics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 5) Autism Speaks, New York, NY, USA; 6) Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St John's NL Canada; 7) Provincial Medical Genetic Program, Eastern Health, St. John's NL, Canada; 8) Autism Research Unit, The Hospital for Sick Children and Bloorview Kids Rehab, University of Toronto, Toronto, ON, Canada; 9) Centre for Addiction and Mental Health, University of Toronto, Toronto, ON, Canada; 10) McLaughlin Centre, University of Toronto, Toronto, ON, Canada.

Autism spectrum disorder (ASD) is genetically heterogeneous with >100 susceptibility genes known. We used whole-genome sequencing (WGS) of 85 quartet families (two parents and two ASD-affected siblings) to comprehensively examine mutation characteristics. In 34/85 (40%) of families, potentially relevant mutations were identified. *De novo* damaging missense and loss-of-function (LoF) mutations (single nucleotide and structural variants) were detected in ASD-risk genes in 11/85 (12.5%) families. We also identified 17 additional families carrying inherited LoF mutations in known ASD-susceptibility genes, and 6 more families with *de novo* or inherited LoF variants in genes linked to autosomal dominant forms of neurodevelopmental disorders. In only one of these families was the same *de novo* mutation (1.8kb deletion in *SCN2A*) found in both ASD siblings; in another 10 families inherited ASD-risk variants were shared between affected siblings. Taken together, in 11/34 (32%) of families a *de novo* or inherited ASD risk variant(s) identified in the index case was also found in the other sibling who developed ASD. Our results emphasize using personalized WGS to maximize the detection of all classes of mutations potentially involved in autism, and to enable the interpretation of that data in confirmatory and predictive diagnosis in different individuals in a family.

632M

Identifying candidate genes and domains for X-linked diseases using population exome data. *X. Ge[1, 2], P. Kwok[2, 3, 4], J. Shieh[1, 2].* 1) Division of Medical Genetics, Department of Pediatrics, University of California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 3) Department of Dermatology, University of California, San Francisco, San Francisco, CA; 4) Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA.

Many new human disease genes can be identified through high-throughput sequencing. Yet variant interpretation for the large amounts of genomic data remains a major challenge. There is an abundance of variation of uncertain significance, and it is difficult to interpret variants in genes that lack disease annotation. As clinically-significant disease genes may be subject to negative selection, we predicted these genes might have a paucity of nonsynonymous variation in the population. The aim of this study was to develop methods to predict pathogenicity using population exome data. We analyzed and integrated human X-chromosome bulk variant data from six thousand individuals (the NHLBI Exome Sequencing Project) to assess intra-human ratios of substitution rates at non-synonymous and synonymous sites (dN/dS) as a potential measure for gene-based pathogenicity. Indeed, we find that the dN/dS ratio is significantly lower in OMIM disease genes, supporting the sensitivity of this intra-human method. Interestingly, genes that are associated with childhood disease outcome also demonstrate a significant relationship to dN/dS, and we have performed some validation using a secondary exome dataset. By analyzing all X-chromosome genes, we identified new candidates for diseases with early mortality, and several of these have been verified by recent exome and gene identification studies. Furthermore, we demonstrate intragenic localized patterns of variants that suggest pathogenic hotspots, which could be used to identify new candidate protein domains linked to diseases. Our results suggest intra-human substitution analysis is a valuable tool to help prioritize novel disease genes in sequence interpretation.

633T

A15924G mt-tRNAT Gene Mutation is not the Primary Cause of Mitochondrial Myopathies. *A. Cakiris[1], N. Abaci[1], Z. Emrence[1], F. Pacal[1], C. Gurses[2], B. Kara[3], D. Ustek[1].* 1) Genetics, Institute for Experimental Medicine, Istanbul, Turkey; 2) Neurology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey; 3) Pediatric Neurology, Kocaeli University Medical Faculty, Kocaeli, Turkey.

Mitochondria are essential organelles present in all nucleated mammalian cells. Mitochondrial genome has a very high mutation rate at 10–17 folds higher than the nuclear genome. Those MtDNA mutations may occur within proteins, tRNA or rRNA genes. But more than half of mutations that are disease related are located in mt-tRNA genes. Some of these are polymorphic and the others are pathogenic mutations. In order to analyze mitochondrial DNA mutations, the entire mitochondrial genome was amplified in two overlapping polymerase chain reactions. Mitochondrial DNA was deep sequenced by next generation sequencing technology. The blood or muscle tissue of the 267 patients diagnosed with undefined mitochondrial myopathy and 330 healthy individuals have been analyzed. A15924G mt-tRNA mutation was detected to be homoplasmic in 12 patients and 6 healthy control samples. A15924G mt-tRNA mutation can alter the free energy of the thermodynamic ensemble for the secondary structure of wild type TRNT and affect the structure of tRNAs. This mutation was previously shown to be associated with the LIMM disease as a pathogenic mt-TRNT mutation. In another study, this mutation has been reported as a common mtDNA polymorphism and not the primary cause of LIMM. In our study, this mutation was found in 12 patients, 7 of which belongs to haplogroup I, 2 to F1b1, 1 to U7b, and 1 to T2h haplogroup. Among the healthy individuals, this haplogroup distribution is as follows; 4 haplogroup I, 1 U7b, and one to haplogroup F1b1. According to the above results, the A15924G mutation in both patients and healthy individuals is considered as a haplogroup marker. These results suggest that the A15924G mutation is a common mtDNA polymorphism, rather than a pathogenic mtDNA mutation.

634S

A Mouse Mutagenesis Scheme to Isolate Lethal X-Linked Recessive Mutations. *F.J. Probst[1], R.R. Corrigan[2], M.N. Bainbridge[3], S.N. Jhangiani[3], H.V. Doddanpaneni[3], J. Hu[3], R.A. Gibbs[3].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Neuroscience, Baylor College of Medicine, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

Large-scale mutagenesis of the mouse has yielded numerous new autosomal mutants, but X-linked recessive phenotypes have been virtually nonexistent. We have devised a two-generation mouse mutagenesis screen to capture lethal and sub-lethal X-linked recessive mutations. C57BL/6J (B6) male mice are mutagenized with the chemical "supermutagen" ENU. These mice are then bred to 129S6/SvEvTac (129) females to generate G1 progeny. The females from this cross will be X(B6*)X(129) (i.e., they will have inherited a mutagenized B6 X chromosome from their father and a 129 X chromosome from their mother). These G1 animals are then bred to 129 males to generate G2 progeny. In theory, the X(B6*)X(129) G1 female from each pedigree will transmit her X(B6*) chromosome to half of her male progeny and her X(129) chromosome to the other half. When meiotic "crossing over" is considered, this means that, for each X-linked allele, about 1/2 of the male progeny should have the B6 allele, and the other half should have the 129 allele. However, if the X(B6*) chromosome carries an X-linked recessive lethal mutation, then there will be no male mice with B6 alleles at the markers flanking the mutation, since all of these mice will not survive to weaning. DNA from 10 G2 male progeny of each G1 female are therefore combined to create a single "pool" of DNAs for each pedigree. Each pool is then run on an Illumina Mouse Medium Density Linkage Panel. If the female carries a lethal X-linked recessive mutation, the male mice that inherited this mutation will have died prior to weaning, and there will be no B6 alleles in the pools for the markers around the mutation. Though some drift is to be expected in the data, the chance of a particular region of the X chromosome appearing to be homozygous for the 129 alleles is less than 1/1,000, and additional mice can be bred from potential mutant G1 carriers in order to confirm the result. DNA from critical G1 female animals is then subjected to exome capture via a Roche Nimblegen EZ Library, and the products are analyzed by next-generation sequencing on an Illumina apparatus. Causative mutations are confirmed via linkage analysis and additional phenotypic and biochemical studies. Many of the resulting mutants are likely to be models of human X-linked diseases, thus providing new insights into the cause and pathophysiology of X-linked diseases.

635M

The relative impact of DNA mutation and RNA editing as sources of somatic sequence variation in the transcriptomes of normal adult tissues. *D. Conrad[1], N. Huang[1], P. Kheradpour[2,3], M. Kellis[2,3], K. Ardlie[3], The GTEx Consortium.* 1) Department of Genetics, Washington University School of Medicine, St Louis, MO; 2) MIT Computer Science and Artificial Intelligence Lab, Cambridge, MA; 3) Broad Institute, Cambridge, MA.

The full extent and origins of somatic mosaicism in humans are unknown but could have profound implications for our understanding of human health and disease. We have developed a statistical framework for inferring DNA mutations from multi-tissue RNA sequencing data, which considers sequencing error, sampling error, RNA editing and allele-specific expression as alternative explanations for apparent sequence differences among transcripts from the same locus. We have applied this framework to the GTEx Consortium data to detect RNA editing and somatic mutation from 1,514 non-cancerous tissues obtained from 175 donors. We identified 174 putative somatic mutations across 77 donors (min/donor=0; max = 8). In contrast, we observed some evidence for RNA editing in one or more tissues at an average of 1,733 sites per donor (min, 669 sites; max, 6,522). Thus, RNA editing appears to dominate over somatic mutation as a process creating high-frequency somatic sequence variation in the transcriptomes of normal cells. We constructed an expression fingerprint combining *ADAR*, *ADARB1* and *ADARB2* that is strongly correlated with the average extent of RNA editing within individual tissues (R=0.88 in brain tissues, $p< 5.7 \times 10^{-5}$ and R=0.76 in other tissues, $p<4.7 \times 10^{-6}$).

Like RNA editing by the *ADAR* family of enzymes, somatic mutations can be "programmed" to some extent by directed enzymatic activity. V(D)J recombination, catalyzed by *RAG1* and *RAG2*, and somatic hypermutation, catalyzed by *AID*, act on the variable regions T-cell receptor and/or immunoglobulin genes. Of the 174 somatic mutation calls, 86 (49%) occur in the immunoglobulin genes, and lung is the single tissue with the most antibody mutations (24/86, 28%). Partly motivated by this observation, we have performed extensive QC on all somatic mutation calls by read realignment and manual curation, but definitive evidence will be provided by experimental validation of the entire callset, which is ongoing. Of the non-antibody mutation calls, 18 (20%) are observed across multiple tissues in the same person, while 26 (30%) are observed only in blood or LCLs. As these data are derived from transcript sequences, many of the mutations are likely to be functional, and we interpret their origin and impact by investigating their sequence context, relation to cancer genes, and medical history of the donors.

636T

Distinct variation in the LILRB3 and LILRA6 genes encoding a myeloid inhibitory and activating receptor pair. *A. Bashirova[1, 2], C. O'hUigin[1], M. Carrington[1, 2].* 1) Leidos Biomedical Research , Frederick, MD; 2) Ragon Institute of MGH, MIT, and Harvard, Boston, MA.

The leukocyte immunoglobulin-like receptor (LILR)B3 and LILRA6 genes encode homologous inhibitory and activating orphan receptors. Both genes are expressed in monocytes and exhibit a strikingly high level of polymorphism at the amino acid level. We have demonstrated recently that LILRA6 (but not LILRB3) can be present in variable copy numbers in the genome, probably due to crossovers between the two genes. In a cohort of 228 healthy white individuals, we observed that 64% had two copies of LILRA6 per diploid genome, while 32% had more than two and 4% had only one LILRA6 copy. We have now characterized SNPs in the two genes in a subset of individuals with two copies of LILRA6 (N=91). Across seven exons encoding the signal peptide and extracellular D1-D4 domains, we have identified 46 and 38 amino acid changing SNPs in LILRB3 and LILRA6, respectively. Among these, 35 SNPs were located at identical positions within the two genes. While minor allelic frequencies (MAFs) of more than two thirds of the SNPs in each gene were greater than 10%, there were substantial differences in MAFs of many SNPs between LILRB3 and LILRA6. In exons encoding the signal peptide and the D3/D4 domains, some SNPs had MAFs of <2% in one gene and up to 40% in the other gene. These differences were also reflected in distinct patterns of linkage disequilibrium (LD) between pairs of SNPs in the two genes, with LILRA6 demonstrating a higher degree of LD. The dataset on LILRB3/A6 SNPs will be useful in further functional and evolutionary characterization of this highly polymorphic locus, which may be involved in regulation of immune responses in disease pathogenesis.

637S

Finding effectively neutral sequence in the presence of coding and noncoding conserved elements. *A.E. Woerner[1,3], K.R. Veeramah[2], M.F. Hammer[3].* 1) Computer Science, University of Arizona, Tucson, AZ; 2) Department of Ecology and Evolution, Stony Brook, NY; 3) Arizona Research Laboratories, University of Arizona, Tucson, AZ.

Early studies that have attempted to utilize putatively neutral loci in the genome for demographic inference have focused on pseudogenes, introns, and repetitive elements such as Alus. More recent studies have attempted to reduce the impact of positive and negative selection as well as background selection and genetic draft by finding loci that are either in regions of low gene density and high recombination rate, or are at a maximal distance from the nearest gene in genetic units. However, nucleotide diversity may not only be reduced near genes, but also near noncoding conserved elements. Because ~5% of primate genomes is composed of evolutionarily conserved elements, with ~2/3 of these elements being noncoding, skews in diversity in regions in linkage with such elements could have large effects on genomic patterns of nucleotide variation. Here, we show that the combination of multiple conserved noncoding elements has more of an influence on nucleotide diversity than the distance to the nearest gene. Using an approach that models the combined effects of linkage to coding and noncoding elements, we predict levels of polymorphism on both the autosomes and the X chromosome. Based on our predictions we identify loci that we consider to be effectively neutral, i.e., regions whose expected diversity is at least 99% of our inferred neutral rate. Using the relationship between diversity and the local recombination rate as a proxy for the effects of directional selection on neutral sites, we show that unlike other loci in the genome, our effectively neutral loci show no correlation between the local recombination rate and diversity. We conclude that "neutral" sequence is extremely rare in the genome. Finally, we use these neutral regions to estimate the relative effective population size of the X chromosome vs. the autosomes in African and non-African human populations.

638M

The associations of multiple genes with systemic sclerosis by next generation sequence technology. *H. Li, X. Guo, S. Assassi, J. Reveille, M. Mayes, X. Zhou.* University of Texas - Houston Medical School, Houston, TX.

Systemic sclerosis (SSc) is a rare and complex immune-mediated disease characterized by vasculopathy, fibrosis of skin and internal organs and the presence of autoantibodies. According to genome-wide association studies, the HLA region contains the strongest SSc-associated loci. Classical HLA class II genes (DPB1, DQB1 and DRB1) have been clearly associated with SSc. However, the non-classical HLA genes have not been systemically investigated. Here we investigated the HLA genes, PSORS1C1, CCHCR1, MICA, MICB, NFKBIL1, NOTCH4, C6orf10, BTNL2, TAP2 and TNFAIP3 in association with SSc by using next generation sequencing technology (NGS, Ion torrent personal genome machine). We sequenced all exons for the above genes in a large US cohort. A total of 979 sequence variants were identified in these genes. Variant counts for PSORS1C1, CCHCR1, MICA, MICB, NFKBIL1, NOTCH4, C6orf10, BTNL2, TAP2, TNFAIP3 were 44, 138, 85, 42, 57, 232, 102, 101, 107, 71 respectively. These variants included 654 exonic and 325 intronic variants. A nominal association with SSc ($p < 0.05$) was observed in 77 variants of the examined genes with 30 variants of NOTCH4. NOTCH4 is part of an ancient signaling system involved in cell fate decision, and it was reported to be involved in kidney fibrosis and patterning and development of blood vessels. A subsequent confirmation study of NOTCH4 with Sanger sequencing was conducted. Seven out of nine exonic variants were confirmed to be correct and two deletions were confirmed to be substitutions. NOTCH4 rs520692 (p. D272G), rs915894 (p. K117Q) and rs422951 (p.T320A) are the missense variants; rs204987 (p.T1367T), rs520803 (p. Q284Q), rs520688 (p.P271P) and rs423023 (p.G348G) are the synonymous variants. The haplotype of rs520692, rs520803 and rs520688 was associated with anti-TOPO I ($p = 1.1 \times 10^{-8}$, OR 0.37). Rs423023 and rs422951 were also associated with anti-TOPO I at $p = 3.3 \times 10^{-8}$, OR 0.36 and $p = 4.9 \times 10^{-5}$, OR 0.44. NOTCH4 rs204987 was associated with the anti-centromere autoantibody subset ($p = 2.1 \times 10^{-6}$, OR 6.59). These variants showed no association with HLA-DRB1 and are reported to be associated with SSc for the first time. Our results indicate there are many susceptible genes in the non-classical HLA region that are associated with SSc, which further supports the concept that SSc is a complex genetic disease with contributions from multiple loci. NOTCH4 is an important and independent gene from HLA-DRB1 associated with SSc.

639T

Evolutionary constraint and disease associations of post-translational signaling sites in human genomes. *J. Reimand, G.D. Bader.* The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada.

Interpreting genome variation in the context of phenotype and molecular function, distinguishing disease variants, and assessing personal risk are central challenges of biomedical genomics, hindered by predominantly rare inter-individual variation. Potentially functional protein-coding variants found from genome sequencing and genome-wide association studies are often ranked by evolutionary conservation and population frequency. However it is also likely that certain intrinsic protein features carry special functional significance. We hypothesized that protein regions involved in post-translational modifications (PTMs) are functional hotspots of human genomes. PTMs are biochemical alterations of amino acids that extend the functional repertoire of proteins, conducted by enzymes that recognize substrate proteins through short linear motifs (SLMs) in protein sequence. Currently we have public experimental data on ~130,000 PTM sites in human proteins. The centrality of PTMs in biological processes suggests the functional importance of underlying genome sequence. Here we found that PTM-associated protein regions representing ~11% of protein sequence follow a specific constraint in the human population. This is distinct from major sources of variation such as conservation, recombination, or GC content. PTM regions are also enriched in disease and cancer mutations, further emphasizing the deleteriousness of their variation. PTM constraint is exemplified in chromosomal context, gene signatures of most human tissues, and across a thousand processes and pathways. PTM regions include ~250,000 protein residues whose substitution would disrupt SLMs bound by PTM enzymes. Intriguingly these residues are enriched in disease mutations, confirming their importance in physiology. As many PTM enzymes are targetable with approved drugs, we use their substrate specificity data to build a network of drug-disease interactions mediated by enzymes that bind mutated sites in disease genes. This represents a useful resource for drug screen design and hypothesis generation. Finally, we applied our ActiveDriver model to study disease mutation hotspots in PTM regions. For example, we propose a new mechanism to variants of the PTPN11 gene responsible for Noonan syndrome. In summary, we describe a novel constraint of the protein-coding genome, and demonstrate use of proteomics data to interpret genome variants, improving associations to phenotypes, molecular mechanism and disease.

640S

Genetic analysis of dendritic cell responses to influenza using RNA sequencing reveals novel genotype by stimulation effects on alternative splicing. *C.J. Ye[1], J. Chen[1,7], A.C. Villani[1], M.N. Lee[1,2,3], T. Raj[1,2,4], B.E. Stranger[5], P. De Jager[1,2,4], C.O. Benoist[2,4], T. Bhangale[6], W. Ortmann[6], T. Behrens[6], A. Regev[1,7,8], N. Hacohen[1,2,3].* 1) Broad Institute, Cambridge, MA; 2) Harvard Medical School, Boston, MA; 3) Massachusetts General Hospital, Boston, MA; 4) Brigham and Women's Hospital, Boston, MA; 5) University of Chicago, Chicago, IL; 6) Genentech Inc., South San Francisco, CA; 7) Massachusetts Institute of Technology, Cambridge, MA; 8) Howard Hughes Medical Institute.

Genetic variation affecting splicing plays an important role in human disease both through cis effects on splice isoforms and by trans effects mediated through the splicing machinery. Genetic variants are known to modulate gene expression levels (eQTLs) in response to in vitro stimulation, yet systematic genome-wide investigation of how variants affect alternative splicing has not been widely explored. RNA-seq opens the possibility to directly estimate isoform usage ratios and assess the underlying genetic variants (irQTL). Here, we sought to identify cis irQTLs, and to elucidate the functional mechanisms by which they act in the anti-viral transcriptional response of innate immune cells using RNA-seq. We sequenced primary dendritic cells from 279 individuals in response to stimulation with influenza virus and the cytokine IFNβ. Post flu stimulation, in addition to 3474 genes with significant cis eQTLs (with 52% replication in GEUVADIS, correlation of effect sizes $\rho(\beta)=0.59$), we identified 1069 genes with significant cis irQTLs, higher than previous estimates. Three lines of evidence support our findings: (i) 67% internal replication with IFNβ stimulation ($\rho(\beta)=0.97$), (ii) 54% replication with the GEUVADIS samples ($\rho(\beta)=0.8$) using our computational pipeline, and (iii) replication with allele specific isoform usage analysis. As a class, cis irQTLs cluster near known splice sites and overlap with GWAS loci at the same frequency (17%) as cis eQTLs suggesting the potential importance of differential splicing in disease pathogenesis. We highlight ERAP2, and a IBD associated loss-of-function (LoF) cis variant, rs2248374, that induces the alternative splicing of a longer isoform predicted to undergo non-sense mediated decay (NMD). We observed the expected ERAP2 cis irQTL post IFNβ stimulation. But remarkably, in flu-stimulated cells, we observed near complete abrogation of the cis irQTL and rescued expression of the NMD isoform, confirmed by allele specific analysis. These results suggest potentially a new host response mechanism of NMD-escape or a flu specific mechanism for NMD inhibition, either of which alters our interpretation of an apparent LoF polymorphism. Taken together, our results suggest that modulation of alternative splicing may be a widespread and important mechanism by which genetic variants control transcript diversity by creating proteins of differing function or transcripts with altered transcript processing or stability.

641M

Expression quantitative trait analysis in human platelets. *L. Simon[1], E. Chen[1], L. Edelstein[2], P. Bray[2], C. Shaw[1].* 1) Baylor College of Medicine, Houston, TX; 2) Thomas Jefferson University, Philadelphia, PA.

Human platelets are responsible for coagulation physiology and maintenance of haemostatic balance. Abnormal platelet reactivity is associated with various disease states such as bleeding disorders, atherosclerosis, occlusive or thrombotic cardiovascular disorders, inflammation and cancer, resulting in significant morbidity and mortality. The Platelet RNA and eXpression-1 study was designed to investigate significant heritable inter-individual variation in platelet functional properties. This study profiled platelet mRNA and microRNA expression as well as genotype of 154 healthy human subjects. To study the regulatory landscape of the human platelet, we conducted expression quantitative trait locus (eQTL) analysis and identified a large number of loci significantly associated with gene expression levels in cis and trans. The comprehensive approach to cataloging all platelet eQTLs can be of great help in future studies, such as the prioritization of trait-associated loci implicated in genome-wide association studies.

642T

Accurate and fast multiple testing correction to identify eGenes in eQTL studies. *J. Sul[1,2], B. Han[1,2], T. Raj[2,3,4], P. de Bakker[1,2,5], S. Raychaudhuri[1,2,6], B. Stranger[7,8], E. Eskin[9,10].* 1) Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA; 2) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 3) Harvard Medical School, Boston, Massachusetts, USA; 4) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham & Women's Hospital, Boston, Massachusetts, USA; 5) Departments of Epidemiology and Medical Genetics, University Medical Center Utrecht, 3584 CG Utrecht, the Netherlands; 6) Arthritis Research UK Epidemiology Unit, Musculoskeletal Research Group, University of Manchester, Manchester Academic Health Sciences Centre, Manchester M13 9PT, UK; 7) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois, USA; 8) Institute for Genomics and Systems Biology, University of Chicago, Chicago, Illinois, USA; 9) Computer Science Department, University of California, Los Angeles, California, USA; 10) Department of Human Genetics, University of California, Los Angeles, California, USA.

Expression quantitative trait loci (eQTL) studies attempt to detect genetic variations called eQTLs that alter expression levels of genes. Those studies are especially interested in identifying genes called eGenes that contain eQTLs because genes are primary entities utilized in a disease pathway analysis. A traditional approach to detect eGenes is to find a genetic variant with the minimum p-value among all variants in *cis* with each gene and to perform a multiple testing correction to obtain a gene-level p-value. The permutation test is widely used for the multiple testing correction because it considers LD structure among variants. However, the sample size of eQTL studies has recently grown larger, and the permutation test has become a computational bottleneck in eQTL studies. In this paper, we propose an efficient approach to correct for multiple testing utilizing a multivariate normal distribution (MVN) that approximates the null distribution of test statistics. Our approach takes into account LD structure among variants, and its time complexity is independent of the sample size. However, one challenge is that the true null distribution of statistics does not exactly follow the MVN, especially in datasets of the small sample size. We propose an accurate method that corrects for this discrepancy between the true null distribution and MVN. We show that our approach yields gene-level p-values very close to those from the permutation test using three human eQTL datasets including Genotype-Tissue Expression project (GTEx). Our method is 38 times faster than the permutation test when the sample size is 300 and much faster for the larger sample size.

643S

Ttn as a likely causal gene for QTL of alcohol preference on mouse chromosome 2. *L. Wang[1], Y. Jiao[1], Y. Huang[1], B. Bennett[2], R.W. Williams[3], D. Li[4], H. Zhao[5,6], J. Gelernter[4,6,7,8], H.R. Kranzler[9,10], L.A. Farrer[11], W. Gu[1].* 1) Orthopaedic Surgery, University of Tennessee Health Science Center, Memphis, TN; 2) Pharmacology, University of Colorado Denver, Aurora, CO; 3) Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, TN; 4) Psychiatry, School of Medicine, Yale University, New Haven, CT; 5) Epidemiology and Public Health, School of Medicine, Yale University, New Haven, CT; 6) Genetics, School of Medicine, Yale University, New Haven, CT; 7) VA Connecticut Healthcare Center, West Haven, CT; 8) Neurobiology, School of Medicine, Yale University, New Haven, CT; 9) Psychiatry, School of Medicine, University of Pennsylvania, Philadelphia, PA; 10) VISN 4 MIRECC, Philadelphia VAMC, Philadelphia, PA; 11) Departments of Medicine (Biomedical Genetics), Neurology, Ophthalmology, Genetics & Genomics, Biostatistics, and Epidemiology, Schools of Medicine and Public Health, Boston University, Boston, MA.

Many quantitative trait loci (QTL), influencing mouse model phenotypes for alcoholism, have been mapped genetically. However, the gene(s) comprising the QTL (QTG) are largely unknown. In previous work, Dr. Bennett and colleagues created congenic strains carrying the DBA/2IBG (D2) region for alcohol preference (AP) on chromosome 2, on a C57BL/6IBG (B6) background. Subsequently, interval specific congenic recombinant strains (ISCRS), in which the full D2 QTL region was broken into smaller, partially overlapping regions of introgression, were generated and tested. With information from two ISCRS ,the QTL has been mapped onto mouse chromosome 2 (Chr2) in a region of 3.4Mb by using C57BL/6J (B6) × DBA/2J (D2) recombinant inbred (RI) strains as well as by using F2 populations. Several candidate genes, Gad1, Atp5g3, Atf2, Sp3 and Sp9, have been evaluated but none of them is confirmed for a definitive role in the regulation of the QTL of AP on Chr2. We have been searching candidate gene for this QTL intensively by using an integrative approach including 1) bioinformatics tools to search potential function relevant genes of alcohol preference within the QTL region; 2) searching for single nucleotide polymorphisms (SNPs) within the exons of every gene between B6 and D2 in the QTL region; 3) conducting real time PCR to examine the differentialyl expressed genes between B6 and alcohol preferred interval-specific congenic recombinant strains (ISCRS); and 4) analysis of association of candidate gene in human population. Titin (Ttn) is known as a giant muscle protein expressed in the cardiac and skeletal muscles. However, its expression level in the tongue is known to be higher than that in the heart. We therefore investigated if Ttn plays a role in the regulation of AP. Our data indicated that 1) the expression level of Ttn in the less AP congenic strains is significantly higher than that in B6; 2) the expression of a Ttn probe in the BXD RI strains is negatively correlated to that of AP; 3) One SNP is in up- and the other is in down-stream of Ttn. The alcohol consumption of the B6 genotype is significantly higher than that of D2 genotype in the BXD RI strains, based on data from multiple reports using two-bottle of choice; and 4) the polymorphism of TTN in human population is highly associated with alcoholism. We conclude that Ttn is a likely causal gene for the QTL on Ch2 for the AP.

644M

Cross-population Meta-analysis of eQTLs: Fine-mapping and Functional Study. *X. Wen[1], G. Moyerbrailean[2], F. Luca[2], R. Pique-Regi[2].* 1) Dept Biostatistics, Univ. of Michigan, Ann Arbor, MI; 2) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI.

Data sets collected from multiple ethnic groups provide great potentials to understand the genetic basis of gene regulations. Nevertheless, the joint analysis of multiple-population eQTL data presents great challenges, mainly due to varying LD patterns and heterogeneity of genetic effects presented in multiple populations. In this study, we develop a set of statistical and computational tools to tackle the problem of cross-population meta-analysis of eQTLs, primarily focusing on the applications of fine mapping and functional study. Our approach holds three distinct advantages over the existing methodologies:

1. Our method effectively identifies eQTL signals that show consistent genetic effects across populations while accounting for the potential heterogeneity based on a powerful Bayesian meta-analysis framework (Wen and Stephens, 2014 AOAS; Flutre et al, 2013 PLoS Genetics).
2. Our approach performs efficient multiple SNP fine-mapping analysis across multiple populations allowing varying LD patterns.
3. Building on the fine-mapping results, our method enables high-resolution functional analysis of eQTL signals accounting for LD.

We apply our methods to re-analyze the GEUVDIS data consisting of five population groups. Furthermore, we use base-pair resolution annotation obtained with an improved CENTIPEDE method using ENCODE DNase-seq data. Our preliminary results show that

1. Meta-analysis is more powerful and robust in identifying eQTL signals. We identify 6,555 genes harbouring at least one eQTL, at FDR 5% level, in their cis-regions (eGenes) using the joint meta-analysis approach. In comparison, separate subgroup analysis identifies 1803 (TSI), 2078 (GBR), 2100 (FIN), 960 (CEU) and 1042 (YRI) eGenes.
2. Many genes contain multiple independent cis-eQTLs in their cis-regions. With the meta-analysis powered fine-mapping approach, we are able to confidently identify 674 genes containing 2 or more independent cis-eQTL signals and 67 genes containing 3 or more signals.
3. Binding variants are significantly enriched in the identified population consistent eQTL signals (p-value < 1.6×10^{-11}).

.

645T

Linking systems genetics and co-expression analysis to elucidate diabetic kidney disease regulatory networks. *T. Leak[1], C. Komorosky[1], V. Nair[1], H. Huang[1], B. Keller[1], A. Randolph[1], R. Nelson[2], M. Kretzler[1], T. Werner[1,3].* 1) Internal Med-Nephrology, Univ of Michigan, Ann Arbor, MI; 2) 2National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, Arizona; 3) Genomatix Software GmbH, Munich, Germany.

Expression quantitative trait loci (eQTL) analysis associates genetic and transcriptomic data, but does not provide insight into how eQTLs influence sets of genes in complex diseases. Using diabetic kidney disease (DKD) as a case study, genome-wide SNPs, protocol biopsy gene expression and clinical data were obtained from 65 Pima Indians. Genome-wide systems genetic analysis paralleled with weighted co-expression network analysis (WGCNA) was performed on gene expression profiles from microdissected glomeruli. Integration of glomerular eQTL analysis and WGCNA was carried out to identify transcriptional coregulation-modules. Matrix eQTL was employed to test allele dosage effect on gene expression levels. An in silico promoter analysis approach was used to identify common transcription factor binding sites (TFBS) frameworks shared between the promoters of cis-eQTL transcripts and potentially within LD blocks of associated expression (e)SNPs. Multivariate linear regression was then used on the module eQTLs, promoter gene matches, TF belonging to TFBS families represented in the framework and additive genotypic model, to test for association with indices of renal function and diabetes traits. A total of 13,179 transcripts were parsed into 37 co-expression modules and Matrix eQTL resulted in 3,112 cis-eQTL gene pairs. To establish a mechanistic link between correlated cis-eQTL modules, a module with a similar number of transcripts and eSNPs was analyzed via an in silico promoter analysis approach (dark magenta module). Framework analysis of tag-eSNP LD block sequences with kidney associated promoters revealed a four-element framework (ETS-SP1-PAX5-HOMF; 5.64 fold over presented) that matched 2 of the 3 eQTLs (TAF9, HPGD). The LD block spans TAF9, RAD17 and MARVELD2, two of which (TAF9 and RAD17) have matches of the framework in their promoters. A genome-wide search of all human promoters (N=101,492) revealed only 17 matches. Four of these genes, ETS TFs and rs6450041 (located in the same LD block and an eSNP for RAD17) showed suggestive evidence of association with decreased renal function ($P \leq 0.000001$). Our integrative framework potentially highlights how eSNPs may influence the expression of several genes linked in a LD block with renal function. Preliminary data indicate that other eQTLs may also act on gene sets employing similar strategies, i.e TFframeworks that could coordinate transcriptional regulation, all of which is linked with DKD.

646S

Mediation analysis identifies causal relationships among SNPs, cis-CpG methylation, and cis- and trans-transcripts. *B. Pierce[1], L. Tong[1], R. Rahaman[1], L. Chen[1], M. Kibriya[1], M. Argos[1], J. Farzana[1], S. Roy[1], R. Paul-Brutus[1], R. Zaman[2], T. Islam[2], M. Rahman[2], J. Baron[3], H. Ahsan[1].* 1) Health Studies, Univ Chicago, Chicago, IL; 2) UChicago Research Bangladesh, Mohakhali, Dhaka, Bangladesh; 3) University of North Carolina, Lineberger Comprehensive Cancer Center, Chapel Hill, NC.

A large fraction of human genes are regulated by genetic variation. However, our understanding of the mechanisms by which genetic variants influence transcription is incomplete. A more in-depth understanding of eQTLs (expression quantitative trait loci) can be gained by identifying cellular features that mediate the effects of genetic variants on gene expression. In this work, we use Sobel tests of mediation to determine (1) if trans-eQTL associations are mediated by transcripts in cis with the SNP showing the trans-association and (2) if cis-eQTL associations are mediated by local chromatin structure (measured as DNA methylation). Using data on 1,799 Bangladeshi individuals with genome-wide SNP and array-based expression data, we demonstrate that among 197 independent trans-eQTLs (FDR= 0.05), 76 were also cis-eQTLs, and evidence of mediation was detected for 55 of these 76 trans-eQTLs (Sobel P <0.01). Using a subset of these participants (n=377) with genome-wide DNA methylation data (Illumina 450K), we detect mediation by cis-CpG methylation status for 22% of 2,598 observed cis-eQTLs, suggesting that alteration of chromatin features is a common mechanism by which genetic variation influences gene expression. In these analyses, "partial" mediation, rather than "full" mediation, is most often observed, and using simulated data, we show that this result is expected. More specifically, in the presence of mediator measurement error or imperfect LD between measured and causal variants, complete mediation will be observed as partial mediation. Our results also show that eQTL associations can become significantly stronger after adjusting for a potential mediator, a counterintuitive observation. Using simulated data, we demonstrate that this phenomenon can occur when the when the analyzed mediator is not the true mediator, but is correlated with the true (unmeasured) mediator. In this work we demonstrate mediation for a large fraction of cis- and trans-eQTLs; however, our ability to detect mediation would be enhanced by the use of more comprehensive data on potential mediators (e.g., RNA-seq and bisulfite-seq). Mediation analysis will be useful for validation and discovery of eQTLs and is a valuable tool for characterizing the biological mechanisms underlying eQTLs.

647M

Characterizing the genetic basis of innate immune response in TLR4-activated human monocytes. *S. Kim[1], B. Pütz[3], M. Bechheim[2], J. Becker[1], M. Nöthen[1], B. Müller-Myhsok[3], V. Hornung[2], J. Schumacher[1].* 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Institute of Molecular Medicine, University of Bonn, Bonn, Germany; 3) Statistical Genetics, Max-Planck-Institute of Psychiatry, Munich, Germany.

Toll-like receptors (TLRs) play a key role in innate immunity. Apart from their function in host defense, dysregulation in TLR-signaling can confer risk to autoimmune diseases, septic shock or cancer. Despite major advancements in our understanding of how the innate immune system recognizes pathogens, the genetic basis for differences in innate immune responses is only poorly defined. This study was aimed to characterize the genetic basis of variation in gene expression in TLR4-stimulated human monocytes. For this purpose we isolated monocytes of 137 individuals and stimulated them with lipopolysaccharide (LPS) to activate Toll-like receptor 4 (TLR4). From these donors, we performed transcriptome profiling and genome-wide SNP-genotyping. Comparing unstimulated versus TLR4-stimulated monocytes revealed expression quantitative trait loci (eQTLs) that are unique to TLR4 stimulation. Among these, we show that SNPs conferring risk to primary biliary cirrhosis (PBC), inflammatory bowel disease (IBD) and celiac disease are immune response eQTLs for novel candidate genes, bringing new insights into the pathophysiology of these disorders in the context of TLR4-activation.

648T

Analysis of genetic and environmental determinants of gene expression. *F. Luca[1], C. Harvey[1], G.O. Davis[1], G. Moyerbrailean[1], D. Watza[1], X.W. Wen[2], R. Pique-Regi[1].* 1) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI; 2) Department of Biostatistics, University of Michigan, Ann Arbor, MI.

The effect of genetic variants on a molecular pathway, and ultimately on the individual's phenotype, is likely modulated by "environmental" factors. However, it is generally difficult to determine in which tissues and conditions genetic variants may have a functional impact. We denote the functional genetic variants that show cellular environment-specific effects as GxE expression quantitative trait loci (GxE-eQTLs). Achieving a better understanding of the mechanisms underlying GxE-eQTLs is a critical step in understanding the link between genotype and complex phenotypes. To identify and characterize GxE-eQTLs we have established a new two-step and cost-effective experimental approach. In the first step, we identify global changes in gene expression using low-coverage sequencing of pools of highly multiplexed samples. In the second step, we select a subset of samples for deep sequencing and allele-specific analysis. For the first step, we generated 1056 RNA-seq libraries in pools of 96 spanning 250 cellular environments across 5 cell-types (3 individuals), and 50 different treatments (including hormones, dietary components, environmental contaminants and metal ions). Relevant GO categories were enriched in the observed global gene expression changes (e.g. immune response for Dexamethasone, ion homeostasis for Zinc). We then analyzed allele specific expression (ASE) using a novel method (QuASAR) that allows for joint genotyping and allele specific analysis on RNA-seq data. Across 21 cellular environments we discovered 2469 instances of ASE (FDR<10%), corresponding to 1402 unique ASE genes. Using Bayesian model across treatments within cell types, we observed that 95.9% ASE signals are shared (95% credible interval [93.4%-99.3%]) and their effect sizes are highly concordant (posterior correlation coefficient 0.9). This is highly consistent with previous analysis of condition-specific eQTLs. Nevertheless, we still identified 270 loci with a Bayes factor supporting GxE (200 sites treatment-specific and 70 sites control-specific GxE-eQTLs). We are now in the process of expanding the ASE analysis to 32 additional cellular environments. Our results constitute a first comprehensive catalog of GxE-eQTLs and we anticipate that it will contribute to the discovery and understanding of GxE interactions underlying complex traits.

649S

Identification of novel trans-acting eQTLs at *SLC25A38-MYRIP-EIF1B* and *CIITA* in subcutaneous adipose tissue. *Y. Wu[1], M. Civelek[2], C.K. Raulerson[1], A. He[3], C. Tilford[3], C. Fuchsberger[4], A.E. Locke[4], H.M. Stringham[4], A.U. Jackson[4], N.K. Saleem[5], N. Narisu[6], P.S. Chines[6], P. Gargalovic[3], T. Kirchgessner[3], F.S. Collins[6], M. Boehnke[4], M. Laakso[5], A.J. Lusis[2], K.L. Mohlke[1].* 1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Department of Medicine, University of California, Los Angeles, CA; 3) Bristol-Myers Squibb, Pennington, NJ; 4) Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI; 5) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 6) National Institutes of Health, Bethesda, MD.

Mapping of expression quantitative trait loci (eQTLs) can identify genetic variants that influence gene expression and subsequent clinical traits. To identify trans-acting eQTLs that affect distal (> 1 Mb) gene expression and to characterize regulatory gene networks, we analyzed subcutaneous adipose tissue from a subset of 1,381 individuals from the METabolic Syndromes In Men (METSIM) study. A total of 43,145 transcripts were assayed using the Affymetrix Human Genome U219 Array and tested for association with ~10 million SNPs genotyped using the Illumina OmniExpress array or imputed (1000 Genomes Project). Among the 427 billion SNP-transcript associations tested for trans eQTLs, 376,941 (0.00009%) exhibited significant evidence of association at $P < 5\times10^{-8}$. In addition to confirming the reported *KLF14* master regulator of adipose gene expression, we identified a potential novel master trans regulator at *SLC25A38-MYRIP-EIF1B* on chr 3. SNP rs11714871 (MAF = 0.40) was strongly associated with transcript levels of 16 distal genes ($P = 3.4\times10^{-29}$-4.1×10^{-8}) and modestly associated with transcript levels of 3 local (<1 Mb) genes, *SLC25A38* ($P = 2.5\times10^{-4}$), *MYRIP* ($P = 0.042$) and *EIF1B* ($P = 0.099$). Of the 16 distal transcripts, levels of 15 were correlated with expression of the local *SLC25A38* (|r| = 0.19-0.07; $P = 2.8\times10^{-13}$-0.014), and levels of 2 were correlated with expression of *EIF1B* (|r| = 0.07, $P = 0.011$). Of these 19 distal and local transcripts, *EIF1B*, *ATXN7L2* and *CD164* were associated with levels of triglycerides and LDL-cholesterol ($P = 0.025$–0.040). We identified a second novel trans-eQTL near *CIITA* on chromosome 16. SNP rs11074932 (MAF = 0.36) was associated with the transcript levels of four class II major histocompatibility complex (MHC) genes on chr 6 and *CD74* on chr 5 ($P = 7.1\times10^{-15}$-1.1×10^{-10}). The rs11074932 allele associated with higher expression levels of these transcripts was associated with higher *CIITA* expression ($P = 1.1\times10^{-4}$). CIITA is a transcriptional coactivator long known as a master regulator for class II MHC gene transcription, thus our result, which may reflect adipose tissue or infiltrated lymphocytes, supports a genetic influence on this regulatory control. These findings of trans regulatory effects should help construct regulatory networks in adipose tissue that may lead to an improved understanding of disease etiology.

650M

Long intergenic non-coding RNA eQTLs are enriched for complex trait-associated SNPs and do not distally regulate the expression of protein-coding genes. *I. McDowell[1], C. Guo[2], G. TEx Consortium[3], A. Pai[4], C. Brown[5], T. Reddy[3], B. Engelhardt[3].* 1) Computational Biology & Bioinformatics, Duke University, Durham, NC; 2) University Program in Genetics and Genomics, Duke University, Durham, NC; 3) Biostatistics & Bioinformatics, Duke University, Durham, NC; 4) Biology, MIT, Boston, MA; 5) Department of Genetics, University of Pennsylvania, Philadelphia, PA.

Long intergenic non-coding RNAs (lincRNAs) participate in a variety of biological processes including mammalian X-chromosome inactivation, telomere maintenance, adipogenesis, and cancer progression. Expression quantitative trait loci (eQTL) studies have proved effective in characterizing the genetic regulation of protein-coding genes but have largely neglected lincRNA genes for a number of reasons. Here, we performed a genome-wide association study to identify regulatory variants associated with lincRNA and protein-coding gene expression in the Genotype-Tissue Expression (GTEx) pilot data. Genotypic variation at 10 M single nucleotide polymorphisms (SNPs) including SNPs imputed from over 3.5 M quality-filtered SNPs was associated with the expression of 4000 lincRNA genes and 18,000 protein-coding genes as quantified by RNA-seq. We performed association mapping in four tissue types independently by an additive effects Bayesian regression model and across all four tissues jointly using a multi-tissue Bayesian regression model. We identified cis-lincRNA-eQTLs (linc-eQTLs) associated with 1,566 lincRNA genes (36%) and cis-protein-coding eQTLs (pc-eQTLs) associated with 9,794 protein-coding genes (54%) at a 5% FDR. We found substantial sharing of linc-eQTLs and pc-eQTLs across tissues, suggesting that the regulation of both lincRNA and protein-coding gene expression relies on regulatory elements that are similarly active across tissues. Nevertheless, linc-eQTLs were more tissue-specific overall than pc-eQTLs after controlling for tissue-specific expression levels, suggesting greater tissue-specificity of linc-eQTLs relative to pc-eQTLs. While both linc-eQTLs and pc-eQTLs were enriched for linkage to complex trait-associated SNPs, pc-eQTLs had greater relative enrichment. These results suggest that genetic variants that alter lincRNA expression may contribute to genetic risk for human diseases. For example, we identified several lincQTLs in adipose tissue in genomic regions associated with adiponectin levels and with anthropomorphic measures of adiposity in adults and newborns. LincRNAs have been speculated to function mainly in the transcriptional regulation of protein-coding genes. We tested this hypothesis by using cis-linc-eQTLs and cis-pc-eQTLs as candidate sets for trans-eQTL association. Through these analyses we did not find support for the hypothesis that lincRNAs play a substantial role in distal regulation of transcription of protein-coding genes.

651T

Massively parallel identification of non-coding causal alleles driving genetic associations. *R. Tewhey[1,2], K.G. Andersen[1,2], E.A. Brown[2,3], S. Winnicki[1,2], T.S. Mikkelsen[2], P.C. Sabeti[1,2].* 1) Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA; 2) Broad Institute, Cambridge, MA; 3) Department of Human Evolutionary Biology, Harvard University, Cambridge, MA.

Although genome-wide association studies (GWAS) have implicated thousands of common variants in various human diseases, our ability to localize association signals to individual causal alleles - and thereby elucidate mechanisms of pathogenesis - remains poor. This is partially due to our inadequate understanding of the effects of non-coding variation, which includes the majority of disease associations. To address this, we extended the use of the massively parallel reporter assay (MPRA) as a tool for the identification of non-coding causal alleles. We introduced modifications to MPRA allowing us to empirically measure the regulatory potential of 29,000 single nucleotide polymorphisms (SNPs) and the differential allelic effects at those sites in a single experimental setup. To evaluate the assay, we tested all variants in perfect linkage disequilibrium with the highest-ranking SNP for each of 3,992 eQTLs found in the 1000 Genomes lymphoblastoid cell lines (LCLs). In addition, we performed more comprehensive testing at eQTL peaks overlapping 163 variants from the NHGRI GWAS catalog. Overall, we successfully interrogated >95% of the targeted loci and detected thousands of regions that enhance expression activity. The identified regulatory loci are highly enriched for DNase hypersensitivity sites, as well as for histone marks predicting enhancer and promoter activity. At the same time, a substantial proportion of regulatory sequence remains unaccounted for by current markers. We identified putative causal alleles for hundreds of eQTL peaks and show strong correlation with traditional luciferase reporter assays. We will also discuss in detail several instances where the causal allele localizes to known disease-associated loci. Having experimentally identified regulatory alleles in two separate 1000 genome cell lines, we evaluated the same regions in a hepatocyte cell line, thus providing direct evidence of hundreds of tissue-specific regulatory elements. This work demonstrates the application of MPRA to functional genetics, allowing us to pinpoint causal alleles and investigate the molecular mechanisms underlying disease/trait associations.

652S

Uncovering expression variability and eQTLs on the X chromosome. *K. Kukurba[1,2], A. Battle*[3], S.M. Montgomery*[1,2].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Pathology, Stanford University, Stanford, CA; 3) Department of Computer Science, John Hopkins University, Baltimore, MD.

The X chromosome plays a distinctive role in genotype-phenotype relationships. However, many existing genetic association and eQTL studies have entirely disregarded it due to the special challenges of conducting analyses on the X chromosome; in particular, accounting for properties including random X-inactivation, dosage compensation, and distinct modes of inheritance relative to autosomes. To deal with these challenges and to better identify genotype-phenotype patterns on the X chromosome, we compared the utility of different statistical models for eQTL discovery on the X chromosome. By applying well-performing models, we studied X chromosome eQTL within the Depression Genes and Networks (DGN) study cohort of 922 genotyped individuals with RNA-sequencing data from whole blood. In the male and female populations, we detected a high proportion of shared eQTLs (85% of X-chromosome eQTLs are shared between sexes) as well as identified sex-specific eQTLs on the X chromosome. We further characterized the properties of these eQTLs by comparing effect sizes and evidence of selective constraint between the X chromosome and autosomes, specifically evaluating variants in lower frequency ranges. In addition, we investigated the contributions of sex to overall expression variance across the transcriptome. We discovered that females have higher variance in expression on the autosomes. Conversely, we found that males have higher expression variance on the X chromosome than females, a pattern expected in males where hemizygosity results in more extreme effects arising from genetic variation. Lastly, we examined the phenomenon of random X-inactivation in females across different ages. We found that older individuals typically have more allelic skewing on the X chromosome compared to younger individuals; an observation which may be explained by clonal selection in the hematopoietic stem cell population. Together, this work advances our understanding of how sex and eQTLs shape human expression variation on the X chromosome contributing to sexual dimorphism.

653M

Universal eQTL: discovery and replication across cell type and population. *S.A. Shenoy[1], T.L. Vincent[1], R.G. Crystal[1], J.G. Mezey[1,2].* 1) Department of Genetic Medicine, Weill Cornell Medical College, New York, NY; 2) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY.

Expression Quantitative Trait Loci (eQTL) describe a fundamental aspect of biology: inherited allelic variation that impacts gene expression. Universal eQTL (e_uQTL) describe cases where the same eQTL relationship, indicated by repeated identification of an association between genotypes in the same genomic region and expression of the same gene, is almost always present regardless of the tissue analyzed or ancestry of the sample population. These e_uQTL are inherently interesting because they indicate a strong connection between the regulation of gene expression and the genetics that influence expression, and because they represent cases of super-replication that can be leveraged to infer the genetic causes and impacts of e_uQTL. While the existence of e_uQTL has been indirectly acknowledged there has not been a targeted effort to discover e_uQTL.

Making use of an eQTL detection pipeline developed in our lab capable of independently testing billions of genotype-expression associations in a few hours, we analyzed >35 publicly available and new datasets, including HapMap lymphoblastoid cell lines and cancer/healthy tissues in the TCGA catalog, where together our combined analysis included over 20 distinct tissue/cell types and 8 distinct ancestry groups. For each dataset, we independently tested every genotype-expression variable measured by microarray or RNA-Seq genome-wide after minimal filtering. By employing stringent false-positive controls and data fitting metrics, we identified 22 e_uQTL that were in almost every dataset and 64 additional e_uQTL that replicate across natural groupings of datasets. Most of these were cis-e_uQTL, where a number of trans-e_uQTL were identified when considering less conservative multiple test corrections. Taking advantage of e_uQTL replication across populations, we leveraged patterns of linkage-disequilibrium to narrow the genomic region containing the causal genotype(s) to <1–5 kb in a number of cases, a "drilling down" approach that is only possible with e_uQTL. The result of this work is a picture of the nature and generality of e_uQTL, the causal polymorphisms responsible, and the regulatory impacts of these e_uQTL, which together provide an understanding of why e_uQTL exist, their regulatory and genetic basis, and their downstream impacts.

654T

Sex-biased genetic effects on the transcriptome of monocytes and T-cells. *N.L. Tignor[1], T. Raj[2,3,4,5], E.R. Gamazon[1], K. Rothamel[6], P.L. De Jager[2,4,5], B.E. Stranger[1,7], The ImmVar Consortium.* 1) Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL; 2) Program in Translational NeuroPsychiatric Genomics, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, MA; 3) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 4) Harvard Medical School, Boston, MA; 5) The Broad Institute of MIT and Harvard, Cambridge, MA; 6) Division of Immunology, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA; 7) Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL.

Sexual dimorphism is observed across a variety of complex phenotypes in model organisms and humans, including immune function in healthy individuals, as well as autoimmune disorders and neuropsychiatric disorders, among others. In model organisms, and to a lesser extent in humans, sex-specific quantitative trait loci have been identified for sexually dimorphic traits. To extend our understanding of the contribution of sex to the genetic basis of human immune function and dysfunction, we performed a sex-aware expression quantitative trait locus (eQTL) study of highly purified CD4(+) T cells and monocytes, representing adaptive and innate immunity, in a cohort of 212 healthy individuals, as part of the Immunological Variation (ImmVar) Project. We estimate that of genes with significant cis-eQTLs in each sex, less than 10% are unique to one sex, with the degree of sex bias approximately equal in each of the two cell types. Where associations are shared across sexes, we see strong correlation between effect sizes of male- and female eQTLs (r^2=0.94, CD4; r^2=0.87, CD14), suggesting little in the way of sex-specific modification of allelic effects. We do not observe cases where cis-eQTLs are shared across sexes but show opposite direction of allelic effects across sexes. We do, however, observe cis-eQTL allelic direction changes across cell-types, as has been reported previously. Although the degree of sex-bias between males and females is similar across the cell-types, we observe tissue-specificity in the genes with sex-biased eQTLs. Functional analysis of sex-specific eQTLs within cell types identifies different categories of significant enrichment, with sex-specific eQTL genes enriched in antigen and immune function in CD14, and glutathione and oxidative stress in CD4. We identified genetic variants previously associated with complex traits in human diseases as cataloged in the NHGRI catalog of genome-wide association studies that are also significant sex-specific eQTLs in our datasets. These include variants associated with diseases that exhibit sexually dimorphic characteristics, including Crohn's disease, bipolar disorder, and rheumatoid arthritis. Our results characterizing the sex-specific genetic architecture of the transcriptome in these immune cells will be compared and contrasted to our parallel analyses of the same cell-types in populations of different ancestry, as well as in tissues of the Genotype-Tissue Expression (GTEx) Project.

655S

ERAL1 Stabilization of miRNA: A Novel Regulatory Mechanism of Gene Expression. *K. Aquino-Michaels[1], V. Trubetskoy[1], E.R. Gamazon[1], M.G. Larson[3], J. Freedman[3], J. Rong[3], S. Lacey[3], T. Huan[3], D. Levy[3], R.S. Huang[2], N.J. Cox[1], H.K. Im[1].* 1) Department of Medicine, Section of Genetic Medicine, University of Chicago, Chicago, IL; 2) Department of Medicine, Section of Hematology/Oncology, University of Chicago, Chicago, IL; 3) Framingham Heart Study, Population Sciences, Branch National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD.

Genetic mechanisms of gene regulation create a rich source of variation between individuals and are essential to complex traits and biological function. These mechanisms can influence local gene transcription or act in trans through complex regulatory interactions. Trans gene regulation can occur through small non-coding RNAs such as microRNAs that act as regulatory hubs for many downstream targets. While cis-acting regulation of microRNA expression shares similar mechanisms with cis-regulation of messenger RNA (mRNA), trans regulation of microRNA remains poorly understood. Furthermore, currently only a handful of trans-microRNA quantitative trait loci (mirQTLs) have been reported and fewer still have been replicated. To systematically identify mechanisms of trans-microRNA expression, we conducted mirQTL mapping in 4902 participants from the Framingham Heart Study--the largest single mirQTL cohort to date. We discover widespread trans-regulation of microRNAs with 720 mirQTLs (q-value <= 0.05) mapping to 90 microRNA transcripts. Of the 720 mirQTLs, 189, corresponding to 78 SNPs and 8 microRNAs, are clustered within 100kb upstream of the gene ERAL1. We replicated 22 of the 189 mirQTLs at the ERAL1 locus (q-value <= 0.05) using the gEUVADIS cohort (n=334). In addition, of the 78 SNPs at this locus we found that 74 are cis-eQTLs for ERAL1 exon specific probes, suggesting coregulation of ERAL1 and microRNA. ERAL1 is an RNA binding GTPase that acts as a ribosomal RNA chaperone. We therefore hypothesize that cis-eQTLs of ERAL1 act as trans-mirQTLs for microRNA through stabilization of RNA structures that in turn regulate downstream mRNA.

656M

Genome-wide Identification of microRNA Expression Quantitative Trait Loci in the Framingham Heart Study. *T. Huan[1,2], J. Rong[3], C. Liu[1,2], X. Zhang[1,2], K. Tanriverdi[4], R. Joehanes[1,2,5], B. Chen[1,2], J. Murabito[3], C. Yao[1,2], P. Courchesne[1,2], P. Munson[5], C. O'Donnell[1,2], N. Cox[6], A. Johnson[1,2], M. Larson[3], D. Levy[1,2], J. Freedman[4].* 1) The National Heart, Lung, and Blood Institute's Framingham Heart Study, 73 Mt. Wayte Avenue, Framingham, MA, USA; 2) The Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD, USA; 3) Department of Mathematics and Statistics, Boston University, Boston, MA, USA; 4) Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA; 5) Mathematical and Statistical Computing Laboratory, Center for Information Technology, National Institutes of Health, Bethesda, MD, USA; 6) Department of Human Genetics, University of Chicago, Chicago, IL, USA.

Introduction: Identification of miRNA expression quantitative trait loci (miR-eQTL) may yield insights into regulatory mechanisms of miRNA transcription, and thus may help elucidate important roles of miRNAs as mediators in complex traits. Several published miR-eQTL studies were based on modest sample sizes (n<200). These studies revealed a few cis-miR-eQTLs, but uncertainty persists regarding the number of miR-eQTLs and their relations to regulatory elements in the human genome. Methods: We analyzed the associations of approximately 10 million 1000 Genomes Project imputed SNPs (at minor allele frequency >0.01) with whole blood miRNA expression levels measured by Taqmen qPCR using a Fluidigm platform in 5239 Framingham Heart Study (FHS) participants. Associated SNPs residing within 1Mb of the mature miRNA (cis) and those occurring more than 1 Mb away (trans) were identified separately using an additive regression model adjusted for age, sex, and family structure. Replicability of the identified cis and trans miR-eQTLs was tested by splitting the overall samples 1:1 into pedigree independent sets. Results: 280 miRNAs were expressed in >200 individuals and were used for analysis. At a false discovery rate <0.1, we identified 5269 cis miR-eQTLs (representing 982 loci at linkage disequilibrium r2<0.7) for 76 mature miRNAs. cis miR-eQTLs showed high concordance across both pedigree independent sets (the concordance rate is 50%). We found that most cis miR-eQTLs were located upstream of their associated miRNAs (57% for intragenic and 84% for intergenic miRNAs). 59% of cis miR-eQTLs were located 300–500kb upstream of their associated intergenic miRNAs, suggesting that distal regulatory elements may affect interindividual variability in miRNA expression levels. We also found that cis miR-eQTLs were highly enriched for cis mRNA eQTLs and regulatory SNPs (e.g., SNPs in promoter, enhancer, and transcription factor binding sites annotated by ENCODE). By cross-linking cis miR-eQTL SNPs with GWAS SNPs and by linking cis miR-eQTL miRNAs with differentially expressed miRNAs for complex traits we identified 245 cis miR-eQTLs that are associated in GWAS with complex traits. Several of the cis miR-eQTL miRNAs displayed differential expression in relation to the corresponding GWAS trait in the FHS. Conclusion: Our study provides a roadmap for understanding the genetic basis of miRNA expression, and sheds light on miRNA involvement in a variety of complex traits.

657T

Dissecting the genetic regulation of exosome RNA cargo in a large family. *E.K. Tsang[1,5], X. Li[2,5], V. Anaya[2], K.J. Karczewski[1], D.A. Knowles[3], K.S. Smith[2,6], S.B. Montgomery[2,4,6].* 1) Biomedical Informatics Program, Stanford University, Stanford, CA., United States; 2) Department of Pathology, Stanford University, Stanford, CA, United States; 3) Department of Computer Science, Stanford University, Stanford, CA, United States; 4) Department of Genetics, Stanford University, Stanford, CA, United States; 5) Co-first authors; 6) Co-senior authors.

Many studies have characterized genetic variation affecting gene expression through mapping of expression quantitative trait loci (eQTLs) in a variety of isolated cell types and conditions. Meanwhile, little attention has been paid to the impact of such genetic variation on intercellular communication. Exosomes are one mechanism through which cells transfer information. They are small extracellular vesicles that package biomolecules, including RNA and protein, from their cell of origin, and can deliver this cargo to influence neighboring cells. Exosomes are produced by many different cell types and are thought to be involved in the progression of certain diseases, including cancer, as well as being important for maintaining homeostasis. However, while exosomes may be relevant to several biological processes, the regulation of exosome contents remains poorly understood. In this work, we sequenced the small RNA transcriptomes of lymphoblastoid cell lines and their exosomes from a seventeen-member three-generational family. We investigated which transcripts were specifically exported in exosomes or retained in cells through differential expression analyses. By combining publicly available whole genome sequence data for the seventeen individuals with our RNA sequence data, we mapped eQTLs in both cells and exosomes and evaluated the amount of sharing between the two groups.

For the different types of small RNAs we studied, between 13% and 40% of genes were differentially expressed between cells and exosomes. This finding suggests that the RNA contents of exosomes are actively regulated and are not simply a reflection of the cellular transcriptome. We also searched for motifs enriched in the transcripts more highly expressed in exosomes than in cells to find evidence of export signals. To characterize the effect of genetic variability on exosome contents, we performed the first eQTL study in exosomes. We found evidence of sharing of small RNA eQTLs between cells and exosomes, but also identified cell- and exosome-specific eQTLs. Since exosomes are mediators of intercellular communication, our characterization of eQTL in exosomes highlights the impact of regulatory variants beyond the level of individual cells.

658S

Analysis of hypothalamus transcriptome and proteome in 100 strains of mice on high fat diet. *Y. Hasin[1], A. Khan[2], V.A. Petyuk[3], B.W. Parks[1], C.D. Rau[1], C. Pan[1], P.D. Peihowski[3], R.D. Smith[3], A.J. Lusis[1], D.J. Smith[2].* 1) David Geffen School of Medicine, Cardiology, UCLA, Los Angeles, CA; 2) Department of Molecular & Medical Pharmacology, University of California, Los Angeles; 3) Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington.

Regulation of body weight and appetite are multifaceted processes, in which hypothalamic nuclei play a pivotal role. We here report the first integrated transcriptome and proteome analysis in the mouse hypothalamus, using paired end RNA sequencing and LC-MS Proteomics Analysis in 100 inbred mouse strains from the Hybrid Mouse Diversity Panel (HMDP). HMDP strains allow high resolution mapping and we were able to identify thousands of new transcripts and isoform variants as well as mapping 283 pQTLs. We are currently constructing gene co-expression network from these data, using linear and non linear correlation measurements. In addition these mice were extensively phenotyped for various of metabolic phenotypes, such as plasma metabolites and lipids, which allows us to associate gene and protein expression with particular phenotypes.

659M

Quantifying Context Specificity of Gene Regulation using Predicted Gene Expression Levels. *S.V. Mozaffari[1,2], E.R. Gamazon[2], K. Aquino-Michaels[2], N.J. Cox[2], H.K. Im[2].* 1) Committee of Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL., USA; 2) Department of Medicine, Section of Genetic Medicine, University of Chicago, Chicago, IL.,.

Genetic regulation of gene expression is likely to play an important role in the etiology of complex traits as indicated by the enrichment of expression-associated variants (eQTLs) and DNase hypersensitivity sites among trait-associated variants. Tissue specificity and other context dependent regulation of the transcriptome is an area of current interest. We propose a method to quantify the degree of context specificity by using prediction models trained and tested in different contexts. For example, we examined the context specificity of gene regulation by quantifying the ability to predict gene expression levels using models trained in whole blood, tibial nerve, and muscle, and tested on expression levels assayed in lymphoblastoid cell lines. We generated predictive models of gene expression levels using single variant regression coefficients. SNPs significantly associated with gene expression levels were kept in an additive model with p-value threshold of 0.05 for variants in the vicinity of the gene and 10-6 for distant ones. We used expression data from whole blood, tibial nerve, and muscle from the GTEx pilot data for training and independent lymphoblastoid cell lines (LCL) from 1000 Genomes as test set. When using whole blood models, we found that for 8.5% of the genes, the predicted levels were significantly associated with the observed levels in LCLs with FDR <5%. For 15% of the genes, correlation was better than 10%. ERAP2 was the best predicted gene with correlation between predicted and observed of 0.87. When using models trained with tibial nerve and muscle data, 7.5% and 7.4% of the genes show correlation with FDR less than 5%. Performance is slightly inferior to whole blood indicating more similarity between LCL and whole blood than with tibial nerve or muscle. As expected, we found that genes with higher heritability, higher expression levels, and higher variability tended to be better predicted. In conclusion, we propose a novel prediction-based method to quantify context specificity of gene regulation and find that a substantial portion of the regulatory mechanism is likely to be common across tissues.

660T

Genetic architecture of the transcriptome of four tissues in a twin cohort. *A. Buil[1], A. Viñuela[2], A.A. Brown[3], M. Davies[2], K. Small[2], R. Durbin[3], T.D. Spector[2], E.T. Dermitzakis[1].* 1) Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 3) Wellcome Trust Sanger Institute, Hinxton, United Kingdom.

Understanding the genetic architecture of gene expression is an intermediate step to understand the genetic architecture of complex diseases. To explore the underlying causes of gene expression regulation we sequenced the mRNA fraction of the transcriptome of ~400 female MZ and DZ twin pairs from the TwinUK cohort in four tissues: fat, skin, blood and lymphoblastoid cell lines (LCLs) (2330 samples in total) and used genotype information imputed into the 1000 Genomes Phase 1 reference panel. Using a linear regression approach with SNPs in a 1Mb window each side of the TSS for each gene and tissue we identified 9166 significant cis-eQTLs in fat, 9551 in LCLs, 8731 in skin and 5313 in blood (1% FDR). But cis-eQTL are only a small part of the genetic effects that affect gene expression. By exploiting the twin structure of our sample, we dissected the proportions of gene expression variation which is due to different genetic and non-genetic causes. We observed that, on average, common cis-eQTL only explained about a 20% of the heritability of gene expression while other genetic variants in cis (mainly rare variants or common variants with small effects) explained about 30% of heritability. The remaining 50% of the heritability was explained by genetic variants in trans. Moreover, we observed that between 70% and 80% of the SNPs that were cis-eQTL in one tissue have at least a weak cis-eQTL signaling another tissue. Genetic variation may also affect gene expression by modifying mRNA splicing processes. We calculated the association of cis SNPs and the frequencies of exon-exon links per individual and identified between 1566 and 4104 splicing QTLs (sQTLs) per tissue. We observed that between 35% to 60% of the sQTLs replicate among tissues, highlighting the importance of tissue specific effects in alternative splicing. Finally, we looked for genetic variants that have a cis effect in one tissue and, at the same time, a downstream effect, in trans, in another tissue. We selected cis-eQTLs that are active in a single tissue (tissue 1) and looked for its associations with trans genes in another tissue (tissue 2). Then, we used causal model inference to test that the association between the eQTL and the trans gene in tissue 2 is more likely mediated through effects from the cis-eQTL in tissue 1 than through a cis effect of the eQTL in tissue 2. We found several examples of these cross-tissue effects and we are now in the process of validating them.

661S

Genome-Wide Analysis of Expression Short Tandem Repeats. *M. Gymrek[1,2,3], S. Georgiev[4], B. Markus[1], D. Zielinski[1], J. Pritchard[5,6], Y. Erlich[1].* 1) Whitehead Institute, Cambridge, MA; 2) Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, MA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Department of Human Genetics, University of Chicago, Chicago, IL; 5) Departments of Genetics and Biology, Stanford University, Palo Alto, CA; 6) Howard Hughes Medical Institute, Chevy Chase, MD.

A central goal in genomics is to elucidate the genetic architecture of complex traits. So far, efforts to discover eQTLs (expression Quantitative Trait Loci) have mainly focused on the contribution of SNPs to gene expression. Several candidate gene studies in human and model organisms have suggested that Short Tandem Repeat (STR) variations can modulate expression levels and splicing of nearby transcripts. Here, we report the first genome-wide survey to identify STRs that regulate the transcriptome. We analyzed STR variations across hundreds of samples from the 1000 Genomes Project using a custom algorithm. We then performed association tests between multi-allelic STR genotypes and expression levels of nearby genes while controlling for SNPs in linkage disequilibrium. This process identified significant associations between STR variations and expression profiles (eSTRs) of more than 1,000 genes. These associations were replicated across populations and in orthogonal expression assays (RNA-sequencing and expression array). Predicted eSTRs are found near transcription start sites and near splice junctions, and are strongly enriched in annotated regulatory elements and regions under purifying selection. We used a variance components analysis to dissect the contribution of STRs to gene expression. This analysis showed that eSTRs explain on average about 22% of the cis-heritability of the expression levels after controlling for nearby SNPs. Overall, our results highlight the contribution of STRs to the genetic architecture of gene expression and raise the possibility that these variants may contribute to the missing heritability of complex traits.

662M

Cross-platform validation and cross-tissue activity of expression quantitative trait loci (eQTL). *A.J. Jasinska[1], S. Service[1], I. Zelaya[1], C. Blum[1], Y. Huang[1], L. Navarro[1], M.J. Jorgensen[2], G. Weinstock[3], R. Wilson[4], W. Warren[4], E. Eskin[5], R. Woods[6], G. Coppola[1,6], N.B. Freimer[1].* 1) UCLA Center for Neurobehavioral Genetics, University of California, Los Angeles, CA; 2) Department of Pathology, Section on Comparative Medicine, Wake Forest School of Medicine, Winston-Salem, NC; 3) The Jackson Laboratory for Genomic Medicine, Farmington, CT; 4) The Genome Institute, Washington University School of Medicine, St Louis, MO; 5) Department of Computer Science and Human Genetics, The University of California, Los Angeles, CA; 6) Department of Neurology, The University of California, Los Angeles, CA.

Genetic regulation of transcript levels by expression quantitative trait loci (eQTL) is a key contributor to the regulatory mechanism underlying higher order traits. Therefore, eQTL can be useful for interpreting and determining the biological relevance of trait mapping results. However, to improve the interpretability of the eQTL results some questions still need to be addressed and answered. For example: How replicable are eQTL, especially trans-acting or distant eQTL? Which eQTL act in a tissue-specific manner, and which act across tissues? In order to provide more insight into gene regulation organization, we conducted a cross-platform and cross-tissue comparison of eQTL signals using data from two gene expression platforms, RNA-seq and expression microarrays, in a genetically-characterized pedigree of vervet monkeys (Chlorocebus aethiops sabaeus) obtained from the Vervet Research Colony (VRC). For eQTL mapping of steady-state transcript levels, we used two data sets: 1) RNA sequencing (RNAseq) in a matched set of five tissues (blood, skin fibroblasts, caudate, adrenal, and pituitary) from each of 35 monkeys (a total of 175 samples), and 2) microarray gene expression in peripheral blood samples from 327 monkeys. The gene expression data was inverse-normal transformed. We used genotype data from a genetic mapping set consisting of 500K common autosomal SNPs that was generated based on whole genome sequencing (WGS) from the VRC monkeys. For association analysis between genotypes and gene expression levels, we used the EMMAX statistical test, using age, sex, batch and pregnancy status as covariates. Among the 5,539 gene expression probes that passed our filter for high quality probes, 3,134 probes (57%) were heritable at FDR<0.01 (SOLAR). Among the heritable probes, we identified a Bonferroni-significant association signal for 281 probes in cis and 37 probes in trans (i.e., on a different chromosome than a probe). The trans signals include two eQTL hotspot regions where five or more probes (gene expression) were mapped (on chromosomes 5 and 9). Forty percent of the trans-regulated probes (15/37) were associated to the same trans SNPs in the RNAseq dataset at the significance level of p<0.05. Despite a small sample size with the RNAseq data, we identified 59–238 genes per tissue that were regulated by eQTL (p<5e-08). That includes genes regulated across tissues: LRRC37A3 in five tissues, three genes in three tissues, and 23 genes in two tissues.

663T

Transcriptome sequencing of a large human family identifies the impact of rare non-coding variants. *X. Li[1], A. Battle[2,3], K.J. Karczewski[2], Z. Zappala[2], K.S. Smith[1], K.R. Kukurba[2], E. Wu[1], N. Simon[4], S.B. Montgomery[1,2,3].* 1) Department of Pathology, Stanford University School of Medicine; 2) Department of Genetics, Stanford University School of Medicine; 3) Department of Computer Science, Stanford University; 4) Department of Biostatistics, University of Washington.

Recent and rapid human population growth has led to an excess of rare genetic variants that are expected to contribute to an individual's genetic burden of disease risk. To date, much of the focus has been on rare protein-coding variants where potential impact can be estimated from the genetic code, while determining the impact of rare non-coding variants has been more challenging. To improve our understanding of such variants, we combine high-quality genome sequencing and RNA-sequencing data within a 17-individual, three generation family to contrast expression and splicing quantitative trait loci (e/sQTLs) within a single family to e/sQTLs within a population sample. Using this design, we identify substantial numbers of genes exhibiting larger effect-sizes within the family than the population. We further observe that those e/sQTLs with large effects in the family are enriched for rare regulatory and splicing variants (MAF<0.01) and they are also more likely to influence essential genes and genes involved in complex disease. In addition, we found rare variants are much more likely to cause e/sQTLs than common variants given same functional annotations. We found that distance to the transcription start site, evolutionary constraint and epigenetic annotation are considerably more informative for predicting e/sQTL of rare variants than common variants. We postulate that lack of selective pressure increased capability for genomic annotation to predict the impact of those rare non-coding variants. These results highlight that rare non-coding variants are significant contributors to individual gene expression profiles, especially to large and potentially deleterious effects.

664S

Allele specific expression and eQTL in diploid genomes. *S.C. Munger, N. Raghupathy, K. Choi, D.M. Gatti, P. Simecek, G.A. Churchill.* The Jackson Laboratory, Bar Harbor, ME.

The emergence of high throughput sequencing (HTS) technologies has coincided with the development of advanced genetic reference populations including the mouse Diversity Outbred (DO) heterogeneous stock. The application of HTS to genetically diverse mapping populations has the potential to provide nucleotide resolution of causal variants underlying phenotypic differences. However the increase in information content comes at the cost of increased analytical complexity. We have developed novel methods and software to exploit the high genetic diversity and heterogeneous diploid genomes of the DO to yield new layers of information and inform fine mapping of phenotypic and expression QTL (eQTL). Importantly, these methods can be readily applied to any human expression dataset where genotyping data is available.

RNA-seq alignment to individualized diploid genomes yields direct, accurate estimates of allele specific expression (ASE), and improves eQTL identification and resolution. To illustrate this, we profiled the liver transcriptomes of 450 DO mice by RNA-seq and estimated gene, isoform, and allele expression. We correlated gene expression differences to genetic variation and identified 9,000 local expression QTL (eQTL) and 900 distant eQTL, showing that most of the variation in transcript abundance derives from segregating local genetic variation. Allelic expression differences confirmed that cis-acting mechanisms underlie most local eQTL, and DO allele estimates correlate well with gene expression in livers from the eight DO founder strains. Most eQTL appear biallelic suggestive of a single causal variant, however complex 3- and 4- allele patterns are observed. Cis-eQTL with allelic expression patterns that deviate from the known strain ancestry are most amenable to fine mapping, and in some cases we predict a single causal variant. Distant eQTL with large effects are rare in the adult liver transcriptome, however we have developed conditioning methods to amplify trans effects and test candidate regulatory genes in the interval.

Seqnature software for constructing individualized diploid genomes and gene annotation files is available at github.com/jaxcs/Seqnature. Analytical pipelines described above can be found at do.jax.org. All eQTL results are available as an interactive web application at do.jax.org.

665M

An integrated framework for evaluating the pathogenicity of rare, population-specific non-coding variation. *Z. Zappala[1], M. Pala[2,6,7], M. Marongiu[2], A. Mulas[2], R. Cusano[2], F. Crobu[2], F. Reinier[3], R. Berutti[3,6], M.G. Piras[2], C. Jones[3], D. Schlessinger[5], G. Abecasis[4], A. Angius[2], S. Sanna[2], F. Cucca[2,6,8], S.B. Montgomery[1,7,8].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Research Institute of Genetics and Biomedicine, Monseratto, Italy; 3) CRS4, Advanced Genomic Computing Technology, Pula, Italy; 4) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 5) Laboratory of Genetics, National Institute on Aging, Baltimore, MD, USA; 6) Department of Biomedical Science, University of Sassari, Sassari, Italy; 7) Department of Pathology, Stanford University, Stanford, CA; 8) co-senior authors.

Large-scale genome sequencing efforts have revealed a great diversity of coding and non-coding genetic variation. Identifying the impact of rare, non-coding genetic variation on phenotypic variation in a population remains an open challenge in population genomics. In the absence of large sample sizes, rare variants do not have an appreciable frequency to power traditional expression quantitative trait loci (eQTL) association tests. Using population-scale full genome and RNA-Seq data for 624 Sardinian individuals, we have developed a framework for combining population structure, genetic background, gene expression, and allele-specific expression (ASE) to find rare variants that are associated with family-specific patterns of gene regulation. Our analysis of 66 trios in the Sardinian population reveals a number of gene expression traits that co-segregate in a family with significant signals of ASE. These signatures are consistent with the model of a rare, causal *cis*-regulatory element that also segregates within the family. Within these outlier individuals, we test for rare non-coding variants that overlap with functional annotations such as conservation, DNase hypersensitivity, and transcription factor binding. We further integrate these variants with genome-wide association study (GWAS) hits and disease annotations to identify rare variants that may contribute to complex diseases. By utilizing this pipeline, we have identified rare variants in the Sardinian population that contribute to family-specific regulatory traits and hold pathogenic potential. We have additionally designed this pipeline so that it will work with any large-scale genome and RNA sequencing project, allowing the general integration of these genome-wide datasets to reveal rare variant associations with complex traits and diseases.

666T

Modeling uncertainty in RNA-seq analysis: Beyond differential expression. *G.E. Hoffman, B.A. Kidd, E.E. Schadt.* Icahn Institute for Genomics and Multiscale Biology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

RNA-seq is becoming the standard technique for assessing gene expression in medical and experimental genomics. The large dynamic range and ability to capture alternative splicing make RNA-seq a rich data source for asking an array of biological questions beyond the standard analysis of differential expression. Yet efforts to address complex biological questions must consider the complexity of RNA-seq data. For example, typical RNA-seq data is generated by sequencing RNA molecules, mapping reads to a reference genome, and counting the number of reads corresponding to each gene. The count nature of the data imply that RNA-seq has a complex error structure where the sampling variance increases with number of read counts per gene. Widely used methods for differential expression have established that appropriately modeling the error structure is essential for maintaining power and reducing false positives. Yet other applications of RNA-seq make strong simplifying assumptions and do not address the complex error structure. Here we consider the general question of how to best utilize RNA-seq data to perform complex analysis such as eQTL detection, principal components analysis (PCA) and coexpression network reconstruction. To date these applications have been applied to processed count data, where processing involves transforming the counts for each gene based on rank (i.e. quantile transformation) or read fraction (i.e. RPKM/FPKM, count per million (CPM)), which do not directly model the error structure. We consider transformations based on a statistical model accounting for uncertainty in counts (i.e variance stabilizing transform, precision-weighted CPM) and demonstrate that the processing of the count data affects the quality of the results. Using simulations, public data from the GEUVADIS project and disease datasets, we illustrate how modeling the error structure of RNA-seq data affects power in cis-eQTL detection, power of PCA to distinguish between know subgroups, and stability of inferred coexpression models. This work proposes a unified framework for addressing an array of biological questions by modeling the complexity of RNA-seq data.

667S

Meta-Analysis of Liver eQTL Studies and Cross-tissue eQTL Comparison using GTEx Data. *E.L. Seiser[1], K. Xia[2], F.A. Wright[3], F. Innocenti[1].* 1) Division of Pharmacotherapy and Experimental Therapeutics, The University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Statistics, North Carolina State University, Raleigh, NC.

Identification of liver-specific expression quantitative trait loci (eQTLs) may aid in determining genes and genetic variants relevant to the pathogenesis of complex liver-associated diseases and the biological mechanisms underlying drug pharmacology by providing a functional link between genetic variation and gene expression levels. Previous liver eQTL studies have resulted in low replication rates, suggesting the need for a larger overall sample size to increase the statistical power for identifying eQTLs. We performed a meta-analysis of four human liver eQTL studies (n= 149, 206, 427, 960), including three previously published, to identify cis-eQTLs (within 100kb of transcription start/stop site). Genotype microarray data from each study was used for determining sex, genetic ancestry, and common samples between studies, leading to a total of 1,226 Caucasian liver samples for analysis. Imputation of genotype data using Impute2 produced a set of 6,684,896 SNPs common among all studies. Probe sequences from each expression microarray platform were mapped to Entrez genes and a set of 13,669 genes common to all platforms was identified. Using expression data from each study, probabilistic estimation of expression residuals (PEER) was employed to determine hidden variance components. Matrix eQTL was used to test for cis- associations, incorporating sex, age, genetic ancestry, and PEER factors into an additive genetic model. A meta-t-statistic was calculated using SNP/gene pair t-statistics from each study, and p-values for the meta-t-statistics were generated using the normal distribution. Testing of 8,236,407 SNP/gene pairs identified significant cis-linked genetic effects on expression levels (Bonferroni corrected p-value<6.1×10–9) for 3,334 genes, including numerous genes related to drug processing. Ongoing comparative analysis of the liver eQTL data with the Genotype Tissue Expression (GTEx) project eQTL data from 9 tissues (whole blood, adipose, muscle, heart, artery, lung, skin, nerve and thyroid) will allow for a comprehensive identification of liver-specific eQTLs, and may help to further elucidate pathogenic and biological processes in the liver.

668M

Functional mapping of eQTL signals for prostate cancer risk SNPs. *L. Tillmans, S. McDonnell, A. French, Y. Zhang, S. Riska, M. Larson, Z. Fogarty, N. Larson, J. Cheville, S. Baheti, S. Middha, D. Schaid, S. Thibodeau.* Mayo Clinic, Rochester, MN.

In order to define the functional role of previously reported risk-SNPs for prostate cancer (PC), we created a normal prostate tissue eQTL database using 471 tissue samples from patients with prostate cancer. Genotyping was performed using the Illumina Human Omni 2.5 M SNP array and expression was measured by RNA sequencing. Our initial eQTL analysis focused on those risk-SNPs previously identified by multiple PC GWAS studies (n= 123), plus all SNPs (included imputed) that were in linkage disequilibrium (r2>0.5) with each risk-SNP. Additionally, we focused on cis-acting associations only testing all transcripts within 2Mb of the risk-SNP interval. Of 5116 SNPs identified in the risk-regions, 1002 demonstrated a Bonferroni significant eQTL signal (p<1.96e-7) and these were associated with 43 genes. To identify possible functional SNPs for each of the risk-regions associated with a candidate gene, we utilized a variety of publicly available databases and bioinformatics tools. These included various GEO datasets that were specific for prostate tissue or PC cell lines (with and without androgen treatment) and a variety of analysis tools (Regulome, Funcpred, GWAS3D, etc.) to obtain regulation potential scores for each of the candidate SNPs. Some of the information retrieved from these databases included transcription factor binding sites and motifs, previous eQTL data, miRNA binding, conservation and reported long range interactions. Results from these analyses were then compiled to determine possible functional regions. For example, for the risk-SNP rs4962416 on chromosome 10, there were 25 SNPs in LD that had significant eQTL signals (all p-values <3.71E-12) with the CTBP2 gene (a transcriptional repressor). Of these, we were able to map the 14 most significant to a probable enhancer region as defined by markers DNase I, H3K27Ac, H3K4me1 and H3K4me2. The top variant (rs12769019) as determined by Regulome score and GWAS3D for LnCap cells is located in an AR binding site (chip-seq) and is shown to possibly disrupt the binding affinities of numerous transcription factors that could be important for prostate cancer regulation through interactions with the AR (ie NKX3-1, FOXA1, GR and POU6F1). Utilizing publically available data and bioinformatics tools provides a powerful strategy for mapping and identifying candidate functional elements for PC susceptibility loci to help guide subsequent laboratory studies.

669T

Sex-specific genetic architecture of the transcriptome. *E.R. Gamazon[1], N.L. Tignor[1], B.E. Stranger[1,2], GTEx Consortium.* 1) Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL; 2) Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL.

Sexual dimorphism has been observed in a broad spectrum of complex phenotypes, including autoimmune disorders, psychiatric disorders, and cancers. Here, in support of our efforts to investigate the sex-specific genetic architecture of complex traits, we examine potential functional mechanisms underlying sexual dimorphism. Using RNA sequencing data from the Genotype-Tissue Expression (GTEx) Consortium, we assessed sex-specific gene regulation and mapped sex-biased expression quantitative trait loci (eQTLs) in a diverse set of human tissues. We identified tissue-specific gene expression patterns with some genes showing differential expression (p<10-6) between the sexes in some tissues but not in others. Our approach facilitated the comprehensive analysis of non-coding RNAs, putatively regulatory, with substantial sexually dimorphic expression profiles. We identified eQTLs associated with the expression of a gene in one sex and not the other, quantified the proportion of shared eQTLs, and highlighted traits with distinct regulatory variation in the sexes. Notably, our unique dataset allowed us to evaluate the tissue-specificity and the degree of sharing between tissues of sex-biased eQTLs. We found in each tissue substantial variation in the presence of sex-specific eQTLs among different classes of genes, including, most prominently, antisense transcripts and long intergenic non-coding RNAs. We identified sex-specific eQTLs among known disease susceptibility and quantitative trait loci (for example, Parkinson's disease, Crohn's disease, systemic lupus erythematosus, aortic dimension, and folate pathway levels) as curated in the NHGRI catalog of published genome-wide association studies. On the basis of eQTL targets, we propose novel genes, distinct from the originally reported genes, for some of these phenotypes. Analyses of WTCCC, the Psychiatric Genomics Consortium, and GIANT GWAS data reveal that sex-biased eQTLs show tissue-dependent enrichment for trait associations. Sex-specific eQTL target genes in such tissues as tibial artery and blood are enriched for antigen processing and presentation as well as immune-related function whereas those in muscle are, in addition, enriched for metabolic processes and for proteins that localize in mitochondria. Our study shows that genotype-sex interactions have a broad influence on the human transcriptome and that sex-specific genetic architecture of gene expression is likely to mediate higher-order traits.

670S

Activating mutations in STIM1 and ORAI1 cause overlapping syndromes of tubular myopathy and congenital miosis. *M. Kousi[1], V. Nesin[2], G. Wiley[3], E.C. Ong[2], T. Lehmann[4], D.J. Nicholl[5], M. Suri[6], N. Shahrizaila[7], P.M. Gaffney[3], K.J. Wierenga[8], L. Tsiokas[2], N. Katsanis[1].* 1) Cell Biology, Duke University, Durham, NC; 2) Department of Cell Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 3) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK; 4) Department of Hematology, Center for Laboratory Medicine, Kantonsspital St. Gallen, Switzerland; 5) Department of Neurology, City Hospital, United Kingdom; 6) Nottingham Clinical Genetics Service, Nottingham University Hospitals National Health Service Trust, United Kingdom; 7) Division of Neurology, Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia; 8) Section of Genetics, Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Signaling through the store-operated Ca2+ release-activated Ca2+ (CRAC) channel regulates critical cellular functions, including gene expression, cell growth and differentiation, and Ca2+ homeostasis. Loss-of-function mutations in the CRAC channel pore-forming protein ORAI1 or the Ca2+ sensing protein stromal interaction molecule 1 (STIM1) result in severe immune dysfunction and nonprogressive myopathy. Here, we identify gain-of-function mutations in the cytoplasmic domain of STIM1 (p.R304W) associated with thrombocytopenia, bleeding diathesis, miosis, and tubular myopathy in patients with Stormorken syndrome, and in ORAI1 (p.P245L), associated with a Stormorken-like syndrome of congenital miosis and tubular aggregate myopathy but without hematological abnormalities. Heterologous expression of STIM1 p.R304W results in constitutive activation of the CRAC channel in vitro, and spontaneous bleeding accompanied by reduced numbers of thrombocytes in zebrafish embryos, recapitulating key aspects of Stormorken syndrome. p.P245L in ORAI1 does not make a constitutively active CRAC channel, but suppresses the slow Ca2+-dependent inactivation of the CRAC channel, thus also functioning as a gain-of-function mutation. These data expand our understanding of the phenotypic spectrum of dysregulated CRAC channel signaling, advance our knowledge of the molecular function of the CRAC channel, and suggest new therapies aiming at attenuating store-operated Ca2+ entry in the treatment of patients with Stormorken syndrome and related pathologic conditions.

671M

High-resolution personal genome-wide maps of meiotic double-strand breaks in humans. *F. Pratto[1], K. Brick[1], P. Khil[1], F. Smagulova[2], G. Petukhova[2], R.D. Camerini-Otero[1].* 1) National Institutes of Health, Bethesda, MD; 2) Uniformed Services University of Health Sciences, Bethesda, MD.

Meiotic recombination is initiated by the formation of DNA double-strand breaks (DSBs) that are directed to specific genomic loci called hotspots by a meiosis-specific protein called PRDM9. To date, the most comprehensive map of hotspots sites in humans was generated using computational analysis of patterns of linkage disequilibrium (LD), however, this method cannot resolve gender-specific hotspots or hotspots defined by different alleles of PRDM9. In this work, we exploit a sequencing based method we recently developed to perform the first direct high-resolution genome-wide analysis of meiotic recombination initiation hotspots in individual human males. We mapped up to 39,000 DSB hotspots in each of five individuals with different combinations of PRDM9 alleles. Our PRDM9-specific DSB maps demonstrate that the LD-derived recombination maps reflect hotspots defined by different alleles of PRDM9 and we find that the relatively infrequent C-allele of PRDM9 (an allele found predominantly in African populations) contributes significantly to the LD map. About 75% of DSB hotspots are found at an LD-defined hotspot, yet DSB sites without an LD-hotspot still show a significant elevation in the recombination rate. Thus, our PRDM9-specific DSB hotspot maps greatly expand our knowledge of sites where recombination occurs in the human genome. We observed significant inter-individual variation in DSB frequency at hotspots, however most differences could not be explained by sequence variation at PRDM9 binding sites. Additional factors therefore dictate the efficiency of DSB formation. While sequence changes can influence recombination, we also found that recombination can influence the genomic sequence. Characteristic patterns of polymorphisms at hotspots offer compelling evidence for both GC-biased gene conversion and for a mutagenic effect of meiotic recombination. DSB hotspots were also enriched at structural variants that arise via homology-mediated mechanisms and at chromosomal breakpoints associated with many well-documented genomic diseases. Finally, an analysis of the relationship between meiotic DSBs and crossovers suggests that meiotic DSB initiation frequency is a primary determinant of the genetic crossover landscape.

672T

Investigating the maternal age effect on meiotic recombination in multiple cohorts. *H.C. Martin[1], J. Hussin[1], J. O'Connell[1,2], S. Gordon[3], K. McAloney[3], H. Mbarek[4], J.J. Hottenga[4], J. Marchini[5], D. Boomsma[4], N.G. Martin[3], P. Donnelly[2,5].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Illumina, Inc; 3) Queensland Institute for Medical Research, Brisbane, Australia; 4) Department for Biological Psychiatry, Vrije Universiteit, Amsterdam; 5) Department of Statistics, University of Oxford, Oxford, United Kingdom.

Multiple studies have reported that recombination rate increases with maternal age, but several have found an effect in the opposite direction. We investigate this question in new cohorts from the Australian and Dutch twin registries. Our rationale was that the use of dizygotic twins would allow measurement of the variability in recombination rate at a single point in a mother's lifetime. We analyse these twin cohorts in combination with several previously published studies, and compare different methods for calling crossovers. We then fit a Bayesian hierarchical model to the recombination rates, accounting for differences in the mothers' baseline rates, and test the hypothesis that the maternal age effect is the same in all cohorts.

673S

Mapping of two neurogenetic disorder genomes with a single molecule nanochannel array platform for genome-wide structural variation analysis. *Y.Y.Y. Lai[1], E.T. Lam[2], A.C.Y. Mak[1], V. Searles[3], C. Chu[1], C. Lin[1], N. Anderson[3], J.M. Sikela[3], P.Y. Kwok[1].* 1) UCSF, San Francisco, CA; 2) BioNano Genomics, San Diego, CA; 3) Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO.

Copy number variation of sequence encoding DUF1220 protein domains has been shown to be associated with human brain size and a number of neuropsychiatric disorders. DUF1220 motifs are embedded in the NBPF and PDE4DIP genes that are mostly found in Chr1q21. They are poorly assembled in the human reference genome due to the repetitive property of that region. Different methods, such as quantitative PCR, array comparative genomic hybridization (aCGH), Sanger or next-generation sequencing have been used for identification of these polymorphic loci. However these methods are too labor-intensive and do not provide positional (break-point) information for the structural variants. Genome mapping utilizes highly parallel nanochannel arrays in which hundreds of very long, fluorescently labeled, single DNA molecules are linearized and imaged. This novel approach is automated on the Irys System (BioNano Genomics, La Jolla, CA), which can scan the entire genome rapidly to generate physical maps that provide a more comprehensive view of the genome. Here, we use this genome mapping approach to detect genome-wide structural variation, including the copy number variation of DUF1220 in the PDE4DIP and NBPF genes, in two diploid cell lines from patients with neuropsychiatric disorders (autism and schizophrenia). To date, we generated over 50X coverage data and constructed de novo assembled genome maps that cover about 90% of hg38 reference genome using an automated assembly pipeline. We identified many structural variants including those are found in the Chr1q21 region. Overall, this genome mapping approach is simple and provides a list of genome-wide candidate structural variations that are associated with autism and schizophrenia.

674M

Association study of COL11A2 with aspirin exacerbated respiratory disease and its FEV1 decline. *J. Kim[1], C. Park[2], H. Shin[1].* 1) Research Institute for Basic Science, Sogang University, Seoul, South Korea; 2) Division of Allergy and Respiratory Medicine, Soonchunhyang University Bucheon Hospital, Bucheon, Republic of Korea.

Aspirin exacerbated respiratory disease (AERD) induces bronchoconstriction in asthma patients and is characterized with a clinical condition of severe decline in forced expiratory volume in one second (FEV1) after ingestion of aspirin. Genetic association studies of several candidate genes within human major histocompatibility complex (MHC) region on chromosome 6p21-24 have reported that this MHC genomic region is implicated in asthma and related respiratory diseases. This study investigated the genetic association of collagen, type XI, alpha 2 (COL11A2) within the MHC genomic region with AERD and its FEV1 decline by genotyping of 19 tagging COL11A2 SNPs in 93 AERD patients and 96 aspirin-tolerant asthma controls. As a result, polymorphisms of COL11A2 showed potential associations with AERD (minimum P = 0.02 in rs2269346), along with the increased significances for the FEV1 decline by aspirin provocation (minimum P = 0.002 in rs2855459). In the haplotype analysis, no significant signals were observed. Despite the need for further replications in larger cohorts and functional evaluations, our preliminary findings suggest that COL11A2 may be a predisposing factor for FEV1 decline-related symptoms.

675T

Bovine animal model for spermatic and scrotal alterations: additional clues for an X-chromosome component. *P.A.S. Fonseca[1], M.P. Almeida[1], G.S. Moura[2], F.C. Santos[1], D.J. Santos[3], G.C. Oliveira[4], V.R. Vale Filho[2], M.V.G.B. Silva[3], M.R.S. Carvalho[1].* 1) Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, 31.270-901, Brazil; 2) Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, 31.270-901, Brazil; 3) Embrapa Gado de Leite, Juiz de Fora, 36038-330, Brazil; 4) Centro de Pesquisas René Rachou (CPqRR)—FIOCRUZ, Belo Horizonte- MG, 30190002, Brazil.

Bovines are a subtilized, but interesting animal model for reproductive defects, mainly in males, because reproductive performance is attentively monitored over the selection process. In addition, male reproductive traits present high heritability in bovines. Genome-wide association studies (GWAS) in bovines can provide important clues for the genetic etiology of reproductive diseases not only in cattle, but also in others species. Here we describe an in-family GWAS for sperm defects and testis hypoplasia in a sample composed by 30 affected and 30 normal bulls of the Dairy-Gir breed (Bos taurus indicus). The phenotypes ascertained were gonadal hypoplasia and spermatic defects analysis. Animals were genotyped using Axiom Genome-Wide BOS 1 Array from Affymetrix, composed by 648K SNPs. Phenotypes were corrected for fixed effects (father, age, farm, and season). Pairwise Association Study (PAS) was developed for quantitative and numeric continuous phenotypes. For Haplotype Association Study (HAS) two dichotomic variable were construct: testes morphology abnormalities plus spermatic defects (ALL) and spermatic defects (Sperm Only). GWAS was conducted with SNP and Variation Suit 7 (Golden Heliz, Inc.). X-chromosomal associations were detected in both PAS and HAS. In PAS, near significant association was detected for four markers after Bonferroni correction. These markers map close to the following genes: serine/threonine protein kinase MST4 (*MST4*), oculocerebrorenal syndrome of Lowe (*OCRL*), *mir*2488 and Nance-Horan syndrome (*NHS*). In HAS, two significant associations were detected: Cullin 4B (*CUL4B*) gene for ALL phenotype and zinc finger and BTB domain containing 33 (*ZBTB33*) for Sperm Only phenotype. *CUL4B* plays a crucial role in post-translational modification, involving ubiquitin, of proteins involved in cell cycle regulation, DNA replication and embryonic development. *CUL4A*, a *CULB4B* homologue, has been associated to spermatic development. *ZBTB33* encodes Kaiso zinc finger protein associated with embryonic development, homeostasis of adult tissues, maintenance of progenitor cells, cell fate determination and differentiation and with reproductive traits in bovine males. These findings, although still in the need of confirmatory functional studies, point to new heterologous candidate genes for human reproductive disturbances.

676S

De novo genome assembly and structural variations detection by genome mapping on nanochannel arrays. *A.C.Y. Mak[1], E.T. Lam[8], Y.Y.Y. Lai[1], A.K.Y. Leung[3], T.P. Kwok[4], C.M.L. Li[6], A.R. Hastie[8], W. Stedman[8], T. Anantharaman[8], W. Andrews[8], X. Zhou[8], H. Dai[8], K.Y. Yip[4,5], T.F. Chan[5], S.M. Yiu[6], J.W. Li[3,5], A.K.Y. Yim[3,5], S. Chan[8], Z. Dzakula[8], H. Cao[8], J. Sibert[7], M. Xiao[7], P.Y. Kwok[1,2].* 1) Cardiovascular Research Institute, Univeristy of California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 3) School of Life Sciences, The Chinese University of Hong Kong, Hong Kong; 4) Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong; 5) Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong, Hong Kong; 6) Department of Computer Science, The University of Hong Kong, Hong Kong; 7) School of Biomedical Engineering, Science & Health Systems, Drexel University, Philadelphia, PA; 8) BioNano Genomics, San Diego, CA.

Despite recent advances in next-generation sequencing technologies, de novo genome sequence assembly and genome-wide structural variation (SV) detection based on 'short reads' remain challenging. For efficient de novo sequence assembly and complete SV characterization, we generated genome maps with long DNA molecules (>150kb) fluorescently labeled at Nt.BspQI sites (GCTCTTCN/), linearized and imaged in massively parallel nanochannel arrays [1]. We obtained single-molecule data at 80X genome coverage on CEPH CEU trio samples, NA12878, NA12891 and NA12892, that were extensively analyzed by many groups, including the 1000 Genomes Project. Single-molecule maps were assembled de novo into consensus genome maps that cover >90% of the reference human genome. We confirmed previously reported and identified novel structural variants that were consistent with Mendelian inheritance. We also analyzed genome features of interest (e.g. subtelomeric and subcentromeric regions). Our genome maps provide valuable information on genome features and structural variants that are otherwise hard to discover with short read sequencing data alone. 1. Lam ET et al. Genome Mapping on Nanochannel Arrays for Structural Variation Analysis and Sequence Assembly. Nat Biotechnol (2012) 30(8):771-776.

677M

An integrated platform for the collection of biospecimens to support the Genotype-Tissue Expression (GTEx) Project. *JC. Keen[1], L. Carithers[1], A. Undale[2], A. Rao[1], M. Barcus[2], P. Branton[1], L. Sobin[2], P. Guan[1], N. Roche[2], C. Shive[2], A. Smith[2], K. Valentino[2], S. Volpi[3], J. Struewing[3], S. Koester[4], R. Little[4], HM. Moore[1] on behalf of the GTEx consortium.* 1) Biorepositories and Biospecimen Research Branch, National Cancer Institute, Bethesda, MD; 2) Leidos Biomedical Research, Inc., Rockville, MD; 3) National Human Genome Research Institute, Bethesda, MD; 4) National Institute of Mental Health, Bethesda, MD.

The Genotype-Tissue Expression (GTEx) project is an NIH Common Fund study that aims to clarify our understanding of how genetic variation influences gene expression in normal human tissues. As of May 2014, GTEx has collected 25+ tissue types per donor from over 650 deceased tissue or organ donors. To facilitate the collection and storage of tissues and data, the NCI Biorepositories and Biospecimen Research Branch (BBRB) has developed an integrated biospecimen collection platform. This platform includes collection and storage facilities, a data warehouse, a team of board-certified pathologists, and a molecular analysis core facility. Together, this system has allowed for the acquisition of high quality, well-annotated biospecimens that met established quality criteria for molecular analysis. Clinical data, genotype and gene expression data are available at dbGaP for each donor and tissue type collected. These data will enable scientists to better understand the genomic and molecular variation in normal human tissues and may facilitate the identification of genes involved in disease states.

678T

Functional characterization of recent single nucleotide mutations on HSPA1A, a human Hsp70 gene. *K. Hess, M. Siracusa, P. Nguyen, A. Bilog, C. McCallister, N. Nikolaidis.* Department of Biological Science, Center for Applied Biotechnology Studies, and Center for Computational and Applied Mathematics, California State University Fullerton, Fullerton, CA.

A fundamental question in molecular evolution is how mutations enable organisms to adapt to their environments and survive. At the cellular level the same essential question relates to the ability of cells to adapt and survive homeostatic imbalances due to stress. To maintain homeostasis the cellular stress response system, which is composed by several protein networks, has evolved. Molecular chaperones and in particular the 70-kD heat shock proteins (Hsp70s) are key regulators of the stress response system and alterations in their function have direct physiological consequences. Therefore, it is of paramount importance to determine how natural mutations alter the function of Hsp70s and how these changes affect cellular and organismal adaptation and survival. To shed light to this important biological question we capitalized on the presence of multiple natural single nucleotide polymorphisms (SNPs) found on HSPA1A, the major stress inducible Hsp70 gene in humans, and determined whether these mutations alter protein function. Specifically, the wild-type HSPA1A sequence was subcloned into both bacterial and mammalian expression vectors and the mutated gene variants were generated using site directed mutagenesis. Recombinant proteins corresponding to the WT and mutated variants were then generated and tested for their ability to hydrolyze ATP. Based on the Km values obtained, the WT had the highest affinity for binding to ATP, but based on the Vmax values, some mutants were able to hydrolyze ATP at a much faster rate. The different affinities for both ATP and ADP were verified using Isothermal titration calorimetry (ITC), which also revealed differences in the reaction entropy. The mutant proteins' ability to properly localize intracellularly within mammalian cells was also tested by tagging HSPA1A to GFP, using fluorescent dyes to stain the mitochondria, lysosomes, and plasma membrane, and viewing where the proteins localized via confocal microscopy. These results revealed that the mutants and the WT protein had similar subcellular localization. However, live-dead and cell proliferation assays revealed that some of these mutations significantly change cellular growth and survival after heat stress. Given that these natural variants are either population-specific or clinical we suspect that the observed functional differences alter the ability of cells and the individuals carrying them to cope with stress or disease.

679S

Impact of genetic polymorphisms in FADS1, FADS2, and FADS3 genes on fatty acid metabolic mechanism on methadone therapy in Taiwan. *R. Wang[1], H.T. Yang[2], H.Y. Lane[3], C.L. Huang[4], Chieh-Liang Huang.* 1) Department of Public Health, china Medical Univ, Taichung, Taiwan; 2) Department of Nutrition , China Medical University, Taiwan; 3) Institute of Clinical Medical Science, College of Medicine, China Medical University, Taichung, Taiwan; 4) Department of Psychiatry, China Medical University Hospital, Taichung, Taiwan.

It had been observed that abnormal fatty acid metabolism in patients with psychological disorders, such as major depression, schizophrenia, and Alzheimer's disease , etc. Illegal drugs is not only associated with the increased risk of the mental health problems, family and social economic burden, but also the elevated risk of HCV and AIDS infection. Drug abuse is an important public health issue. Heroin is one of the most common increasing abuses in Taiwan. Methadone, a synthetic μ-opioid receptor agonist, is extremely effective to treat to reduce illegal drug use, AIDS-caused infectious transmission and mortality in heroin dependent abusers patients. The aim of the study is to explore the association between the 7 tagSNPs (rs174547, rs174550, rs174570, rs174602, rs498793, rs526126, rs174634) of FADS1, FADS2, and FADS3 (fatty acid desaturase 1, 2, and 3) and fatty acid variation before and after methadone maintenance treatment in drug abusers. A total of 89 subjects with heroin dependence or methamphetamine were recruited from the methadone clinic in China Medical University Hospital. Based on the Chinese Han hapmap LD structure, seven tagSNPs were selected from FADS1, FADS2 and FADS3 genes. Genotyping were carried out using the Applied Biosystem Assay on Demand reagents and were implemented using an ABI Prism 7900HT Sequence Detection System. Fatty acid profiles of RBC were analyzed with gas chromatography. This study was approved by the institutional review board of China Medical University Hospital in Taiwan. All the informed consent was obtained from all participants during their initial clinic visit. All data analyses were performed using SAS version 9.1.3 (SAS Inc., NC, USA). The study results showed that all the 7 tagSNPs are significantly associated with the variation of PUFA and AA/EPA ratio ($p<0.0001$) but not associated with C18:2n6, C18:3n6 before and after methadone treatment. The data pointed out that methadone treatment would interfere fatty acid metabolism, especially PUFA desaturation in subjects. As abnormal fatty acid profiles were readily associated with several psychological disorders, the correlation between fatty acid metabolism and drug abusers with certain tag SNPs. In conclusion, our study may provide further information regarding the 7tagSNPs of the FADS1, FADS2 and FADS3 were associated with fatty acid metabolism for drug abuser with MMT treatment in Taiwan.

680M

Drug Metabolizing Enzyme and Transporter Gene Variation, Nicotine Metabolism, and Cigarette Consumption. *A.W. Bergen[1], M. Michel[1], D. Nishita[1], R. Krasnow[1], H.S. Javitz[1], K.N. Conneely[2], C.N. Lessov-Schlagger[3], H. Hops[4], A.Z.X. Zhu[5], J.W. Baurley[6], J.B. McClure[7], S.M. Hall[8], T.B. Baker[9], D.V. Conti[10], N.L. Benowitz[11], C. Lerman[12], R.F. Tyndale[13], G.E. Swan[14].* 1) Center for Health Sciences, SRI International, Menlo Park, CA; 2) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 3) Department of Psychiatry, Washington University School of Medicine, St Louis, MO; 4) Oregon Research Institute, Eugene, OR; 5) Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON; 6) BioRealm, LLC, Monument, CO; 7) Group Health Research Institute, Seattle, WA; 8) Department of Psychiatry, University of California, San Francisco, CA; 9) Center for Tobacco Research and Intervention, Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI; 10) Department of Preventive Medicine, University of Southern California, Los Angeles, CA; 11) Departments of Medicine and of Bioengineering & Therapeutic Sciences, University of California San Francisco; 12) Department of Psychiatry, University of Pennsylvania, Philadelphia, PA; 13) Centre for Addiction and Mental Health, and Departments of Psychiatry, and of Pharmacology and Toxicology, University of Toronto, Toronto, ON; 14) Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine.

The Nicotine Metabolite Ratio (NMR), the ratio of trans-3'-hydroxycotinine and cotinine, two nicotine metabolites, is significantly associated with CYP2A6 activity, response to smoking cessation treatments, and cigarette consumption. We searched for drug metabolizing enzyme and transporter (DMET) gene variation associated with the NMR and with prospective abstinence in 2,946 participants of clinical laboratory studies of nicotine metabolism ("laboratory studies") and of randomized comparative efficacy trials of smoking cessation therapies ("clinical trials") using a two-stage design. The first stage was a meta-analysis of the association of 507 common single nucleotide polymorphisms (SNPs) at 173 DMET genes, using the Affymetrix DMET™ Plus Array, with the NMR in 449 participants of two laboratory studies. Nominally significant associations with the laboratory-based NMR were identified in ten DMET genes after adjustment for intragenic tested SNPs, and two CYP2A6 SNPs attained experiment-wide significance after adjusting for 507 SNPs (rs4803881, PACT=4.5E-5, PBonferroni=6.7E-5; rs1137115, PACT=1.2E-3, PBonferroni=1.6E-3). In the second stage, we genotyped selected DMET SNPs and performed mega-regression analyses with the pretreatment NMR and with prospective abstinence in clinical trial participants. rs1137115 was significantly associated (P=4.2E-14, N=614) with pretreatment NMR. DMET SNPs analyzed in the validation stage were not associated with prospective abstinence outcomes in up to 2,497 participants of eight clinical trials. rs1137115 accounts for 8% of pretreatment NMR and 0.33% of cigarette consumption (P=2E-3), among clinical trial participants. We have identified and validated significant associations of one CYP2A6 SNP with the NMR at experiment-wide and at genome-wide significance in laboratory study and clinical trial participants.

681S

Genetic variation in glutamate signaling influences response to cognitive behavioral therapy in pediatric anxiety. *C.G. Bhat, C.P. Laughlin, G.S. Hellemann, J.C. Piacentini, J.T. McCracken, E.L. Nurmi, The CAMS Study Group.* University of California Los Angeles, Los Angeles, CA.

Background: Selective Serotonin Reuptake Inhibitors (SSRIs) and cognitive-behavioral therapy (CBT) are effective therapies for pediatric anxiety; however, response to treatment is highly variable. Differential treatment response may be explained by genetic factors. Due to the role of glutamate in learning and memory, we examined genetic variation in the glutamate pathway on the effectiveness of sertraline vs. CBT vs. combination therapy for pediatric anxiety in the randomized, placebo-controlled Child/Adolescent Anxiety Multimodal Study (CAMS). Methods: We examined 15 key variants in glutamate signaling candidates for association with treatment response: NMDA glutamate receptor GRIN2B and transporters regulating synaptic glutamate levels (the cysteine-glutamate exchanger (SLC7A11) and neuronal (SLC1A1) and glial (SLC1A2) glutamate transporters). Results: Influences of the glutamate system were apparent only in the CBT treatment group. One SLC1A1 variant was significant in omnibus analyses across all groups. Minor allele carriers (C+) at the glutamate transporter (SLC1A1) rs3933331 variant demonstrated minimal symptom reduction over time with CBT monotherapy ($p= 9.06 \times 10^{-5}$), whereas GG homozygotes showed an excellent response. Interestingly, the opposite allele showed a greater symptom drop with placebo. Within the CBT group, SLC7A11 rs11723658 ($p<0.0005$) and GRIN2B rs1806195 ($p<0.005$) also predicted treatment response. Conclusion: While no effects on medication response were observed, our data support a possible role for genetic variation in glutamate signaling pathways in response to CBT monotherapy. Our results show promise for eventual personalization of anxiety treatment algorithms and warrant replication in independent samples.

682M

Clinical and pharmacogenomic features of cisplatin-induced ototoxicity in Asian nasopharyngeal carcinoma patients. *S.L. Chan[1], L.S. Ng[2], C.H. Siow[2], C.J. Ross[3,4], B.C. Carleton[4,5], M.R. Hayden[1,3], B.C. Goh[6,7,8,9], K.S. Loh[2], L.R. Brunham[1,3,10].* 1) Translational Laboratory in Genetic Medicine, Agency for Science, Technology and Research, Singapore; 2) Department of Otolaryngology-Head & Neck Surgery, National University Health System, Singapore; 3) Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada; 4) Division of Translational Therapeutics, Department of Pediatrics, Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada; 5) Pharmaceutical Outcomes Programme, BC Children's Hospital, Vancouver, BC, Canada; 6) Department of Haematology-Oncology, National University Cancer Institute, National University Health System, Singapore; 7) Cancer Science Institute, Singapore; 8) Department of Pharmacology, Yong Loo Lin School of Medicine, National University Health System, Singapore; 9) Haematology Oncology Research Group, National University Cancer Institute, National University Health System, Singapore; 10) Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

Background: Cisplatin-induced ototoxicity (CIO) is a common and debilitating adverse drug reaction (ADR) that affects up to a quarter of adults and more than half of children receiving the drug, and Asian populations may be at increased risk. Inter-individual variability in susceptibility to CIO suggests a genetic component to this ADR that has not been fully defined. In European populations, *TPMT* rs12201199 and *COMT* rs9332377 have been found to be associated with CIO. However, these variants are very rare or absent in healthy individuals from the Chinese and Malay populations. We therefore hypothesized that other pharmacogenomic variants influence the risk of CIO in these Asian populations. **Methods:** Nasopharyngeal carcinoma patients were recruited from the National University Hospital of Singapore and genotyped on a custom Illumina Infinium panel containing 7907 SNPs in genes involved in drug absorption, distribution, metabolism and excretion (ADME). Clinical and audiometric data were collected from all patients. Trends of hearing loss were first delineated in terms of threshold change from baseline averaged across both ears over time. Genetic association was then performed only on patients who received cisplatin. CIO cases were defined as >15dB mean threshold shift at 1 year and controls defined as ≤15dB shift. **Results:** Out of 115 patients with complete clinical and audiometric data, 24 received radiotherapy (RT) only and 91 received both RT and cisplatin (chemoRT). There was no significant threshold shift at lower hearing frequencies (0.5–2kHz). However at 4kHz, there was a substantial threshold increase from as early as 3 months after initiation of therapy in patients who received chemoRT. At 12 months, 32% of patients who received chemoRT compared to 0% who received RT displayed a threshold increase of >15db. After QC, 82 patients (24 cases and 58 controls) and 4281 SNPs were included in the genetic association study. *COMT* and *TPMT* variants reported previously were extremely rare in this population. None of the ADME variants reached statistical significance after correction for multiple testing. **Conclusion:** High frequency hearing loss in NPC patients at time points up to one year is due primarily to cisplatin. Established risk variants for CIO are rare in this population, suggesting the presence of novel genetic factors. Given the small sample size of this preliminary study we did not detect evidence of ADME variants with large effect size on CIO.

683S

Genetic moderators of treatment response to dexmethylphenidate in children and adolescents with ADHD. *S.N. Chang, E.L. Nurmi, C.P. Laughlin, G.S. Hellemann, S.K. Loo, J.J. McGough, J.T. McCracken.* Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA.

Background: The stimulant dexmethylphenidate (d-MPH) is an effective treatment for hyperactive and inattentive symptoms associated with pediatric attention deficit hyperactivity disorder (ADHD). Treatment-induced growth slowing, however, is a common adverse effect associated with morbidity and poor treatment adherence. Substantial variation in response and growth effects is common and may be explained by genetic factors. This study investigated whether genetic variation in drug targets (SLC6A3/DAT1, SLC18A2/VMAT2, TAAR1) and metabolic enzymes and transporters impacting drug disposition (ABCB1, NR1I2) could help explain differential outcomes in treatment response to d-MPH. Methods: In 202 children participating in the NIMH TRECC Study, height and weight were measured at baseline and at Weeks 4 and 8. A subset continued treatment for 18 months and growth was measured regularly. Treatment response was determined by scores on the ADHD-rating scale (ADHD-RS) and Clinical Global Impression Improvement (CGI-I) Scale. Variants previously associated with similar phenotypes or shown to impact protein function were genotyped using TaqMan allelic discrimination assays. Results: Homozygotes for the minor allele (GG) of exonic NR1I2 variant rs7643645 demonstrated rare guanfacine-associated growth suppression ($p=1.78 \times 10^{-13}$). A second NR1I2 variant (rs1523130) showed differential effects across treatment groups ($p=4.09 \times 10^{-10}$). Heterozygotes for the intronic ABCB1 variant rs1128503 ($p=1.17 \times 10^{-10}$) and 5'-UTR TAAR1 variant rs5008782 ($p=3.24 \times 10^{-6}$) revealed amplified effects of growth suppression when treated with d-MPH and acceleration when treated with guanfacine. Homozygotes for the minor allele (TT) of 3'-UTR SLC6A3 variant rs27072 ($p=4.27 \times 10^{-7}$) and intronic SLC18A2 variant rs363227 showed a paradoxical growth acceleration when treated with d-MPH (rs363227). Minor allele carriers (A+) of the SLC18A2 variant rs363227 show a relatively poor response to either guanfacine or stimulant monotherapy. In contrast, SLC18A2 A+ carriers achieved a significantly greater symptom reduction to combination treatment (p=0.01). Conclusion: Genetic variants impacting d-MPH pharmacokinetic and pharmacodynamic effects influenced treatment outcomes and adverse effects to common ADHD treatments. Understanding the pharmacogenomic factors moderating treatment response and adverse effects could lead to strategies to individualize treatment matching and develop improved interventions.

684M

Genetic vulnerability to adverse child attachment quality and stress reactivity. *E.H. Gail[1], J.L. Borelli[2], A. Nguyen[1], C.M. Laubacher[2], G.S. Hellemann[1], E.L. Nurmi[1].* 1) University of California, Los Angeles, Los Angeles, CA; 2) Pomona College, Claremont, CA.

Background: Stressful parent-child interactions are known to have effects on both child attachment quality and physiological autonomic reactivity. In this pilot study, we examined the relationship between environmental stress and genetic vulnerability at four polymorphisms that have been linked to stress and attachment phenotypes: serotonin transporter promoter polymorphism (HTT-LPR), oxytocin receptor (OXTR) rs53576, mu-opioid receptor (OPRM1) rs1799971, and glucocorticoid receptor chaperone (FKBP5) rs3800373. Methods: The sample consisted of 45 mother-child dyads recruited from the community. The Child Attachment Interview (CAI) and the Parental Stress Index were collected, as well as cortisol saliva levels during a stressful task. Individual genotypes were assayed and combined genetic load was calculated by adding the number of risk alleles carried by each child. Results: The minor allele at FKBP5 rs3800373 was associated with cortisol reactivity (p=0.038) and attachment quality (p=0.015), both with medium effect sizes (f2 = 0.1). Combined genetic load positively correlated with higher scores on CAI (r=0.32, p=0.036). The correlation of a higher number of risk alleles with greater attachment security may be due to a reversal of effects in the absence of stress, which has been previously observed. While this preliminary study was underpowered to sufficiently analyze gene-environmental interactions, the level of parent-child stress moderated the interaction between risk alleles and both child attachment and cortisol reactivity at the trend level (p<0.1). Conclusions: These data suggest that genetic risk alleles may have opposite effects on child attachment quality and stress reactivity in the context of high versus low parent-child stress. Our findings are consistent with and extend prior observations. A larger study to further explore these potential gene-environment interactions is currently underway. An understanding of the interactions of parenting behavior and genetic load may help predict the development of child anxiety and inform the subsequent development of targeted interventions.

685S

Genetic variation in mitochondrial pore protein ANT1 predicts antipsychotic-induced weight gain in the NIMH RUPP Autism sample. *K.J. Hwang, K.S. Mallya, G.S. Hellemann, J.T. McCracken, E.L. Nurmi.* Semel Institute for Neuroscience at University of California, Los Angeles, Los Angeles, CA.

Background: The atypical antipsychotic drug, risperidone, has been shown to decrease irritability and aggression in children with autism spectrum disorder (ASD); however, weight gain is a common adverse event. Genetic background may explain the substantial variability in this treatment-limiting side effect. As previous data suggests a potential mitochondrial contribution to antipsychotic-induced weight gain (AIWG), we attempted to replicate prior findings at three specific mitochondrial gene variants (SLC25A4, TSPO and NDUFS1). Methods: 225 subjects participating in the NIMH Research Units on Pediatric Psychopharmacology (RUPP) Autism Risperidone trials were genotyped and individual BMIs were measured weekly for eight weeks to examine association of AIWG with mitochondrial gene variants. Results: The rs10024068 promoter variant in SLC25A4 strongly moderated AIWG over the course of just 8 weeks of treatment in both the complete sample ($p = 3.21\times10^{-8}$) and the Caucasian subset ($p = 2.49\times10^{-6}$). Individuals homozygous for the minor G-allele gained significantly more weight (Z-Score change GG = 1) than A-allele carriers (Z-Score change AG=0.58 and AA = 0.61). This single variant explains 6% of the variation in AIWG in our sample. SLC25A4 encodes the Adenine Nucleotide Translocator 1 (ANT1), part of the mitochondrial permeability pore. There was no evidence for association of weight gain with genetic variants in NDUFS1 or TSPO. As seen in prior studies, however, a significant interaction between SLC25A4 and TSPO was observed ($p = 0.018$), which is especially relevant given the biological association of ANT1 and translocator protein TSPO in forming the mitochondrial permeability pore. Conclusion: Genomewide significant association at SLC25A4 supports the involvement of mitochondrial energy pathways in AIWG and suggests the examination of additional mitochondrial candidates as well as replication in independent samples. These data may help uncover pathways underlying AIWG and may eventually be used to optimize treatment matching.

686M

The application of personalized medicine for understanding clinical variability in rasagiline response in early Parkinson's Disease. *J. Knight[1,2,3], S. Collinson[1], N. Freeman[1], M. Tampakeras[1], J. Levy[4], A. Tchelet[4], E. Eyal[4], E. Berkovich[4], R. Eliaz[4], V. Abler[4], I. Grossman[4], A. Tiwari[1], J.L. Kennedy[1,2,3], M. Masellis[3,5,6,7], A.E. Lang[5,8,9], ADAGIO investigators.* 1) Campbell Family Mental Health Research Institute, Centre of Addiction and Mental Health, Toronto, Ontario, Canada; 2) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 3) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 4) Teva Pharmaceutical Industries, Israel; 5) Department of Medicine, Division of Neurology, University of Toronto, Toronto, Canada; 6) Movement Disorders Clinic, Sunnybrook Health Sciences Centre, Toronto, Canada; 7) Neurogenetics Section, Centre for Addiction and Mental Health, Toronto, Canada; 8) Morton and Gloria Shulman Movement Disorders Clinic, Toronto Western Hospital, University Health Network, Toronto, Canada; 9) The Edmond J. Safra Program in Parkinson's Disease, Toronto Western Hospital, University Health Network, Toronto, Canada.

Objective: To identify candidate gene polymorphisms associated with peak motor benefit to rasagiline at 12 weeks and associated with sustained benefit over a 36-week evaluation period. **Background:** Rasagiline, a selective irreversible monoamine oxidase-B (MAO-B) inhibitor, has been shown to be safe and effective for the treatment of Parkinson's disease (PD). As with all drugs, there is variability among individuals in their clinical benefit to rasagiline, this study is designed to identify genetic factors that may influence this variability. **Methods:** We performed a retrospective genetic association study using clinical data from the ADAGIO trial. Candidate genes encoding proteins involved in catecholamine synthesis and metabolism, rasagiline metabolism, as well as those reported in Genome Wide Association Studies (GWAS) of PD susceptibility were selected. We examined association between genetic polymorphisms and peak change in Unified Parkinson Disease Rating Scale (UPDRS) score from baseline to 12 weeks, using a linear model. We also examined association between genetic polymorphisms and change in UPDRS score from baseline over 12, 24, and 36 weeks, using a mixed effects linear model. Both models controlled for placebo response. **Results:** 204 Single Nucleotide Polymorphisms (SNPs) and 5 Variable Number Tandem Repeat (VNTR) from 28 candidate genes were successfully genotyped in 694 samples Analysis 1: Two SNPs in strong linkage disequilibrium within the dopamine D2 receptor gene (DRD2) were found to be significantly associated with peak change in UPDRS scores at 12 weeks (rs1076560 and rs2283265, False Discovery Rate [FDR] p=0.045 for each) as well as one in Norepinephrine Transporter (SLC6A2) (rs36023, FDR-corrected p=0.045). Analysis 2: No allelic associations were identified in the model assessing longitudinal data. **Conclusions:** To our knowledge, this is the largest pharmacogenetic study of an anti-Parkinsonian drug conducted to date. The D2 receptor is a major target of dopamine and its stimulation relates to motor benefits in PD. Prior literature indicates that the two associated SNPs alter transcriptional processing of DRD2. Together with the dopamine transporter, the norepinephrine transporter is involved in the pre-synaptic reuptake of catecholamines, including dopamine. Further investigation of these genes is clearly warranted.

687S

Evaluation of candidate gene SNPs in the International Clopidogrel Pharmacogenomics Consortium (ICPC) reveals genetic variation that significantly impacts on-treatment platelet reactivity. *J. Lewis on behalf of the International Clopidogrel Pharmacogenomics Consortium.* Endocrinology, Diabetes & Nutrition, University of Maryland, Baltimore, Baltimore, MD.

Antiplatelet therapy with clopidogrel is standard of care in improving outcome in secondary prevention for cardiovascular patients. While generally effective, substantial inter-individual variation in response to clopidogrel exists, and high on-treatment platelet reactivity to adenosine diphosphate (ADP) is associated with recurrent cardiovascular events. To better assess the genetic determinants of clopidogrel response, we evaluated 28 previously reported single nucleotide polymorphisms (SNPs) on platelet reactivity in 2054 coronary artery disease patients of the ICPC. Consistent with previous investigations, we detected strong association between the loss-of-function *CYP2C19*2* [rs4244285] variant and increased on-clopidogrel ADP-induced platelet aggregation ($P=9.83 \times 10^{-25}$, Beta = 0.48). Conversely, on-treatment ADP-induced platelet aggregation was significantly decreased in subjects with the gain-of-function *CYP2C19**17 variant both before and after adjustment for *CYP2C19*2* ($P=2.33 \times 10^{-7}$, Beta = -0.23 and $P=3.54 \times 10^{-4}$, Beta = -0.12, respectively), as well as a functional SNP in *CES1* (G143E [s71647871]), which metabolizes clopidogrel to an inactive form ($P=7.76 \times 10^{-6}$, Beta = -0.62). A variant in *P2YR12* (rs1472122), which encodes the platelet ADP receptor was also significantly associated with on-clopidogrel platelet reactivity ($P=3.82 \times 10^{-4}$, Beta = 0.13). We observed no evidence of association between platelet aggregation and genetic variants in other candidate genes including *ABCB1* C3435T [rs1045642] and C1236T [rs1128503]), *PON1* (Q192R [rs662] and L55M [rs854560], *ITGB3* (rs5918), and other CYP enzymes (e.g. *CYP2B6*, *CYP2C9*, and *CYP1A2*). Our data suggest that, in addition to *CYP2C19*, genetic variation in genes responsible for generation of the inactive clopidogrel metabolite formation and regulating the P2Y12 ADP-receptor impact on-treatment platelet reactivity and may influence overall clopidogrel efficacy.

688M

Genetic variation in the oxytocin/vasopressin locus is associated with treatment-dependent improvement in social functioning in autism spectrum disorder. *W. Li, E.L. Nurmi, C.P. Laughlin, J.T. McCracken, NIMH RUPP Autism Network.* Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA.

Background: While social functioning is a key deficit in autism spectrum disorders (ASD), few current treatments specifically target these impairments. In the NIMH Research Units on Pediatric Psychopharmacology (RUPP) study, risperidone reduced social withdrawal ratings in children with ASD; however, variability was substantial. We examined genetic variability in the oxytocin and vasopressin signaling pathways, which are known to play a role in social behavior, for hypothesized association with treatment response. Methods: Risperidone response in 225 children and adolescents with ASD in the NIMH RUPP Risperidone trials was measured by weekly Aberrant Behavior Checklist (ABC) Social Withdrawal subscale (II) ratings over 8 weeks of treatment. Complete common variation at the Oxytocin (OXT), Arginine Vasopressin (AVP), and Arginine Vasopressin Receptor (AVPR1 and 2) loci was captured with 16 tSNPs. Additionally, we included 10 Oxytocin Receptor (OXTR) SNPs with evidence for prior association with various social phenotypes. A repeated measures general linear mixed model was used to examine genetic association with symptom improvement. Analyses were restricted to the Caucasian subset to limit effects of ethnicity. Results: Three independent variants (rs2740204, rs2770378 and rs4813625) tagging the genomic locus containing the adjacent AVP and OXT genes predicted improvement in ABC Social Withdrawal ($p<0.002$). Individuals homozygous for the C-allele at rs4813625 ($p= 0.0014$) and the A-allele at rs2770378 ($p= 0.0019$) showed greater improvement in social withdrawal than G-carriers. The T-allele at rs2740204 demonstrates an allele dosage-dependent effect on symptom improvement ($p=0.0017$). No effects at either OXTR or AVP receptors was observed. Conclusions: Our results suggest that variants in the OXT/AVP systems influence social withdrawal treatment response to risperidone in children with ASD. If replicated in independent samples, these results may help guide future clinical treatment algorithms.

689S

Genetic variation in dopaminergic and adrenergic receptor targets moderates weight loss during ADHD treatment. *N.A. Nguyen, C.P. Laughlin, G.S. Hellemann, S.K. Loo, J.J. McGough, J.T. McCracken, E.L. Nurmi.* Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA.

Background: Growth effects of ADHD pharmacotherapies are important, treatment-limiting side effects. Previously we reported genetic influences on dexmethylphenidate (d-MPH) and guanfacine treatment of children with ADHD in the UCLA TRECC sample, including variants associated with treatment response and altered height trajectories. The present study focuses BMI changes with stimulant treatment. Due to monoamine and acetylcholine involvement in growth hormone regulation and energy balance, we examined genetic variation in these signaling pathways for association with growth effects. **Methods:** 202 subjects between 7–14 years of age were recruited for the acute phase of a randomized, double-blind, placebo-controlled trial of d-MPH and guanfacine for pediatric ADHD. Three treatment groups included guanfacine monotherapy, d-MPH monotherapy, and combination guanfacine plus d-MPH. Medication responders continued in the trial for approximately 14 months (n=99) and height and weight were tracked regularly in addition to ADHD symptoms. We tested association of genetic variation in monoamine and energy balance candidate systems with height and BMI changes during medication exposure. Due to the large number of markers tested (n=95), we only report results meeting genomewide significance. **Results:** In our sample, guanfacine monotherapy was associated with height and BMI increases over predictions from CDC growth charts (Z-score increase of 0.14 for height and 0.41 for BMI). Both d-MPH monotherapy (Z-score decrease of -0.12 for height and -0.84 for BMI) and combination treatment (Z-score decrease of -0.19 for height and -0.83 for BMI) were associated with growth slowing. Variation in treatment group trajectories for height ($p=6.47\times10^{-15}$) and BMI ($p=5.58\times10^{-34}$) were highly significant. Minor allele (A) carriers at DRD3 rs3732790 showed allele dosage-dependent clinically-significant weight loss (2.6 Z-Score loss) on d-MPH ($p=1.53\times10^{-10}$). BMI change was moderated by ADRA2A rs1800544, with profound weight loss on d-MPH in C-allele carriers and combination treatment in G-allele carriers ($p=2.73\times10^{-9}$). Effects are similar and maintain Bonferroni significance in the Caucasian subset, ruling out population effects. **Conclusions:** The results presented here suggest that genetic factors contribute significantly to stimulant-mediated weight loss pediatric subjects with ADHD and appear to be able to identify patients more likely to be affected.

690M

Differential genetic effects of the melanocortin system in antipsychotic-induced weight gain in children versus adults. *A.K. Subhash, K.S. Mallya, C.P. Laughlin, G.S. Hellemann, J.T. McCracken, E.L. Nurmi.* Semel Institute for Neuroscience, University of California, Los Angeles, 760 Westwood Plaza, 47-376, Los Angeles, CA 90024.

Background: Antipsychotic-induced weight gain (AIWG) is a common, treatment-limiting adverse effect of this widely prescribed drug class. Prior examination of the melanocortin system in adults with AIWG has demonstrated a consistent contribution of the melanocortin 4 receptor (MC4R), which is central to energy balance, growth and body weight regulation. We therefore queried 3 genes in the melanocortin system in our sample of children and adolescents with autism spectrum disorder (ASD) treated with risperidone for aggression and irritability. Methods: 225 youth in the NIMH Autism Risperidone studies were genotyped using Life Technologies' Taqman platform. Complete common variation in key melanocortin signaling receptors was captured with 15 tSNPs spanning the MC3R, MC4R and neuropeptide Y (NPY) loci. Variants were tested for association with BMI and height change (measured weekly across 8 weeks of treatment) using a repeated measures general linear mixed model. Results: In contrast to previous literature, MC4R was not associated with AIWG in our sample. Common allele homozygotes (GG) at MC4R promoter SNPs rs8087522 ($p=7.65\times10^{-7}$) and rs8093815 ($p=6.70\times10^{-9}$) however, showed greater change in height over 8 weeks of treatment in both the complete dataset and the Caucasian subset. On the other hand, we found strong ethnicity-specific association of AIWG with two MC3R gene variants: 3' variant rs6024733 ($p=1.2228\times10^{-5}$) in African Americans and promoter SNP rs6127698 ($p=8.78\times10^{-7}$) in Caucasians. Common allele homozygotes (GG) at rs6127698 and minor allele homozygotes (AA) at rs6024733 were relatively protected from AIWG. While no NPY effects on AIWG were observed, one NPY SNP (rs16141, $p=5.01\times10^{-5}$) predicted height change in Caucasians. These results survive correction for the 120 tests that have been performed in this sample in related studies. Conclusions: Contrary to prior literature in adults showing a substantial contribution of MC4R to AIWG, our data suggest that genetic variation in the peptide hormone receptor MC3R influences AIWG in children. Furthermore, variations in MC4R and NPY show strong height effects in children. Distinct but related genetic effects on AIWG between adults and children may help explain the increased risk of children to this serious adverse event. Thoroughly understanding the pharmacogenomic factors involved in AIWG may guide clinical treatment matching and minimize adverse effects for patients taking anti-psychotic drugs.

691S

Metformin functional pharmacogenomics: STUB1 functions as an E3 ligase for cyclin A and affects metformin sensitivity. *N. Niu, T. Liu, M. Deng, R. Ly, L. Wang.* MPET, Mayo Clinic, Rochester, MN.

Metformin, a widely used anti-diabetic drug, is being considered as a highly promising chemopreventive and therapeutic agent. Although the exact mechanism of action of metformin is unknown, it is thought that metformin could activate AMPK and cause G1 cell cycle arrest. In order to identify biomarkers contributing to metformin response, we took advantage of our Human Variation Panel, a 300 EBV-transformed lymphoblastoid cell line (LCL) model with extensive omic data, and conducted a metformin pharmacogenomic study. The expression of *STUB1*, known as an E3 ubiquitin-protein ligase, was found to be associated with metformin cytotoxicity (IC50 value) with p-value $<10^{-4}$. Functional validation study using siRNA showed that knockdown of *STUB1* significantly altered metformin response in two basal-like breast cancer cell lines, MDA-MB-231 and Hs578T. To further investigate the mechanism by which *STUB1* influences metformin response, we performed mechanistic studies using MDA-MB-231 breast cancer cell line as well as *STUB1* knock-out mouse embryonic fibroblast (MEF) cells. Knockdown of *STUB1* in MDA-MB-231 cell showed an increased proportion of cells in the S phase by flow cytometry and a significant increase in cyclin A protein level by Western Blot. Cyclin A is known to be involved in G1/S transition. In our study, we found that STUB1 interacted with cyclin A. In MEF cells, knockout of *STUB1* caused an increase in cyclin A protein level as well as a decrease in the ubiquitin level of cyclin A. In addition, cyclin A protein was found to be degraded in a proteasome-dependent manner. Thus, we propose that the E3 ubiquitin ligase STUB1 influences metformin response by facilitating proteasome-mediated degradation of cyclin A. In conclusion, genome-wide association study (GWAS) using a genomic data-enriched LCL model system, together with functional validation and mechanistic study using cancer cell line, could help us to identify novel genetic biomarkers involved in metformin response.

692M

Association study of polycystic ovary syndrome with two single nucleotide polymorphism of follicle stimulating hormone receptor in Iranian patients: rs1394205 and rs6166. *B. Dehghan Banadaki[1,2], P. Afsharian[2], M. Shiva[3], SH. Zarei Moradi[2].* 1) Faculty of basic sciences and advanced technologies in biology ,University of science and culture, Tehran, Iran; 2) Department of Genetics at Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran; 3) Department of Endocrinology and Female Infertility at Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran.

Background: Polycystic ovary syndrome (PCOS) is the most common form of WHO II anovulatory infertility, affecting women in reproductive age. The clinical or biochemical hyperandrogenism, menstrual dysfunction with anovulation and presence of polycystic ovaries are the features for PCOS evaluation. Gonadotropins e.g. Follicle-stimulating hormone (FSH) are fundamental for patients undergoing assisted reproductive technology (ART). Therefore FSH receptor (FSHR) has a critical role in follicular development and ovulation induction. The FSHR gene is highly polymorphic and we have studied two functional polymorphisms (rs6166 (N680S), rs1394205 (-29G>A)) in this gene. Methods: The Genomic DNA was extracted from peripheral blood of candidates. The target zones containing single nucleotide polymorphisms (SNPs) -29 and 680 were amplified by polymerase chain reaction (PCR). The allelic and genotypic distributions of these two SNPs were determined by restriction fragmental length analysis (RFLP) technique. We studied rs1394205 on 102 PCOS patients, 100 drug controls (women with normal ovulation and referred for ART because of husband's infertility disorder) and 90 fertile controls. For evaluating rs6166 we experiment on (73 PCOS patients, 54 drug control, 100 fertile controls). Both PCOS and drug-control groups were undergone same IUI protocols and the clinical, surgical and drug history of all subjects were observed and number of follicles $\geq$ 15, as a feature of drug response, were evaluated. Final data were analyzed by chi-square test in SPSS statistical software. Results: PCOS patients in comparison to other two control groups, showed significantly association with N680S (p=0.0008). Genotype distribution of SS, SN and NN in PCO group was 37%, 42% and 15% respectively. According to drug response (exogenous FSH), 30% of PCOs bearing SS genotype were resistant to drug in comparison to 5% of drug controls with the same genotype. This difference in drug resistance was observed in SN genotype too, 42% in PCOs vs. 20% in drug controls. In SS genotype, 70% of both groups respond to treatment. 29G>A polymorphism did not show any significant difference in three groups. Clinical data shown that Patients with normal BMI (18–25 kg/m2) had a better chance to ovulation induction response. Conclusion: Our observation showed great correlation between rs6166 polymorphism and ovulation induction response in PCOS.

693S

***In Vitro* system to assess functional impact of heterozygous variants and complex combinations of variants in dihydropyrimidine dehydrogenase.** *C.R. Jerde, S.M. Offer, R.B. Diasio.* Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN.

Dihydropyrimidine dehydrogenase (DPD, encoded by *DPYD* [MIM 612779]) is the rate-limiting enzyme of the uracil catabolic pathway and, as such, is additionally responsible for the catabolic degradation of the commonly prescribed chemotherapy drug 5-fluorouracil (5-FU). DPD deficiency [MIM 274270] has been associated with increased incidence of severe adverse toxicity (grade 3 or greater), and in some cases death, following treatment with the standard dose of 5-FU. Clinical studies have established that a 30% reduction in DPD function is adequate to increase the risk for toxicity. While there is growing evidence that certain rare genetic variants in *DPYD* are risk alleles for DPD deficiency, the exact contribution of these variants to the disorder is not well understood, particularly in the heterozygous state. The objective of the present study was to develop a reliable system to rapidly evaluate genetic variants that impair DPD activity by 30% or greater as heterozygous alleles. Previously, our lab has developed and optimized methods for evaluating single variants in vitro that rely upon transient expression of DPD in mammalian cells that have undetectable endogenous levels of the protein (*Cancer Res.* March 15, 2013. 73:1958–1968). In the current study, we describe a method that represents an improvement upon that system and permits the study of heterozygous *DPYD* alleles, as well as complex DPYD genotypes. Using this improved method, twice as many cells were double-positive for both copies of the gene compared to co-transfection with two vectors. Additionally, both copies of the gene were shown to be equally expressed at the mRNA and protein levels in target cells. The modeling of a catalytically-null allele in the heterozygous state resulted in in vitro enzyme activity that was 50% of that for the wildtype positive control. Collectively, these data suggest that this system may be applicable for determining which single variants in the heterozygous state and in complex *DPYD* genotypes contribute to DPD deficiency.

694M

***In silico* gene expression results enhance understanding of association studies.** *L. Li[1], E.R. Gamazon[2], T. Johnson[3], C.F. Xu[4], D.M. Waterworth[5], M.G. Ehm[6], N.J. Cox[2].* 1) StatGen, PCPS, GlaxoSmithKline, RTP, US; 2) Section of Genetic Medicine, Department of Medicine, University of Chicago; 3) StatGen, PCPS, GlaxoSmithKline, UK; 4) Genetics, PCPS, GlaxoSmithKline, UK; 5) Genetics, PCPS, GlaxoSmithKline, Philadelphia, USA; 6) Genetics, PCPS, GlaxoSmithKline, RTP, USA.

With the wealth of publically available gene expression data, there is growing excitement in the genetics field that integrating expression information into analyses of genetic and clinical phenotype data will enable a more informative interpretation of findings, and clearer hypotheses for follow-up studies. When expression, genetic and clinical data are all available for the same individuals, a three-way association analysis can be used to test the hypothesis that a given genetic variant affects the clinical phenotype via gene regulation. However, this approach is limited by the infeasibility of obtaining expression data from relevant tissues in the large sample sizes needed to identify associations between genotypes and clinical phenotypes. As a practical alternative, numerous previous studies have asked whether genetic variants associated with clinical phenotypes are also associated with gene expression in independent convenience samples; so-called expression quantitative trait loci (eQTLs). Here, we further refine this approach by developing a polygenic prediction model for expression levels of genes of interest, which can be constructed using publicly available expression data, and using this to derive in silico expression data of the study sample, enabling further analysis to understand findings. The STabilisation of Atherosclerotic plaque By Initiation of darapLadIb TherapY (STABILITY) trial was a cardiovascular outcomes study involving 15,828 patients which evaluated the efficacy of darapladib (an Lp-PLA2 inhibitor) for stable coronary heart disease (CHD). We conducted genome-wide association study for clinical and biomarker traits to identify genetic variants associated with differential drug response. We identified associated variants ($p < 1e\text{-}5$) that were known eQTLs and therefore sought to determine whether these variants contributed to drug response via gene regulation. To address this question, we used publicly available expression data to construct polygenic prediction models for expression levels of the relevant genes. We then used these models to predict expression levels for individuals in the STABILITY trial. Stronger association of these "in silico gene expression levels" with clinical phenotype than for the single genetic variants would highly suggest the effect of gene regulation, warranting further follow up analysis with clearer hypothesis.

695S

Mapping histone modifications provides novel insights in pharmacogenomic discovery. *C. Liu[1], X. Zhang[1], L. Godley[2], M. Dolan[2], W. Zhang[1].* 1) University of Illinois, Chicago, Illinois, USA; 2) University of Chicago, Chicago, Illinois, USA.

Personalized medicine proposes the customization of healthcare and optimal therapies based on the context of an individual patient's genetic make-up and/or other clinical information. Recent progress in pharmacogenetic and pharmacogenomics studies has indicated that both genetic and epigenetic factors may influence drug response to therapeutic treatments. Our objective was to identify histone modification markers associated with drug toxicities using the Lymphoblastoid cell lines (LCL) model, on which we have access to a variety of genetic, epigenetic, and phenotypic data, including the genetic variants associated with chemotherapeutic agents. Specifically, the ChIP-seq data for four post-translational modifications of histone H3 (H3K4me1, H3K4me3, H3K27ac, and H3K27me3) in ten Yoruba LCLs were collected from a publicly available dataset (GSE47991). Drug response association results for five drugs (carboplatin, cisplatin, etoposide, daunorubicin, and ara-C) in the same cell lines were collected from our previously developed pharmacogenetics-cell line database: PACdb (http://www.pacdb.org/). Association studies revealed a substantial number of genes whose local histone modification associated with cytotoxicities based on Akaike information (AIC) and p-value mixed criteria. For example, local H3K4me1 modification levels in 72 genes were observed to be associated with the IC50 of carboplatin. Group comparisons showed that different histone markers exhibited different strength of association. Overall, H3K4me4, a histone mark known to be associated with active gene expression, outperformed other histone markers as a variable for cytotoxicity. Our previous studies had identified a list of potential genes whose expression levels were associated with these drug response phenotypes. We showed that a subset of those associations could be explained by variations of its local histone modification levels. For example, the expression of ARL4A (encoding ADP-ribosylation factor-like 4A) had been identified to be highly associated with the IC50 of cisplatin ($p < 0.001$). Our results further showed this association could be related to the modification in its local H3K4me3. In summary, our whole genome approach provided a global picture of the contributions of histone modification markers to a series of chemotherapeutic-induced cytotoxicity phenotypes. The mapping of histone modification markers provided novel insights into our previous pharmacogenomic findings.

696M

Gene Expression Based Models for Cancer Drug Sensitivity. *U. Ozbek.* Icahn School of Medicine at Mount Sinai, New York, NY.

Pharmacogenomic studies aim to elucidate the genetic or genomic features contributing to the efficacy of medications. The analysis of these data are challenging due to the large number of genomic features that must be investigated. Shrinkage or regularization techniques can help identify the most important features that can predict sensitivity to drug. This work reviews the properties of several regularization methods, such as ridge regression, lasso, and elastic net, using gene expression data from the Cancer Cell Line Encyclopedia project for predicting the sensitivity of 74 lung cancer cell lines to the drug nilotinib. An empirical illustration of the predictive performance of the regularization methods are provided based on three approaches - the first that directly uses gene expressions, and two methods that use principal components of the gene expressions. In this study, ridge regression models fitted slightly better than lasso and elastic net models.

697S

Statin-induced expression change of *INSIG1* in lymphoblastoid cell lines is correlated with plasma triglyceride statin response in men. *E. Theusch, K. Kim, K. Stevens, D. Naidoo, M. Medina.* Children's Hospital Oakland Research Institute (CHORI), Oakland, CA.

Statins are widely prescribed to reduce cardiovascular disease (CVD) risk. Statins primarily lower plasma low-density lipoprotein cholesterol (LDL-C) levels, but they also cause modest reductions in plasma triglycerides (TG), an independent CVD risk factor, in most people. There is extensive inter-individual variability in statin response, but only a small proportion of this variance has been explained by phenotypic characteristics or by genetic variants identified from association studies. As an alternative approach, we studied statin response variability using RNA-seq data from control and statin-treated lymphoblastoid cell lines (LCLs) derived from 100 Caucasian and 50 African American participants of the Cholesterol and Pharmacogenetics (CAP) simvastatin clinical trial (40 mg/day for 6 weeks). 38% of human genes were statin responsive, and 64% of Epstein-Barr virus genes were upregulated with statin treatment ($q<0.0001$, $N=150$). To identify candidate statin efficacy genes, statin-induced changes in gene expression were correlated to plasma LDL-C response and plasma TG response. Initially, we found 151 genes with expression changes that were significantly correlated with LDL-C response (FDR=5%). Next, we adjusted the gene expression changes for potential confounders using probabilistic estimation of expression residuals (PEER) and found that the first two PEER hidden factors were significantly correlated with LDL-C response ($p<0.05$) but not TG response. Using the PEER-adjusted data, changes in 23 genes were significantly correlated with TG response (FDR=15%). Insulin-induced gene 1 (*INSIG1* [MIM 602055]), a well-known regulator of cholesterol homeostasis, had expression changes strongly correlated with TG response (Spearman's $\rho=0.32$, $q=0.11$, $N=150$), and this correlation was driven by men (sex interaction $p=0.0055$). In Caucasians, we identified a SNP that was associated with *INSIG1* expression changes ($N=99$ LCLs, $p=5.4\times10^{-5}$) as well as with TG response in the combined CAP and PRINCE statin trial populations ($p=0.0048$, $N=1890$), predominantly in men. Statin-induced changes in *INSIG1* alternative splicing were also correlated with TG response in men, and a combined model including *INSIG1* expression level and splicing changes accounted for over 28% of the variance in plasma TG statin response in men ($p<5\times10^{-6}$, $N=86$). In summary, our results strongly suggest that variation in *INSIG1* contributes to statin-induced changes in plasma TG in a sex-specific manner.

698M

Accelerating Drug Development with 23andMe Phenome-Wide Association Studies. *F. Sathirapongsasuti, B.T. Naughton, J.L. Mountain, D.A. Hinds, J.Y. Tung, C.Y. McLean.* 23andMe Inc., Mountain View, CA.

Ninety percent of drugs that are deemed successful in animal models fail human clinical trails, because of adverse drug reactions or lack of efficacy. Human genetics can be a powerful tool in the efforts to increase success of drug development through validation of target genes and prioritization of research programs. 23andMe phenome-wide association studies (PheWAS) leverage extremely large sample sizes and phenotypic breadth to yield valuable insights with direct relevance to potential effects in humans. By mining publicly available drug databases for associations between drug target genes and genes harboring variants significantly associated with one or more 23andMe phenotypes, we found over 2,800 drug-phenotype links. Both positive validation of likely successful drug targets and prediction of adverse drug reactions are seen in the 23andMe data, including a replication of association between rare loss-of-function variants in PCSK9 and LDL cholesterol level. A retrospective look at both successful and unsuccessful drug trials confirms our findings, with substantial increase in specificity of predicting successful trials as well as predicting trail failure based on lack of association. Our results support using PheWAS to inform ongoing target validation and identify new drug targets.

699S

Association of C677T polymorphism of the MTHFR gene with toxicity in breast cancer patients treated with FEC chemotherapy. *A. Ramos-Silva[1 y 2], R. Ramírez'[1 y 3], IA. Gutíerrez[1 y 3], OM. Soto[1 y 2], DI. Carrillo[1 y 3], AM. Puebla[4], AR. Rincón[2 y 5], MP. Gallegos[1].* 1) Lab de Genética Molécular. Div Med Mol, CIBO, IMSS, Guadalajara, Jal,. México; 2) Doctorado en Farmacología. CUCS, U de G; 3) Doctorado en Gentica Humana. CUCS, U de G; 4) Laboratorio de Inmunofarmacología. Departamento de Farmacología. CUCEI, U de G; 5) Departamento de Biología Molecular y Genómica. CUCS. U de G.

Background: The influence of C677T polymorphism in the MTHFR gene involved in metabolism of chemotherapeutic agents has been studied in different cancers. MTHFR plays a major role in folate metabolism and consequently could be an important factor for the efficacy of a treatment with FEC (fluorouracil-epirubicin-cyclophosphamide) in breast cancer patients. Our aim was to evaluate the association of C677T polymorphism with toxicity effects in breast cancer patient's neoadjuvantly treated with FEC chemotherapy. Methods: DNA genomic samples from 525 patients (UMAE gynecology and obstetrician Hospital, CMNO, IMSS), that received FEC neoadjuvant chemotherapy; were included in the study. Protocol was support by FIS/IMSS/PROT/G13/1231. The C677T polymorphism was determinate by polyacrylamide gels electrophoresis, previously PCR and HincIII restriction enzyme analysis. The association was determinate by odds ratio. Results: The genotype 677TT was associated with gastrointestinal toxicity (diarrhea grade III) in breast cancer patient response at FEC chemotherapy [1.9(IC95% 1.06–3.53), p=0.028] and genotype 677CT was associated with hematological toxicity (neutropenia) in breast cancer patients non responder at FEC chemotherapy [1.8(IC95% 1.03–3.41), p=0.046]. Conclusion: The polymorphism C677T could be a good marker of toxicity in breast cancer patients treated with FEC chemotherapy in the analyzed sample.

700M

Integration of Genetics into Clinical Development. *D. Waterworth[1], L. Li[2], L. Warren[2], A. Yeo[3], J. Aponte[4], M. Nelson[2], S. Chissoe[4].* 1) Genetics, PCPS, GlaxoSmithKline, King of Prussia, PA; 2) Statistical Genetics, PCPS, GlaxoSmithKline, RTP, NC; 3) Genetics, PCPS, GlaxoSmithKline ,Stevenage UK; 4) Genetics, PCPS, GlaxoSmithKline, RTP, NC.

Pharmacogenetic analysis is commonly initiated after a clinical trial completes and questions arise around interindividual variation in safety or efficacy. This approach can limit the number of pharmacogenetic studies conducted, but it can also limit the opportunity to identify important genetic predictors and delay their potential impact on drug development. We sought to circumvent these issues in the STABILITY study that tested darapladib, an Lp-PLA2 inhibitor, against standard of care for stable coronary heart disease (CHD). We genotyped 13,557 patients with the Illumina Omniexpress Exome chip prior to study completion. The trial failed to demonstrate efficacy for the primary endpoint of major adverse coronary events (MACE). Despite the large size of this study, no genetic variants reached genome-wide significance for MACE or secondary endpoints. However, we demonstrate that there is little power to detect genetic effects unless they are larger than the drug effect assumed in the design the study. Power was further compromised in this study by the limited effects of darapladib on efficacy outcomes. In spite of the failure to identify pharmacogenetic predictors of darapladib efficacy, the information created through prospective genotyping can be valuable in support of further drug discovery and development efforts. CHD is a particularly heterogeneous condition and having the genetic data along with extensive clinical and biomarker data enables us to try to identify patient subgroups that might benefit future CHD outcome trials. Given the exome content of the genotyping array used, we can determine the proportion of subjects who may have rare conditions leading to CHD, which may be commingled with more common forms. In performing analysis of baseline biomarker data, we identified 24 subjects with a somatic JAK2 mutation (V617F) using platelet count as a trait (causes erythrocytosis) and 37 subjects with dominant familial hypercholesteremia due to an APOB mutation (R3527Q). We are performing a thorough review of such conditions to determine their frequency within the trial population, their effects on intermediate traits and their collective impact on outcomes. This experience, along with improvements in genotyping technology and reduction in costs, has added to our understanding of the value of integrating genetics into clinical development and we are currently transitioning toward a fully integrated approach.

701S

A genome-wide association study on antipsychotic-induced body weight gain dissecting the CATIE sample. *M. Maciukiewicz[1], D. Mueller[1], EJ. Brandl[1], AK. Tiwari[1], CC. Zai[1], NI. Chowdhury[1], T. Arenovich[1], JJ. Chen[2], JL. Kennedy[1].* 1) Pharmacogenetics Research Clinic, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada; 2) University of Hong Kong, Department of Pathology,Pokfulam, Hong Kong.

Antipsychotic-induced weight gain (AIWG) is a severe and common side effect often leading to patient non-compliance and increased morbidity and mortality due to metabolic syndrome and cardiovascular events. Genetic predictors would be extremely helpful to identify subjects at risk prior to antipsychotic exposure and first genetic tests are now becoming commercially available. Sample (n=358) consisted of 251 individuals of European and 107 of African-American ancestry with well-documented information on weight changes. Exclusion criteria included a) related individuals, b) available genotype rate of less than 95%, c) mismatch between genetic markers and assigned sex status and d) heterozygosity of more than four standard deviations from the mean. SNPs with MAF < 5%, and genotypes which deviated significantly from HWE were removed. A series of mixed models was used including all potentially relevant covariates and change in BMI over a period of three months as the outcome. In addition, interactions between time and all other predictors were considered. The final model included a random intercept and slope associated with time, study medication group and baseline BMI. Standard quality control workflow was applied to the genotype data and we analyzed 328,733 SNPs. To avoid population stratification, we plotted the MDS components and selected patients within the cluster corresponding to European ancestry as the largest cohort. The GWAS analysis was conducted on 189 individuals. The SNP vs BMI association analysis was carried out using the R package: 'nlme' None of the SNPs was significant at the genome-wide threshold of p = 5×10–8. The top hit of the GWAS was rs12924003 (p=1.06×10–5) located downstream of the SAL-1 gene. The second hit, SNP rs4771655 (p=1.91 ×10–5) is ~194kb upstream of IRS2 gene. The third hit, rs4751427 (p=2.4 ×10–5) is located ~59kb upstream of the Neuropeptide S gene. Our analyses did not detect an association when considering the commonly applied genome-wide correction threshold for multiple testing. Two of our top hits, IRS2 and NPS, were previously shown to be involved in regulation of insulin sensitivity and food intake in other populations.

702M

Genome-wide association study of resistant hypertension in INVEST. *N. El Rouby[1], CW. McDonough[1], Y. Gong[1], CJ. Pepine[2], A. Takahashi[3], T. Tanaka[3], M. Kubo[3], RM. Cooper-DeHoff[1,2], JA. Johnson[1,2].* 1) Pharmacotherapy and Translational Research, University of Florida, Gainesville, FL; 2) Division of Cardiovascular Medicine, Department of Medicine, University of Florida, Gainesville, FL; 3) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

Resistant hypertension (RHTN) is defined as a blood pressure (BP) that remains ≥140/90 mm Hg despite the use of ≥ 3 antihypertensive drugs or BP < 140/90 mm Hg requiring ≥ 4 antihypertensive drugs. RHTN is associated with increased risk for adverse cardiovascular outcomes, especially stroke; therefore, we aimed to identify the association between genetic markers and RHTN using a genome wide association study (GWAS) to further characterize those at risk for RHTN. We genotyped 1194 participant samples from the INternational VErapamil-SR trandolapril STudy - GENEtic Substudy (INVEST-GENES) on the Illumina OmniExpressExome chip. INVEST was a clinical trial that compared cardiovascular outcomes in patients with both hypertension and stable coronary artery disease (CAD) after randomization to either a calcium antagonist strategy "CAS" (verapamil-SR ± trandolapril) or non-calcium antagonist strategy "NCAS" (atenolol ± HCTZ). Non-study drugs were used for BP control except for atenolol in CAS and verapamil in NCAS. In this analysis, a case was defined as a BP ≥ 140/90 despite use of ≥ 3 drugs, or use of ≥ 4 drugs regardless of BP. Overall, BP control in both treatment strategies was similar, which allowed for pooling participants. Logistic regression analysis was conducted separately in 657 European Americans and 537 Hispanics using an additive genetic model, adjusting for age, gender, body mass index, diabetes, smoking, left ventricular hypertrophy, congestive heart failure, stroke, percutaneous coronary intervention, principle components, and treatment assignment. Results from both race groups were combined using a meta-analysis. None of the SNPs reached genome wide significance. The meta-analysis identified a missense SNP (rs3732378) in a chemokine receptor gene *CX3CR1* leading to a Threonine to Methionine substitution. The A allele was associated with lower risk of RHTN in Europeans (OR 0.51, $P = 2\times10^{-4}$). This association was replicated in Hispanics (OR 0.47, P=0.005) with a meta-analysis $p = 8\times10^{-6}$. *CX3CR1* is a specific receptor for Fractalkine, a chemoattractant chemokine involved in leukocyte trafficking, inflammation and endothelial dysfunction. The SNP has been reported to alter the ligand-receptor affinity and has been associated in previous studies with reduced risk for atherosclerosis and CAD. In conclusion, we identified a missense SNP in *CX3CR1* that may be associated with reduced risk for RHTN. Additional research is needed to validate our findings.

703S

Identification of nonresponders through the use of EMRs identifies MSRA as a potential controller of response of asthmatics to inhaled corticosteroids. *M.E. March, P.M.A. Sleiman, F. Mentch, E. Hysinger, K. Thomas, C.E. Kim, C. Hou, H. Hakonarson.* Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

Asthma is a chronic inflammatory condition of the lungs, characterized by excessive responsiveness of the lungs to stimuli, in the forms of infections, allergens, and environmental irritants. Currently, 22.9 million Americans suffer from asthma, and the prevalence has increased dramatically since 1980. Asthma is the leading chronic illness in U.S. children, with 6.8 million affected in 2006. Inhaled corticosteroids (ICS) are a common, effective treatment for asthma. ICS act locally in the lungs to reduce the chronic inflammation that characterizes asthma; this reduction in inflammation reduces the frequency and severity of asthma flares that occur in response to acute triggers. Most studies examining the genetic components of the response of patients to ICS treatment focuses on changes in pulmonary function test results that occur following weeks or months of treatment. In the current study, we wished to determine the possibility of identifying responders and nonreponders to ICS therapy through the use of abstracted electronic medical records (EMRs). We developed an approach to score a patient's response to ICS by evaluating changes in the frequencies of hospitalization and emergency room visits, along with changes in ICS dose and the need for additional prescription of stronger therapy in the form of oral steroids. The use of this response score as a quantitative trait in an African American cohort in a genome wide association study revealed a single significant locus, represented by four significant SNPs located within the MSRA gene on chromosome 8. Replication efforts are currently underway.

704M

A genome-wide association study of hepatic toxicity of methotrexate therapy for rheumatoid arthritis. *H. Mo[1], N. Braggs[2], L. Bastarache[1], R.J. Carroll[1], A. Shah[5], D. Roden[2], S. Raychaudhuri[3,4], E.W. Karlson[3,4], J.C. Denny[1,2].* 1) Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Department of Medicine, Vanderbilt University, Nashville, TN; 3) Brigham and Women's Hospital, Boston, MA; 4) Harvard Medical School, Harvard University, Boston, MA; 5) Prince of Wales Clinical School, University of New South Wales, Sydney, NSW, Australia.

Low dose methotrexate (MTX) is the first-line disease modifying anti-rheumatic drug (DMARDs) for rheumatoid arthritis (RA) and other inflammatory arthritides. While it is generally well tolerated, hepatic toxicity is among one of the most concerning side effects with a cumulative incidence of 31%. We studied individuals identified from BioVU, a DNA biobank linked to de-identified electronic medical records (EMR) at Vanderbilt, to identify genetic susceptibility of hepatic toxicity of MTX therapy. **Methods:** We identified and genotyped individuals with RA in BioVU using a previously validated algorithm. We evaluated RA individuals for presence of MTX exposure, and then used a combination of algorithms and manual review to ascertained 66 cases and 308 controls with OmniExpress GWAS data and 58 cases and 256 controls with Illumina HumanExome array data. Only individuals of European ancestry were included. MTX hepatic toxicity was defined as liver enzymes (LE, either AST or ALT) greater than 2 times the upper limit of normal (ULN, i.e. 80 IU/L) without other identified causes. Control groups were RA patients who had liver enzyme measurements without abnormal values and who were compliant with MTX treatment. Genotyping was done with Illumina OmniExpress and HumanExome platforms. **Result:** The SNPs with the best P-values in GWAS study was rs913898 (P=9.3e-7, OR=15.51, between *KLF5* and *KLF12*). Other top hits in GWA study include: rs564745 near *EVI5L* with p=4.1e-6 (OR=4.11) and rs12537903 in *NPSR1-AS1* with p=5.0e-6 (OR=4.3). The strongest association on the Exome study is rs7335046 (P=6.6e-5 and OR=4.59, in *MIR548AN* and near *UBAC2*). Among the missense Exome markers with negative PAM30 amino-acid replacement scores, the top hits were *CEBPZ* (P15S, p=0.00041), *LSG1* (L92P, p=0.00063), *ACOXL* (T255M, p=0.00077), *WSB1* (L16S, 0.00082). Previously, *ABCC1*, an efflux transporter involved in MTX metabolism, has been associated with MTX toxicity. In this region, our strongest signal was with rs4781712 with p=0.00026 (OR=0.49). **Conclusion:** In this GWAS of MTX-induced liver injury, we replicated a previously-reported association with *ABCC1* and identified several new potential signals that require further investigation.

705S

A genome-wide association study identified variants in KCNIP4 associated with ACE inhibitor induced cough. *J.D. Mosley[1], C.M. Shaffer[1], S.L Van Driest[2], P.E. Weeke[1], Q.S. Wells[1], J.H. Karnes[1], D.R. Velez Edwards[3], W. Wei[4], P.L. Teixeira[4], L. Bastarache[4], D.C. Crawford[3], J.A. Pacheco[8], T.A. Manolio[5], E.P. Bottinger[6], C.A. McCarty[7], J. Linneman[7], W. Thompson[8], R.L. Chisholm[8], G.P. Jarvik[9], D.R. Crosslin[9], D.S. Carrell[10], E.B. Larson[10], H. Jouni[11], I.J. Kullo[11], G. Tromp[12], K.M. Borthwick[12], H. Kuivaniemi[12], M.D. Ritchie[13], J.C. Denny[1,4], D.M. Roden[1], eMERGE II Network.* 1) Department of Medicine, Vanderbilt University, Nashville, TN; 2) Department of Pediatrics, Vanderbilt University, Nashville, TN; 3) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 4) Biomedical Informatics, Vanderbilt University, Nashville, TN; 5) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 6) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, NY; 7) Essentia Institute of Rural Health, Duluth, MN; 8) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 9) Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington, Seattle, WA; 10) Group Health Research Institute, Seattle, WA; 11) Division of Cardiovascular Diseases, Mayo Clinic, Rochester MN; 12) The Sigfried and Janet Weis Center for Research, Geisinger Health System, Danville, PA; 13) Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA, USA.

The most common side effect of angiotensin converting enzyme inhibitors (ACEis) is a persistent cough, which effects up to 8% of whites, and often requires a switch to other drug classes. The etiological mechanism of the cough is unknown. We conducted a genome wide association study (GWAS) to identify common SNPs associated with ACEi-induced cough. The discovery cohort comprised 6,007 whites of European ancestry (EAs) from six sites in the electronic MEdical Records & GEnomics (eMERGE) network. The phenotype was defined as a clinically diagnosed cough attributed to an ACEi and recorded in the drug sensitivity/allergy section of the electronic medical record (EMR). Controls were individuals with at least 6 months of ACEi use and no history of an ACEi-associated cough. We performed single SNP tests of association using logistic regression with 3,194,795 SNPs, assuming an additive genetic model, and adjusting for principal components. The most significant SNPs were evaluated in an replication set of 926 eMERGE EA subjects. The GWAS in the discovery set of 1,346 cases and 4,661 controls identified significant associations located in a single region of chromosome 4 within an intron of the gene KCNIP4. The strongest association was in rs145489027 (MAF=0.33, OR=1.3 [95%CI: 1.2–1.4], p=2.4×10–8). Four of the six most significantly associated SNPs in this region with p<1×10–6 were available for analysis in a replication set of 157 cases and 769 controls. The strongest association among these 4 SNPs was in rs7675300, also in KCNIP4, (MAF=0.33, OR=1.3 [1.01–1.70], p=0.04). KCNIP4 is a member of the KChIP family of EF hand-containing calcium-binding proteins and is primarily expressed in neuronal structures where it regulates Kv4 potassium channels. These data suggest that ACEi-induced cough may be mediated, in part, through a neuronally-mediated mechanism involving KCNIP4.

706M

International Clopidogrel Pharmacogenomics Consortium Genome Wide Association Study Identifies Novel Variants for Clopidogrel Response. *A. Shuldiner, International Clopidogrel Pharmacogenomics Consortium.* Prof/Med,Endocr/Diabetes/Nutr, University of Maryland, Baltimore, Baltimore, MD.

Clopidogrel (Plavix) is a commonly used antiplatelet medication in coronary artery disease patients to prevent recurrent thrombotic events. It is a pro-drug that requires activation by CYP2C19, and common loss of function variants of CYP2C19 (e.g., CYP2C19*2) are associated with decreased antiplatelet responsiveness phenotype but explained only a marginal part of this trait. Heritability estimates are high and suggest that there are additional genetic determinants of clopidogrel response. To identify these variants, the International Clopidogrel Pharmacogenomics Consortium (ICPC) amassed clopidogrel response phenotypes (namely on-treatment ADP-induced platelet reactivity and cardiovascular outcomes) and DNA from more than 6,000 clopidogrel-treated patients. A GWAS (Omni Express with exome coverage) was conducted in 2,721 Caucasian patients. We observed strong association between on-treatment platelet reactivity phenotypes and the CYP2C19 locus on chromosome 10 (best SNP rs1926711; $p=6.25\times10^{-32}$). After adjustment for the known CYP2C19*2 loss of function variant (rs4244285; $r^2=0.963$ with rs1926711), there remained residual association just upstream of CYP2C19 (e.g., rs1998591; $p=1.66\times10^{-05}$) suggesting other variants in this region as determinants of clopidogrel response. Other loci nominally associated with platelet response to clopidogrel ($p<10^{-5}$) included rs12651351, intergenic between FHDC1 and TRIM2; rs10505836 in PLEKHA5; and rs2295306 in DHRS1. While we conclude that CYP2C19*2 remains a major genome-wide determinant of clopidogrel response, other loci are likely to exist, but play a lesser role. Replication studies, now in progress, will be required to affirm potentially novel loci for clopidogrel response.

707S

Class II HLA variants are associated with differential antibody response to toxoid vaccinations in a cohort of Ugandan infants. *A.J. Mentzer[1,2], A. Muhwezi[3], T. Carstensen[4], G. Smits[5], A. Rautanen[1,2], D. Gurdasani[4], T. Mills[1,2], C. Pomilla[4], H. Akurut[3], D. Kizito[3], S. Lule[3], K. Mohammed[3], K. Elliott[1,2], F. van der Klis[5], P. Kaleebu[3], A. Elliott[3], M. Sandhu[4], AVS. Hill[1,2].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) The Jenner Institute, University of Oxford, Old Road Campus Research Building, Oxford, United Kingdom; 3) MRC/UVRI Uganda, Entebbe, Uganda; 4) The Sanger Institute, University of Cambridge, Hinxton, Cambridge, United Kingdom; 5) National Institute for Public Health and Environment, Bilthoven, The Netherlands.

It is estimated that up to 70% of the variability in response to some vaccines administered in childhood is inherited but the identification of the genetic factors responsible for this variation has not yet been prioritised. We have conducted the first genome-wide association study seeking to identify the genetic variants associated with multiple vaccine responses using 2,181,930 autosomal markers genotyped in 1391 Ugandan children enrolled in the Entebbe Mother and Baby cohort study. Vaccine response was defined as quantitative antibody levels measured against 7 antigens targeting 5 separate pathogens which demonstrated little to moderate correlation within individuals (r=0.02 to 0.58). Strong Class II HLA associations were observed with response to diphtheria toxin (most associated SNP rs2647060 P= 1.70×10^{-26}), pertussis toxin (rs9270894 P=4.14×10^{-27}), pertussis pertactin (rs9275391 P=5.32×10^{-25}), and pertussis filamentous haemagglutinin adhesin (rs3104406 P=1.10×10^{-10}). Little overlap was observed for associated SNPs between each vaccine response measured. No associations of genome-wide significance were observed in HLA with response to tetanus toxin, or against attenuated measles or the conjugated Haemophilus influenzae vaccines. Class II HLA genes are critically involved in the stimulation of the adaptive immune response including antibody production against foreign antigens. Toxoid peptide antigens represent the model on which this mechanism was discovered and this study therefore provides a proof-of-concept for the analytical method employed. We will use the same approach to analyse the data from 10,000 children recruited worldwide as part of the newly developed VaccGene initiative coordinated by the Universities of Oxford and Cambridge, UK. The results from this large-scale analysis will be used to understand why HLA is strongly associated with some, but not all, toxoid vaccine responses and to identify the other elements of the immune system associated with inter-individual variability to response to multiple vaccines. These results may have important implications for future vaccine design.

708M

A genome-wide meta-analysis of corticosteroid response in asthmatics identifies a novel associated haplotype on chromosome 4. *Q.L. Duan[1], B.E. Himes[1,2,3], J. Lasky-Su[1], K.G. Tantisira[1,4], S.T. Weiss[1,2].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Partners Personalized Medicine, Partners Health Care, Boston, MA; 3) Children's Hospital Informatics Program, Boston, MA; 4) Pulmonary Division, Brigham & Women's Hospital, Boston, MA.

Corticosteroids are the most common, anti-inflammatory therapy for controlling asthma, a disease affecting over 300 million people worldwide. However, substantial inter-individual variability in corticosteroid response has been described, with up to 25% non-responsive patients. Multiple loci have been associated with this drug response, but account for only part of the heritability. The current genome-wide association study (GWAS) of corticosteroid response tests nearly five million markers in 723 asthma patients from four independent clinical trials: Childhood Asthma Management Program (CAMP); Leukotriene modifier or Corticosteroid Salmeterol (LOCCS); Childhood Asthma Research and Education (CARE) Network and the Asthma Clinical Research Network (ACRN). All participants had clinically diagnosed mild to moderate asthma. The physiological response to inhaled corticosteroids (ICS) was measured as a percent change in pre-bronchodilator FEV_1 (forced expiratory volume in one second) following four to eight weeks of drug treatment. Genome-wide SNP data was available for all trials, which were previously generated using two genotyping platforms: the Illumina HumanHap550 or 610 BeadChips (CAMP and LOCCS) and the Affymetrix Genome-Wide Human SNP Array 6.0 (ACRN and CARE). Due to the poor overlap among SNPs from the two platforms, however, imputations was performed using the Markov Chain Haplotyping (MaCH) software based on SNPs available in the June 2010 data release of the 1000 Genome Project. Approximately 5 million imputed SNPs (minor allele frequency > 0.05) were tested for association with ICS response using dosage data and an additive, linear regression model in PLINK, with adjustments for covariates (i.e. height, age and gender). Combined P-values were calculated using the Liptak's weighted Z method. While no single SNP achieved genome-wide significance, we identified three SNPs on chromosome 4 that tag a common (35%) haplotype associated with ICS response across our four asthma trials (combined P = 3.13E-11). Replication analysis of this haplotype is underway in an additional asthma drug trial. The current GWAS identifies a common haplotype associated with variable ICS response in asthma, which may have wider health care implications as corticosteroids are prescribed for the treatment various inflammatory diseases.

709S

Additional genetic risk factors for carbamazepine-induced cutaneous adverse drug reactions detected by conditional analysis using HLA-A*31:01 as a covariant in Japanese population. *T. Ozeki[1], T. Mushiroda[1], A. Takahashi[2], M. Kubo[3], Y. Shirakata[4], Z. Ikezawa[5], M. Iijima[6], T. Shiohara[7], K. Hashimoto[4], Y. Nakamura[1].* 1) Laboratory for Pharmacogenomics, RIKEN Cen. Integr. Med. Sci., Yokohama, Kanagawa, Japan; 2) Laboratory for Stastical Analysis, RIKEN Cen. Integr. Med. Sci., Yokohama, Kanagawa, Japan; 3) Laboratory for Genotyping Development, RIKEN Cen. Integr. Med. Sci., Yokohama, Kanagawa, Japan; 4) Department of Dermatology, Ehime University Graduate School of Medicine, Ehime, Japan; 5) Department of Dermatology, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 6) Department of Dermatology, Showa University School of Medicine, Tokyo, Japan; 7) Department of Dermatology, Kyorin University School of Medicine, Tokyo, Japan.

An anticonvulsant, carbamazepine (CBZ), is known to show incidences of cutaneous adverse drug reactions (cADRs) including Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN) and drug-induced hypersensitivity syndrome (DIHS). Previously, to identify a susceptible gene(s) for CBZ-induced cADRs, we conducted a genome-wide association study (GWAS) and subsequent HLA typing in 61 cases and 376 patients who showed no cADRs by administration of CBZ (CBZ-tolerant controls), and found that HLA-A*31:01 was present in 60.7% (37/61) of the patients with CBZ-induced cADRs, but in only 12.5% (47/376) of the CBZ-tolerant controls (odds ratio (OR) = 10.8, 95% confidence interval (CI) = 5.9–19.6, $P = 3.64 \times 10^{-15}$). To detect additional genetic marker for CBZ-induced cADRs, we performed genome-wide imputation using the GWAS data and genotype data from 1000 Genomes Project as reference panels by MaCH-Admix software. For a conditional logistic regression analysis of the imputed genotype data, we selected rs1633021 as the covariant with the firmest association observed in the analyzed GWAS peak of association. After the conditional analysis, there were 22 loci that reached GWAS-level significant association ($P < 5 \times 10^{-8}$). SNPs in 17 loci were validated and 2 SNPs rs4488702 and rs16899783 were replicated (Combined population: *P*-logistic = 4.10×10^{-9} and 2.08×10^{-18}). When 2 SNPs rs4488702 and rs16899783 were included as the risk marker for CBZ-induced cADRs in addition to HLA-A*31:01, the population attributable risk percentage was increased from 57% (A*31:01) to 65% (A*31:01 and 2 SNPs).

710M

A Phenome-Wide Association Study of Numerous Laboratory Phenotypes in AIDS Clinical Trials Group (ACTG) Protocols. *S. Pendergrass[1], A. Verma[1], E. Daar[2], R. Gulick[6], R. Haubrich[4], G. Robbins[3], D. Hass[5], M. Ritchie[1].* 1) Center for Systems Genomics, Department Biochemistry & Molecular Biology, Pennsylvania State University, University Park, PA; 2) Department of Medicine, Los Angeles Biomedical Research Institute, Harbor-UCLA Medical Center, Torrance, CA; 3) Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA; 4) University of California San Diego, San Diego, CA; 5) Vanderbilt University, Nashville, TN; 6) Weill Medical College, Cornell University, New York, NY.

Phenome-Wide Association Studies (PheWAS) have the potential to efficiently discover novel genetic associations across multiple phenotypes. Prospective clinical trials data offer a unique opportunity to apply PheWAS to pharmacogenomics. Here we describe the first PheWAS to explore associations between genotypic data and clinical trial data, both pre-treatment and following initiation of antiretroviral therapy. A "pre-treatment" PheWAS considered 27 laboratory variables from 2807 subjects who had participated in 4 ACTG protocols (ACTG384, A5142, A5095 and A5202), and analyzed ~ 5M imputed SNPs. Lowest p-values were for pre-treatment bilirubin, neutrophil counts, and HDL cholesterol levels. These and multiple other laboratory variables matched associations in the NHGRI GWAS Catalog. An "on-treatment" PheWAS considered data from 1181 subjects from A5202. We considered 838 phenotypes and sub-phenotypes derived from 6 variables: CD4 counts, HIV control, fasting LDL cholesterol levels, fasting triglyceride levels, efavirenz pharmacokinetics (PK), and atazanavir PK. We considered 2,374 annotated drug-related SNPs from PharmGKB. Of 23 associations with the lowest p-values (by phenotype), 21 (91%) were with genes with matching biological plausibility: LDL cholesterol levels with *LPL* and *APOE*; triglycerides with *LPL*; CD4 counts with innate immune response gene *TNF*, HIV control with adaptive immune response gene *HLA-DRQA1*, efavirenz PK with *CYP2B6*; atazanavir PK with drug transporter gene *ABCC4*. This analysis highlights the potential utility of PheWAS to evaluate clinical trials datasets for genetic associations.

711S

Genome-wide association study of warfarin maintenance dose in a Brazilian sample. *G. Suarez-Kurtz[1], A.B. Santoro[1], M.R. Botton[2], M.H. Hutz[2], S. Krithika[3], E.J. Parra[3].* 1) Instituto Nacional de Cncer, Rio de Janeiro, Rio de Janeiro, Brazil; 2) Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 3) Department of Anthropology, University of Toronto at Mississauga, Mississauga, Canada.

The anticoagulant warfarin combines several characteristics that make it a case study for the relevance of pharmacogenomics in personalized drug therapy. We carried out a genome-wide association study of warfarin maintenance dose in a sample from Brazil, comprising 177 individuals receiving low warfarin dose (<20 mg/week) and 188 individuals receiving high warfarin dose (> 40 mg/week). The samples were genotyped with Affymetrix Biobank microarrays. Prior to imputation, SNPs were removed from the initial list of autosomal markers based on the following criteria: 1/ Markers classified as CallRateBelowThreshold, OffTargetVariants or Other by the program SNPolisher, 2/ minor allele frequency <1% , 3/ Hardy-Weinberg p-values < 10-5 in the low dose or high dose sample and 4/ missing genotyping rate per SNP > 5% . The program EIGENSOFT was used to carry out a Principal Component Analysis (PCA), in order to identify the presence of population stratification. Imputation was carried out with the program IMPUTE2, using the combined 1000 Genomes panel as reference samples. Four individuals were eliminated from the statistical analysis due to cryptic relatedness (pi-Hat > 0.2) or evidence of substantial East Asian ancestry, based on the PCA plots. For the statistical analysis, we used the program SNPTEST2 (ml method), including as covariates sex, age, BMI, treatment with Amiodarone and the first two PCA scores. An analysis with the program ADMIXTURE indicates that the average European, African and Native American admixture proportions in this sample are 76.8%, 15.1% and 8.1%, respectively. We identified two genome-wide signals, corresponding to the well-known genes VKORC1 (lead SNP rs749671, p=1.08×10−33) and CYP2C9 (lead SNP rs9332220, p=2.5×10−13), as well as numerous suggestive signals (p<10-5). The lead SNP within the CYP2C9 gene seems to be capturing the effect of two known functional alleles (CYP2C9*2 and CYP2C9*3). The minor allele of rs9332220 is in almost perfect linkage disequilibrium with both CYP2C9*2 and CYP2C9*3.

712M

Swedegene: Genome-wide association study of drug-induced agranulocytosis. *M. Wadelius[1], N. Eriksson[2], Q-Y. Yue[3], P.K. Magnusson[4], E. Eliasson[5], H. Melhus[1], P. Hallberg[1].* 1) Dpt of Medical Sciences, Clinical Pharmacology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 2) Uppsala Clinical Research Center and Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 3) Medical Products Agency, Uppsala, Sweden; 4) Swedish Twin Registry, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 5) Clinical Pharmacology, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden.

Background: Agranulocytosis is rare, serious and often immune-mediated condition with a case-fatality rate of around 7–10%. It can be caused by a multitude of drugs, for example sulfasalazine, antithyroid agents, antibiotics, clozapine, aspirin, amiodarone and antiplatelet agents. However, few patients develop agranulocytosis after ingesting these drugs, which suggests that genetic factors could play a role in the pathogenesis. The major histocompatibility complex has been the main implicated region, but genetic studies have so far been small and often lacked replication cohorts. The Swedegene (www.swedegene.se) biobank is collecting patients with adverse drug reactions on a nation-wide basis using spontaneous reports sent to the Swedish Medical Products Agency. Blood for DNA extraction is collected and clinical data is obtained from patient interviews, medical and laboratory records. **Objective**: Our aim is to discover genetic and clinical factors that predispose to drug-induced agranulocytosis with the purpose to prevent this reaction in the future. **Method**: A genome-wide association study (GWAS) on 101 agranulocytosis cases was performed with 5 000 unrelated individuals from the Swedish Twin Registry as controls. Cases were genotyped with the Illumina HumanOmni2.5 array at the Uppsala University SciLife SEQ&SNP Technology Platform. Analyses were performed using the R-package GenABEL, imputations were done with Impute v2 (1000 genomes as reference) and HLA-B imputations with the R-package HIBAG. After quality control and imputation, the merged genotype set contained 6 million SNPs. The significance level was set at $p < 2\times10^{-8}$. **Results**: In the pilot study of the first 101 cases of drug-induced agranulocytosis, the majority of the patients were women and the main causative drugs were sulfasalazine, antithyroid agents and antibiotics. Statistical analyses identified significant associations in the HLA-B region on chromosome 6. After HLA-B genotype imputation, the strongest signal came from HLA-B*27 (OR [95% CI] = 3.67 [2.39, 5.63], $p = 2.99 \times 10^{-9}$). The HLA-B locus remained significant when utilizing 233 controls matched for disease or treatment. **Conclusion**: We have found an association between drug-induced agranulocytosis and the HLA-B locus. The results need to be replicated, and we are proceeding with additional cases from Sweden as well as cases and controls obtained through collaboration with other European countries.

713S

DMET platform identifies unique changes in metabolic gene variants in ethnic Arabs. *S. MAJID[1], PN. Muiya[1], S. Hagos[1], E. Andres[1], B. Baz[1], G. Morahan[2], N. Dzimiri[1].* 1) KFSHRC, RIYADH, India; 2) The University of Western Australia, Perth Australia.

Background: The Affymetrix Drug Metabolizing Enzymes and Transporters (DMET) Plus Premier Pack platform has been designed to genotype about 1936 genes/gene variants thought to be essential for screening patients in personalized drug therapy. These variants include among others, the cytochrome P450s (CYP450s), the key metabolic enzymes for several therapeutic agents, many other enzymes involved in phase I and phase II pharmacokinetic reactions as well as signaling mediators commonly associated with variability in clinical response to numerous drugs not only among individuals, but also between ethnic populations. Materials and Methods: We genotyped 600 Saudi individuals for 1936 variants on the DMET platform to evaluate their clinical potential in personalized medicine in the Saudi population. Results: Approximately 49% each of the 437 cytochrome P450 (CYP450) variants, 56% of the 581transporters, 56 % of 419 transferases 48% of the 104 dehydrogenases and 58% of the remaining 390 remaining variants were detected. A great number of these variants exhibited either significantly higher or lower minor allele frequencies (MAFs) than in other ethnic groups, with more than 70 of them, including the CYPs rs7797834 (CYP51A1_c.1359T>C), rs2470890 (CYP1A2_5347T>C), rs7793861 (CYP51A1_c.*251G>C), rs16947 (CYP2D6_2850C>T) and some NAT2, SLC, UGTB and ABC variants exhibiting MAFs of >0.49. Discussion: The distribution of several variations appears to be unique for our study compared to other ethnic populations. The results point to the need to verify and ascertain the importance of a SNP profile in any given population prior to engaging it in pharmacogenetic screening in personalized medicine. Acquiring adequate prevalence data is, therefore, prerequisite for their employment in clinical routine in any given population.

714M

Selection of cancer patients based on tumor *AKT1* or *PIK3CA* mutation status. *J. Fox[1], P. Kirk[1], J. Whiteley[1], H. Ambrose[2].* 1) Personalized Healthcare, AstraZeneca Pharmaceuticals, Macclesfield, Cheshire, United Kingdom; 2) Innovative Medicines, AstraZeneca Pharmaceuticals, Macclesfield, Cheshire, United Kingdom.

The identification of patients who may experience differential benefit from new cancer medicines as a consequence of the presence of a genetic alteration e.g. mutation in their tumour, requires the timely and accurate implementation of appropriate patient selection methods. Both sourcing and validation of assays for use during drug development bring particular challenges and different solutions are required depending on the clinical situation in which assays will be deployed. We have employed a number of approaches to deliver tests for patient selection suitable for implementation during the initial, exploratory, stages of drug development in man. For logistical reasons a PCR based test was developed in house for *AKT1* whereas for *PIK3CA* a commercially available (also PCR based) Research Use Only (RUO) assay was adopted. Both assays underwent preliminary internal validation prior to transfer to external partners for formal validation before their application for patient testing. Identifying acquired mutations, not present in normal patient tissue, in tumour derived material brings additional considerations to DNA testing beyond that necessary for diagnosis of common inherited disorders. First, tests need to have sufficient analytical specificity and sensitivity to allow accurate assignment of mutation status when the mutant sequence exists in only a minority of the DNA sample. Second, the assays need to deliver accurate results when the quantity and/or quality of the DNA for evaluation is less than optimal due, in part, to the tumour tissue preservation steps used in routine pathology practice. Finally, where tumours originating from different anatomical sites are to be examined, the mutation spectrum covered by the test(s) should be sufficient to identify as many suitable patients as possible. The latter element of test selection/development can be challenging where data defining mutation spectra are limited. Here we describe the steps we have followed to successfully deliver PCR based *AKT1* and *PIK3CA* mutation assays to test a patient selection hypothesis.

715S

Impact of regular physical activity on weekly warfarin dose requirement. *P. Shahabi[1,2], E. Rouleau-Mailloux[1,3], S. Dumas[1,4], Y. Feroz Zada[1], S. Provost[1], J. Hu[1], J. Nguyen[1], N. Bouchama[1], I. Mongrain[1], M. Talajic[1,2], J.C. Tardif[1,2], S. Perreault[2,4], M.P. Dubé[1,2].* 1) Beaulieu-Saucier Pharmacogenomics Centre , Montreal Heart Institute, Université de Montréal , Montreal, QC, Canada; 2) Department of Medicine and Social and Preventive Medicine, Université de Montréal, Montreal, Quebec, Canada; 3) Department of Pharmacology, Faculty of Medicine, Université de Montréal, Montréal, Quebec, Canada; 4) Faculty of Pharmacy, Université of Montréal, Montreal, Quebec, Canada.

Background and aim: Warfarin is the most widely prescribed oral anticoagulant worldwide for the treatment and the prevention of thromboembolic disorders. However, there is a marked inter- and intra-individual variability in the warfarin dose requirement. Genetic factors, including the common variants of vitamin K epoxide reductase complex subunit 1 (VKORC1*2) andcytochrome P-450 2C9 (CYP2C9*2 and *3), can predict up to 40 percent of the variability in warfarin dose and the known non-genetic factors, including weight, height, age and drug interaction, are responsible for another 30%. Hence, the sources of 30% of warfarin dose variability are still unknown. We first investigated the association of regular physical activity (RPA) with warfarin dose in a large cohort of warfarin new users, the Quebec Warfarin Cohort (QWC). We then replicated our findings in an independent on-warfarin hospital-based population, the Montreal Heart Institute (MHI) Biobank. Methods: The QWC isan observational, community-based, prospective cohort of new warfarin users who were recruited at 18 anticoagulation clinics in the Quebec province of Canada. The level of RPA was assessed in 1,064 patients of the QWC using the Stanford Brief Activity Survey. A regression model was used to evaluate the association between the baseline level of RPA and the warfarin dose requirement at 3-month time point after the initiation of treatment. The model was adjusted forheight, weight, age as well asCYP2C9*2, *3 and VKORC1*2 variants.For replication, warfarin dose of prevalent users was modeled in 681 patients of the MHI Biobank in whom the level of RPA was assessed with the Global Physical Activity Questionnaire.The model was adjusted for height, weight, age and CYP2C9 variants in the replication population. Results: Higher levels of RPAwere associated with higher doses of warfarin in both cohorts. In the QWC, univariate regression analysis showed that RPA could explain 5.4% of variance in warfarin dose (P value <0.001); the association was replicated in the MHI Biobank (P value =0.0012; R2=1.7%). In the multiple regression model, RPA remained statistically significant and was found to be independently associated with warfarin dose inboth the QWC (P value < 0.001) and the MHI Biobank (P value = 0.0391). Conclusion: RPA was associated with higher warfarin dose requirement. However, the clinical implications of the results of this work require to be directly investigated by further studies.

716M

Comparison of parents' initial intent and reported sharing of their children's *CYP2D6* research results at three month follow up. *C.A. Prows[1,2], B. McLaughlin[1,4], X. Zhang[1], D. Kissell[1], K. Zhang[1,4], J.B. Harley[3,4], M.F. Myers[1,4].* 1) Division of Human Genetics, Children's Hospital Medical Center, Cincinnati, OH; 2) Department Patient Services, Children's Hospital Medical Center, Cincinnati, OH; 3) Division & Center for Autoimmune Genomics and Etiology, Children's Hospital Medical Center, Cincinnati, OH; 4) Department of Pediatrics, College of Medicine, University of Cincinnati, Cincinnati, OH.

The Clinical Pharmacogenetics Implementation Consortium published guidelines in 2012 regarding the use of codeine based on genotype predicted CYP2D6 phenotype. The recommendations refer to other CYP2D6 opioid substrates; tramadol, hydrocodone and oxycodone. However, *CYP2D6* testing to inform opioid therapy is rare in clinical settings. As a component of an eMERGE genomics pilot study, we aimed to determine parents' intent and actual practice of sharing their child's CYP2D6 research result, as it pertained to codeine. Enrolled parents had previously given permission for their child to participate in a genomics study or for their child's leftover clinical samples to be stored in a voluntary bio bank for future genomic research. For our study, children's CYP2D6 research results were returned by an advanced practice nurse to parents by phone. Immediately after results, parents completed a telephone survey that included items about their intent to share their child's result. The questions were asked again at least 3 months after learning results. To date, 136 parents enrolled, *CYP2D6* testing has been performed on 102 banked children's samples and results returned to 62 parents. Initial surveys were completed by 60 mothers, 1 father, of whom 92% reported their race as White, 3% as Black, 5% declined to report. The majority of parents (80%) strongly agreed their child's doctor could use the CYP2D6 result when providing care to their child. Parents reported they were extremely likely to share the result with their child's doctor (66%), pharmacist (23%) and their child (46%) within the year. Of the 31 parents who have completed 3 month follow up telephone surveys, 19 (61%) reported they were extremely likely to share the result with their child's doctor in the initial survey, and 42% of these had actually shared the result by the time of the 3 month call. Two parents had shared their child's results with a pharmacist and only one of these had reported being extremely likely to do so at the initial survey. Although 39% of the parents reached for follow-up had initially reported they were extremely likely to share the result with their child within a year, only 1 had done so at the 3 month follow up. Parents' intent to share and perceived utility of *CYP2D6* testing is high. Enrollment and data collection from 200 participants is planned to determine factors that influence parents' transition from intention to actually sharing their children's CYP2D6 results.

717S

Software for the clinical implementation of pharmacogenomic testing. *R. Ammar[1], I. Cohn[2], T.A. Paton[2], S. Bacopulos[2], J. Vlasblom[2], M. Fiume[3], C.R. Marshall[2], P. Ray[2], R. Cohn[2], S. Ito[2], M. Brudno[1,2], G.D. Bader[1,3].* 1) Donnelly Centre for Cellular and Biomolecular Research, Toronto, ON, Canada; 2) The Hospital for Sick Children, Toronto, ON, Canada; 3) Department of Computer Science, University of Toronto, Toronto, ON, Canada.

In a medical context, a genome can be used to tailor treatment strategies to deliver more precise healthcare with the promise of improved diagnoses and responses to medical intervention. Pharmacogenomics (PGx), the study of how genes influence individual response to drug therapies, has a major direct impact on choice of drug dosing in many medical contexts. We present an open-source software tool, the *Pharmacogenomics app* for *MedSavant*, an open-source system for interpretation of clinical genomes. The app interprets pharmacogenomic variants with medically actionable output based on the guidelines established by the Clinical Pharmacogenetics Implementation Consortium (CPIC). All pharmacogenomic variants are computationally phased using an HMM-based model (BEAGLE) trained on a reference panel from the 1000 Genomes project. Phasing of haplotypes is critical to prevent misinterpretation of mutations. Haplotypes are constructed without imputation to ensure that the diagnostic results are based entirely upon observed genotypes. Patient diplotypes are reported using standard * nomenclature (i.e. *CYP2C19 *1/*17*) along with metabolizer class (*Ultrarapid*, *Extensive*, *Intermediate*, *Poor*), based on haplotype activity scores from the CPIC guidelines. If non-CPIC, rare (allele frequency < 5%) coding variants are discovered, they are reported independently alongside the patient diplotype. Our tool is agnostic to sequencing technologies, and can be used with data generated by targeted PCR-based methods (e.g. Sequenom), whole-exome sequencing (WES) and whole genome sequencing (WGS). The tool can be used with WES or WGS data regardless of why the data was generated. In the event of missing data, our tool will display possible and similar haplotypes. The *Pharmacogenomics app* is currently being piloted as part of a clinical diagnostic exam at The Hospital for Sick Children in Toronto for the evaluation of PGx in paediatric clinical practice.

718M

Implementation of pharmacogenomics into clinical practice: Mayo Clinic experience. *P.J. Caraballo, S. Bielinski, J. Pathak, R. Weinshilboum, G. Farrugia, C. Chute, J. Black, M. Elliott, C. Schultz, C. Rohrer-Vitek, D. Gabrielson, D. Blair, M. Parkulo, I. Kullo.* Mayo Clinic, Rochester, MN.

Background: Pharmacogenomics (PGx) is increasingly gaining acceptance as a guide to prescribing. However, PGx guidelines are complex and human cognition alone is not sufficient to implement them in already busy clinical workflows. Methods: To implement PGx in the clinical setting we developed clinical decision support (CDS) using synchronous and asynchronous rules integrated in two commercially available electronic medical record (EMR) systems supported by clinically driven governance and online education. We leveraged the Mayo Clinic Clinical Decision Support Program which uses multiple applications of the EMR including problem list, allergy module, computerized physician order entry, messaging system, and expert rule system. Advanced interactive functionality and actionable alerts enable busy clinicians to better understand the alerts. Embedded in the alerts and inboxes are web-links to specific online PGx education material including frequently asked questions. We also established a PGx Task Force to seek consensus among knowledge experts and clinical practice groups affected by the implementation of reactive and preemptive PGx. Results: Ten rules (*HLA-B*57:01*-abacavir, *HLA-B*15:02*-carbamazepine, TPMT-thiopurines, *CYP2D6*-codeine/ -tramadol/ -tamoxifen, *SLCO1B1*-simvastatin, *CYP2C19*-clopidogrel, *CYP2C9/VKORC1*-warfarin and *IL28B*-PEG-IFN alfa) have been approved by the PGx Task Force and seven are currently live. One rule (*IL28B*-PEG-IFN alfa) was abandoned as clinical experts felt it was no longer relevant given the availability of new drugs. Two rules (clopidogrel and warfarin) have not yet been implemented as they encountered clinical resistance in view of controversies and uncertainties in clinical guidelines from medical societies. The acquisition of actionable laboratory results has been the main technical challenge, as quality assurance processes require continuous monitoring to ensure appropriate performance. The activation rate of these rules has ranged from 6/year (carbamazepine) to 180/month (thiopurines). In the first quarter of 2014, approximately 390 unique providers and 675 unique patients have been involved in the activation of these rules. Conclusion: Our comprehensive implementation model provides insights into the challenges of incorporating PGx data preemptively in an academic medical center. A coordinated and multidisciplinary effort is needed to facilitate clinical adoption and technical feasibility of EMR-driven PGx-guided therapy.

719S

The Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging: Characterization of Clinically-Actionable Pharmacogenetic Alleles in over 100,000 Patients with Biobank-linked Electronic Medical Records. *N. Gonzaludo[1], T.J. Hoffmann[2], P.Y. Kwok[2], N. Risch[2,3], C. Schaefer[3].* 1) Bioengineering & Therapeutic Sciences, University of California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 3) Kaiser Permanente Northern California Division of Research, Oakland, CA.

Pharmacogenetic information can be extremely useful in optimizing patient therapy and avoiding adverse clinical outcomes, potentially reducing the cost burden of hospitalizations and treatment of adverse drug events. However, incorporation of pharmacogenetic information into routine clinical practice faces many hurdles, one of which involves the timely availability of accurate genetic information to clinicians when it is needed at point-of-care. With electronic medical record (EMR) systems linked to biorepositories that have generated genotypic information, it is possible to preemptively genotype patients for known drug-gene pairs before these drugs are administered. This information can enable prediction of how patients will respond before they are prescribed a drug, and thus be used to tailor drug therapy for each patient. As part of the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH), members in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort have undergone genome-wide genotyping using Affymetrix Axiom technology. Leveraging genotype data for 103,006 individuals, we phased and imputed a panel of 41 SNPs, which were translated to star allele nomenclature (e.g. CYP2C19*2) in 7 pharmacogenes with established clinical implementation guidelines. Using star allele diplotypes for each GERA cohort member, we assigned metabolizer status phenotypes for several drugs, including clopidogrel, warfarin, and simvastatin. On average, approximately 20–30% of cohort members had actionable metabolizer phentoypes for different drugs, suggesting that a change in drug choice or dose may benefit the patient. Additionally, we performed a similar characterization study using 1000 Genome samples, and compared metabolizer status frequencies across different cohorts, ethnicities, and published frequencies associated with certain drugs. By linking GERA members' metabolizer status to prescription data in the EMR, we were also able to retrospectively assess the impact of this information on prescribing patterns. Our study demonstrates the utility of biobank-linked EMRs for deriving pharmacogenetic information, as well as the potential impact of this information for personalized drug therapy.

720M

Evaluating the Application of Star Allele Nomenclature for Pharmacogenomics in the Era of High-Throughput Sequencing. *A. Gordon[1], Q. Zhu[2], C. Chute[2], D. Nickerson[1], R. Freimuth[2].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Health Sciences Research, Mayo Clinic, Rochester, MN.

As costs for sequencing patient samples within a clinically-relevant timeframe continue to decrease, more and more healthcare institutions are implementing clinical pharmacogenomics tests. Much of the decision support associated with these tests assign a diplotype to each test result, translating that result into a predicted phenotype, and recommending a certain course of action to the provider based on a predicted outcome. The diplotypes are often expressed as a pair of 'Star Alleles,' a nomenclature system that was originally developed to describe variation within the Cytochrome P450 genes and has since spread to many other genes of pharmacogenetic interest. Although this nomenclature system has been useful for pharmacogenomic implementation when individual sites were interrogated, its ability to represent high-throughput sequencing results has not been systematically evaluated. In particular, the ability to distinguish alleles defined by a single variant from those defined by multiple variants and the ability to handle rare variation must be supported. In order to quantitatively analyze these phenomena, we developed an algorithm to assess the strengths and weaknesses of the Star Allele system. The algorithm attempts to assign star allele-based diplotypes using high-throughput sequencing data for pharmacogenomically-relevant genes. We applied this algorithm to phased exome sequence data from 6503 individuals of European-American or African-American ancestry drawn from the NHLBI Exome Sequencing Project. Preliminary analysis of results from CYP2C9 indicate that 25.7% of African-Americans and 14.6% of European-Americans in our dataset carry at least one allele that cannot be named using the canonical star allele definitions, highlighting the need to quantify and understand these nomenclature errors. To that end, we are expanding our analysis to 4 other pharmacogenes of clinical importance that also use this nomenclature system: CYP2C19, CYP3A5, SLCO1B1, and TPMT. We are cataloging the type and frequency of errors encountered, including those due to rare variants, potential phasing errors, and partial or mixed haplotypes. We believe that assessing why and how often such errors occur will help inform a discussion regarding the use of nomenclature to support the clinical implementation of pharmacogenomics in the era of high-throughput sequencing.

721S

A sequence-based pharmacogenomic (PGx) panel: determining CYP2D6 sequence variants and copy number variation. *A.E. Kwitek[1,2], D.L. Kolbe[2], R. Zahr[5], C.A. Campbell[2], K.L. Knudtson[2], K. Popp[5], A.M. Mansilla[2], S.O. Mason[2], E. Snir[2], C.J. Nishimura[2], S.D. Rose[5], S. Gunstream[5], R.J. Smith[2,3,4].* 1) Pharmacology, University of Iowa, Iowa City, IA; 2) Iowa Institute of Human Genetics, University of Iowa, Iowa City, IA; 3) Pediatrics, University of Iowa, Iowa City, IA; 4) Internal Medicine, University of Iowa, Iowa City, IA; 5) Integrated DNA Technologies, Coralville, IA.

Because genetic variation is a major contributor to Adverse Drug Reactions (ADRs), the FDA has incorporated pharmacogenomic (PGx) information for over 100 drugs and recommends genetic testing for nearly 10% of all drugs. Sequence-based pharmacogenetic tests offer tremendous opportunities to transition PGx testing to standard-of-care. However, many sequence-based approaches have difficulties in accurately detecting copy number variants (CNVs) in CYP2D6, a key gene associated with the metabolism of 25% of all drugs, including opioids. We developed a comprehensive PGx test that combines targeted genome enrichment and massive parallel sequencing (TGE+MPS) that can accurately determine CYP2D6 genotypes including SNVs and CNVs. A custom designed xGen® Lockdown® Probe panel (Integrated DNA Technologies) comprised of non-overlapping 120-mer, 5' biotinylated oligonucleotides was used in a 4 hour hybridization reaction to capture over 300 genes and non-coding SNPs (590 kb) associated with drug Absorption, Distribution, Metabolism, and Excretion (ADME). As a pilot study, a validation cohort consisting of 24 CEPH individuals with known CYP2D6 genotype (SNV and CNV) was sequenced in pools of 6 or 8 on the Illumina MiSeq® and subject to both variant calling and copy number analysis. Of the targeted regions, 574 kb were successfully captured. There was over 99.8% coverage at 30x in pools of 6 samples. Variant calling identified, on average, 3000 high-quality variants, and CYP2D6 genotypes were concordant with those determined by multiple methods. CNV analysis was performed on 7 PGx-related genes previously determined to have copy number variation, including known CNVs in CYP2D6. We were able to detect homozygous and heterozygous deletions, duplications, as well as gene-gene and gene-pseudogene fusions involving 10 genes on the panel. The CYP2D6 copy number results were validated by TaqMan® assay and correspond to publically available data. We anticipate the IIHG-PGx panel will provide a cost-effective and rapid test to be used for pre-emptive testing for drug response. PGx test results can assist the clinician in determining the most effective drugs while minimizing ADRs. The utility of this panel spans both clinical testing as well as research studies to determine novel single nucleotide and copy number variants and associations with drug ADME. (*equal contribution).

722M

Prevalence of CYP2C19 * 2 polymorphism in the population with a clopidogrel prescription tertiary Clinic. *J. Martinez, I. briceno, j. Garcia, A. Tellez, P. Galvez, D. Uricoechea.* Bioscience, Univ de La Sabana, Chia, Colombia.

Clopidogrel is an antiplatelet agent used concomitantly with aspirin or as monotherapy in secondary prevention of cardiovascular complications. However, between 5 and 40% of patients develop resistance identified by persistent high platelet reactivity (1). Clopidogrel is metabolized by the hepatic cytochrome P450 (CYP 2C19 enzyme). CYP 2C19 exhibits genetic polymorphisms responsable for the presence of poor metabolizers, intermediate metabolizers and extensive metabolizers. In recent years, research has focused on the CYP2C19 enzyme encoded on chromosome 10. Its polymorphisms may reduce the formation of the active metabolite of clopidogrel and reduce platelet effect. The *2 allele carriers have increased risk of cardiovascular death, myocardial infarction and stent thrombosis compared to non-carriers (2). The defective mutations of the enzyme and their frequencies change between different ethnic groups; however, the polymorphisms of the CYP2C19 gene has not been studied in colombian population. The objective of this study is to determine the prevalence of the polymorphism CYP2C19*2 in a population treated with clopidogrel in a tertiary clinic. Follow up on this population, in order to associate the presence of genetic factors with clinical outcomes will be a second phase of this study. If confirming a significant association, this will contribute significantly so that cardiologists may determine an more appropriate therapeutic strategy for patients. On the other hand, it would be possible to include other antiplatelet agents whose pharmacokinetic and pharmacodynamic parameters are not affected by genetic polymorphisms in the Colombian Health Plan. General objective: Determine the prevalence of CYP2C19 * 2 polymorphism in the population treated with clopidogrel in a tertiary clinic. Specifics: 1.To evaluate the genotype and phenotype status of CYP2C19 in colombian population, in order to contribute to the use of appropriate strategies of drug therapy for this population. 2.To determine the frequency of the polymorphisms associated to variability in response to clopidogrel. 3.To identify the phenotype of patients according to the polymorphisms found. 4.Determining the population at high risk of thrombotic events and cardiovascular death according to genotype / phenotype. 5.To calculate the frequency of demographic variables and non-genetic factors which have been associated with variability in response to clopidogrel in the study population.

723S

Clinician Views about Implementation of Pharmacogenomics into Practice. *J.F. Peterson[1], J.R. Field[2], Y. Shi[3], J.S. Schildcrout[3], J.C. Denny[1], T. McGregor[4], S.L. VanDriest[4], J.M. Pulley[2], I. Lubin[5], M. Laposata[7], D.M. Roden[6], E.W. Clayton[8,4].* 1) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 2) Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN; 4) Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN; 5) Genetics, Centers for Disease Control and Prevention, Atlanta, GA; 6) Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 7) Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN; 8) Center for Biomedical Ethics and Society, Vanderbilt University Medical Center, Nashville, TN.

Attitudes of clinicians participating in pharmacogenomics implementation projects are not well-described but are likely key to the success of translational programs. Using a 54-item online instrument, we surveyed physicians and nurse practitioners in 2013 taking part in a large pharmacogenomics program within a tertiary care academic medical center. Domains surveyed included attitudes towards ordering pharmacogenomics testing, prior pharmacogenomics education, major factors influencing use of the results, expectations of efficacy of genotype-tailored therapy, and responsibility for applying significant results from a multiplexed genetic panel to clinical practice. Of 121 clinicians solicited, 80 (66%) physicians and nurse practitioners gave a complete survey response. Virtually all respondents (99%) agreed that pharmacogenomic variants influence patients' response to drug therapy, and a majority agreed with statements regarding the clinical utility of *CYP2C19* variants to tailor antiplatelet therapy following percutaneous coronary interventions (80%) and *VKORC1* and *CYP2C9* variants to tailor warfarin dosing (86%). Strength of evidence for drug-genome interactions (DGIs) was cited by 88% of respondents as the most important factor influencing the ordering of pharmacogenomic testing; systematic reviews and national standards-setting organizations were the most frequently trusted sources for evidence-based guidance. The majority (92%) of respondents were in favor of immediate notification when a clinically significant drug-genome interaction was present and relevant to the patient's therapeutic course, and most felt specialists were in the best position to apply a pharmacogenomic result to a new prescription (74% agreement). However, survey respondents were divided on whether the provider who originally ordered the test was responsible for acting on a result when a prescription change was indicated (51% agreement). When presented with a scenario where the prescription change was affected by a genetic testing performed 6 months earlier, a majority of respondents (67%) favored assigning responsibility to the implementation program to find and contact the patient's most current providers. Among clinicians practicing within a pharmacogenomics implementation program, genotype results were valued for tailoring prescriptions, but respondents were divided on who is clinically responsible for pharmacogenomic results from a multiplexed panel.

724M

Role of APOE4 genotypes in predicting cardiometabolic outcomes in individuals with metformin and metformin-sulfonylurea combination therapy. *G. Priamvada, B.R. Sapkota, A. Subramanian, P.R. Blackett, D.K. Sanghera.* Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Genetic polymorphisms in apolipoprotein E (APOE) locus are known to have a significant impact on various inflammatory and metabolic diseases. Apoε4 isoform is a well-known marker associated with increased risk of coronary artery disease (CAD) and late-onset of Alzheimer's disease. Recent evidence suggests improved clinical outcomes with metformin in patients with type 2 diabetes (T2D) and heart failure. However, the role of anti-diabetic medications in response to APOE genotypes is less understood. The objectives of this investigations were: 1) to evaluate the distribution of APOE polymorphisms in a large diabetic case-control cohort of the Asian Indian Diabetic Heart Study, 2) to evaluate the impact of APOE polymorphisms on quantitative risk factors of T2D and CAD, and 3) to examine the role of APOE genotypes in response to anti-diabetic therapy. We assessed the role of APOE genotypes with disease and treatment outcome in a total of 3,569 individuals (2,289 T2D cases and 1,280 controls). No significant difference in the distribution of APOε4 was observed among T2D cases and controls. However, APOε4 carriers had higher fasting glucose (p=0.021), higher diastolic blood pressure, higher LDL cholesterol and lower HDL cholesterol (p=0.031) compared to non-APOε4 carriers. Further stratification of data from diabetic patients by APOE genotypes and anti-diabetic treatments revealed a significant decrease in fasting glucose (p<0.0001), 2 hours glucose (p<0.0001), systolic blood pressure (p=0.007), diastolic blood pressure, and LDL cholesterol (p<0.0001) among the APOε4 carriers compared to non-APOε4 carriers on metformin monotherapy. Similar improved clinical outcomes were observed in patients with metformin-sulfonylurea combination therapy (n=618). Our study suggests APOε4 to be a potential risk factor for cardiometabolic susceptibility in patients with diabetes. Our findings also report significantly improved cardiometabolic outcomes among APOε4 carriers in response to metformin and metformin-sulfonylurea combination therapy. These findings warrant confirmation in a large independent dataset. This study was supported by NIH grants -R01DK082766 (NIDDK) and NOT-HG-11-009 (NHGRI), and VPR Bridge Grant (OUHSC).

725S

Genome Liberty: Direct-to-Consumer Pharmacogenetics. *J.A. Rosenfeld[1,2], C.E. Mason[3].* 1) Department of Medicine, Rutgers Medical School, Newark, NJ; 2) American Museum of Natural History, New York, NY; 3) Department of Department of Physiology and Biophysics, Cornell University, New York, NY.

For many medications, there are genetic markers whose genotype will aid in proper dosing. Certain genotypes will rule out the use of a particular medication, while other genotypes will indicate the use of either a higher or a lower dose. There are a wide range of medications that can be tested in this way including Plavix, codeine, Coumadin and estrogen. While the links between the cytochrome p450 genes and proper dosing of medications are widely known, their clinical use has been extremely limited. The vast majority of doctors refrain from performing this or other genetic tests due to their limited knowledge of genetics and their reticence to order a test they do not completely understand. This extreme restraint by doctors has prevented the widespread adoption of pharmacogentics testing. We have therefore developed a rapid, and low-cost direct-to-consumer test that will empower individuals to help their physicians make proper prescribing decisions. Our test will look at all of the clinically actionable pharmacogentic markers and produce both a physician-friendly and a patient-friendly report that will allow individuals to know which medications they should avoid, and which they should use at a dose different than that recommended for the general public.

726M

CHRNA4 rs1044396 is associated with smoking cessation with varenicline therapy. *P.C.J.L. Santos[1], J.R. Santos[1], P.R.X. Tomaz[1], J.S. Issa[2], T.O. Abe[2], J.E. Krieger[1], A.C. Pereira[1].* 1) Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of Sao Paulo Medical School, Sao Paulo, Brazil; 2) Smoker Assistance Program (PAF), Heart Institute, So Paulo, SP.

Background: Smoking is one of the biggest risk factors for developing cardiovascular disease, and is considered the most preventable cause of premature death worldwide. It is estimated that there are currently over 1.3 billion smokers globally and although there are several therapies for smoking cessation, the most smokers find quitting difficult. Aim: to evaluate the association of CHRNA4 polymorphisms with the success rate of smoking cessation therapy and with the Fagerstrom Test for Nicotine Dependence (FTND) in patients treated with varenicline. Methods: Four hundred eighty-three smoker patients were evaluated by FTND questionary and treated with varenicline and/or bupropion. They were followed for 12 weeks. CHRNA4 rs1044396 and rs2236196 polymorphisms were genotyped by melting curve analysis. Results: We observe that patients with rs1044396 CC genotype had lower frequency of success rate (29.5%) compared with patients with rs1044396 CT or TT genotypes (50.9%), during varenicline therapy (p= 0.007; n=167). We also observed that CHRNA4 rs1044396 was associated with smoking cessation success, even in multivariate model (OR: 2.28; 95% CI: 1.03–5.06). However, we did not observe association of CHRNA4 rs1044396 and rs2236196 genotypes with FTND and with bupropion therapy. Conclusion: Our findings indicate that CHRNA4 rs1044396 might be a pharmacogenetic marker for smoking cessation with varenicline therapy since this gene encodes a subunit of the nicotinic acetylcholine receptors more abundant and higher activity.

727S

Early drug responses that are followed by an acquired drug resistance in non-small cell lung cancer cells exposed to gefitinib. *M. Takahashi[1,2], M. Fukuoka[1], H. Hohjoh[1].* 1) Department of Molecular Pharmacology, National Institute of Neuroscience, NCNP, Kodaira, Tokyo, Japan; 2) Division of RNA Medical Science, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan.

Acquired drug resistance is a major problem in medication, particularly in cancer treatment. Early drug responses followed by a drug resistance may provide us with a clue for understanding the mechanism of acquired drug resistance and for developing a therapeutic strategy to overcome such a resistance. We attempted to examine early drug responses of cancer cells that were treated with gefitinib [an epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor]. Non-small cell lung cancer (NSCLC) PC-9 cells that bear an oncogenic mutant EGFR and have the potential for acquiring a resistance to gefitinib were used in this study, and we established gefitinib-resistant PC-9 cells (PC-9/GR) by a long-term exposure of a low-dose gefitinib. Comprehensive gene expression analyses against naïve PC-9 cells and PC-9/GR cells were carried out, and the results indicated that fibroblast growth factor 2 (FGF-2) and FGF receptor 1 (FGFR1) were significantly increased in PC-9/GR cells. In addition, while the FGF2 and FGFR1 protein levels were increased in PC-9/GR cells, oncogenic EGFR dimer appeared to be decreased. The findings suggested the possibility that a survival signaling system might have been changed from an EGFR dependent signaling to a FGFR dependent signaling in PC-9/GR cells. To see early responses of naïve PC-9 cells against gefitinib further, we investigated FGFR1 and FGF-2 in naïve PC-9 cells exposed to gefitinib. The analysis revealed that FGFR1 was increased in an exposure time- and gefitinib dose-dependent manner. As for FGF-2, although its apparent gene expression level appeared to remain unchanged, the FGF-2 protein in culture medium was markedly increased by gefitinib treatment in a dose-dependent manner: this may mean FGF-2 that was released from dead cells. Additionally, naïve PC-9 cells that were treated with exogenous FGF-2 appeared to decrease in sensitivity to gefitinib. Taken together, the results suggested that naïve PC-9 cells exposed to gefitinib might be survivable by an up-regulated FGFR1 and by an increase in FGF-2 that was released from neighboring dead cells, and that PC-9/GR cells, after acquisition of gefitinib resistance, would establish a FGF autocrine signaling system. In addition to the above findings, we would like to further discuss about the possibility that thyroid hormone-related factors might contribute to up-regulation of FGFR1 in gefitinib-treated cells.

728M

Modeling the pharmacological response to advance the research in pharmacogenetics. *J. Bertrand[1], M. de Iorio[2], D.J. Balding[1].* 1) Genetics Institute, University College London, LONDON, --., United Kingdom; 2) Statistical Science Dept, University College London, LONDON, --., United Kingdom.

Context: Pharmacogenetics studies the genetic part of interindividual variability in drug response. Its main challenges are: i. the phenotype results from physiological processes (e.g. drug elimination or distribution) not directly observed, ii. genetic variation markers lead to uneven and small sample size subgroups and iii. one has to face the dimensionality curse when dealing with high throughput genetic screening. Nonlinear mixed effect models (NLMEM) handle challenge i. and ii. by capturing the dynamical nature of the phenotype and fitting patients with sparser profiles while borrowing information from more informative ones. Objective: In this work we propose to address challenge iii. with an integrated approach, that combines the strengths of NLMEM with penalised regression, and compare its performance to a classical stepwise procedure through a simulation study. Methods: We simulated pharmacokinetic profiles in association with genotypes for the 1200 single nucleotide polymorphisms (SNP) of the DMET Chip (Daly, 2007). Several scenarios were studied in terms of design: a typical phase II study or a combined design of such a study with additional through concentrations collected in clinical routine and in terms of genetic association: causal variants affecting one or multiple model parameters. Results: Both approaches showed low power estimates on the phase II study design. Adding patients with only a through concentration led to much higher power estimates for both approaches, with the stepwise procedure detecting more false positives. Further, the integrated approach was more powerful at detecting SNP effect on multiple model parameters. Conclusion: Classical phase II study designs are critically under powered to detect realistic PK parameter-SNP associations. The integrated approach showed better performance on a design combining full-profile and sparse sampling which enables the necessary increased in sample size.

729S

"A Tale Of Genetic Variation In The Human SLC22A1 Gene Encoding OCT 1 Among Type 2 Diabetes Mellitus Population Groups Of West Bengal, India". *D. Sur.* C.U., Kolkata, Westbengal, India.

The organic cation transporter 1, OCT 1 (also called SLC22A1-Solute Carrier Family 22 member 1), appears to play a role in the efficacy and disposition of variety of organic cation including drugs. Genetic polymorphisms in the drug transporter have been increasingly recognized as a possible source of variation in drug disposition and response. Genetic variants in OCT1 have been identified largely in European, Asian (Japanese, Chinese and Korean) populations. Interestingly, eight genetic variations were found in the human SLC22A1 gene, which encodes OCT 1, from 50 type 2 diabetes mellitus individuals (T2DM), in West Bengal population. The purpose of this study was to investigate genetic variants of OCT1 in West Bengal populations. The study detected the three previously reported nonsynonymous variations, 480 G>C (L160F); 1022 C>T (P341L); 1222 A>G (M408V) and one synonymous variations 156 T>C (S52S) at a minor allele frequencies (MAF) of 0.63, 0.20, 0.43 and 0.27 respectively. This study also found four previously reported intronic variations: IVS1-43(T>G), IVS2 -99(C>T), IVS5 -61(G>A), IVS9 +43(C>T) with minor allele frequencies of 0.20, 0.17, 0.18, and 0.37 respectively. This is the first report of SLC22A1 variations among Indian, especially West Bengal's type 2 diabetes mellitus patients. The present results would be useful for haplotype analysis and pharmacogenetic studies on OCT1.

730M

A role for B cells in Progressive Multifocal Leukoencephelopathy revealed by comprehensive genomic analysis. *J. Carulli[1], L. Fugger[2], A. Day-Williams[1], A. Haghikia[2,4], C. Sun[1], N. Allaire[1], H. Li[1], C. Dendrou[2], T. Plavina[1], H. McLaughlin[1], P. Cullen[1], M. Liu[1], A. Thai[1], R. Gold[4], R. Martin[3], I. Jelcic[3], T. Compton[1], T. Harris[1].* 1) Translational Medicine, BiogenIdec, Cambridge, MA; 2) The Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DS; 3) Department of Neurology, St Josef-Hospital, Ruhr University Bochum, Bochum, Germany; 4) Department of Neurology University Hospital Zurich Frauenklinikstrasse 26 CH-8091 Zurich.

Progressive multifocal leukoencephalopathy (PML) is a rare and potentially lethal infectious disease caused by JC virus in individuals with compromised immune systems or who take specific immune modulating drugs. JC virus (JCV) has a seroprevalance of 60% in adult populations worldwide, but PML is very rare. Prior to the development of anti-HIV drugs, roughly 5% of AIDS patients developed PML. More recently, PML has been observed at a rate of approximately 1/1000 or lower in patients taking immune-modulating drugs, including natalizumab (anti-VLA4) for multiple sclerosis. To identify novel biomarkers of PML risk, we performed whole exome sequencing, whole genome sequencing, and genome wide transcription profiling of multiple sclerosis patients who developed PML while taking natalizumab with appropriately matched MS patients who did not develop PML while taking natalizumab for two years or more. We also performed whole genome sequencing of two subjects who developed PML while not taking any immune modulating drugs. Single SNV association testing and gene-based collapsing tests of nearly 100 natalizumab PML cases and 100 natalizumab controls did not reveal any significant associations with PML. One of the two "spontaneous" PML subjects had a complete deletion of the Dock8 gene on chromosome 9p. Dock8 deficiency is a primary immune deficiency with pleiotropic effects, including B cell migration defects. Genome-wide transcript profiling of 18 natalizumab PML subjects and 150 natalizumab controls revealed that several transcripts, including IgHM, FcRLA, FcRL3, CD72 and CD22, exhibited lower expression levels in PML compared to non-PML patients. These findings were confirmed by qPCR, and the IgM observation was confirmed at the protein level. Pathway analysis of the transcript data identified B cell pathways significantly deregulated in subjects who got PML. Taken together, the genetic data and the transcript data suggest that variation in maturation and/or migration of B cells plays a role in the development of PML and may provide biomarkers for PML risk.

731S

Association between *CYP2B6* +516 G>T polymorphism and response to first-line therapy in Brazilian HIV-1+ individuals. *T.B. Almeida, M.B. Arruda, R.M. Brindeiro, A. Tanuri, C.C. Cardoso.* Laboratrio de Virologia Molecular, Instituto de Biologia, UFRJ, Rio de Janeiro, Brazil.

According to World Health Organization, 35 million individuals are infected by HIV-1 in the world and about 9.5 million are currently undergoing highly active antiretroviral therapy (HAART). However, 10–20% of these individuals patients do not reach therapeutic success, mainly as a consequence of low adherence to the treatment and emergence of drug resistant viral strains. HAART effectiveness is also influenced by host factors that affect drug absorption, activation and metabolism. Single nucleotide polymorphisms (SNPs) in *CYP2B6* gene can directly influence the expression and function of the enzyme responsible for the metabolism of efavirenz (EFZ) and nevirapine (NVP). The aim of this study was to investigate the association between 14 candidate SNPs in *CYP2B6* gene and HAART failure in EFZ or NVP containing first line regimens. A total of 111 HIV-1+ individuals under HAART selected from the Brazilian cities of Porto Alegre and Curitiba were included in the study. From these, 73 individuals reached therapeutic success, defined by the reduction of viral loads (VL) to undetectable levels (<50 copies/mL). The remaining 38 individuals exhibited HAART failure, with detectable VL after 6 months of treatment. SNP genotyping was performed by SNaPshot® or TaqMan® assays (Life Technologies, USA). Nine SNPs were excluded from the study due to deviations from Hardy-Weinberg Equilibrium or because they were monomorfic in our study population. The frequencies of the remaining 5 SNPs were compared using logistic regression models adjusted for sex and HIV-1 subtype. Our results showed an increased risk of HAART failure (OR=2.21; 95%CI= 0.96–5.04; p=0.048) among carriers of the allele +516T (rs3745274). The same allele has been previously associated to higher EFZ and NVP plasma levels, suggesting that it might favor therapy effectiveness. However, the effect of this SNP may be influenced by additional variations in *CYP2B6* or genes coding for drug transporters. It suggests that the genetic background of each population, including linkage disequilibrium patterns, may influence data interpretation, reinforcing the importance of replicating genetic association results in different populations. To our knowledge, this was the first study to describe the association of CYP2B6 gene to HAART response in Brazilian HIV-1+ individuals. The use of pharmacogenetic markers may be crucial to define personalized regimens, which may be more effective and better tolerated by the patients.

732M

Identifying Differentially Expressed Genes Associated with Extreme Blood Pressure Response to Hydrochlorothiazide monotherapy. *A.C. Costa Sa[1-3], A. Webb[4], Y. Gong[1,2], T. Langaee[1,2], S. Turner[5], A. Chapman[6], J. Gums[1,2,7], S. Scherer[8], R.M. Cooper-DeHoff[1,2,9], W. Sadee[4], J.A. Johnson[1-3,9].* 1) Pharmacotherapy and Translational Research, University of Florida, Gainesville, FL; 2) Center for Pharmacogenomics, University of Florida, Gainesville, FL; 3) Graduate Program in Genetics and Genomics, University of Florida, Gainesville, FL; 4) Department of Pharmacology, College of Medicine, Ohio State University, Columbus, OH; 5) Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN; 6) The Renal Division, Department of Medicine, Emory University, Atlanta, GA; 7) Department of Community Health and Family Medicine, University of Florida College of Medicine, Gainesville, FL; 8) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 9) Division of Cardiovascular Medicine, Department of Medicine, University of Florida, Gainesville, FL.

Hypertension (HTN) is the most common chronic disease in the world, estimated to affect nearly 1 billion individuals worldwide. Hydrochlorothiazide (HCTZ) is one of the most commonly prescribed antihypertensive (anti-HTN) medications in the US with over 100 million prescriptions annually. Even though HCTZ is one of the preferred options for initial HTN therapy, patient's response varies widely, and less than 50% of them achieve blood pressure (BP) control. We hypothesize that interpreting the functional elements of the genome will help us in identifying candidate biomarkers associated with variability in the efficacy of HCTZ therapy. The present work aims to identify genes with differential expression levels between responders and nonresponders after HCTZ monotherapy in Caucasian hypertensive PEAR (Pharmacogenomic Evaluation of Antihypertensive Responses) participants. Uncomplicated hypertensive patients were recruited and randomized to either HCTZ or atenolol monotherapy. BP and adverse metabolic responses were assessed after 9 weeks of monotherapy. On 50 European American participants, classified as responders and nonresponders to HCTZ, total RNA was isolated from whole blood and used for transcriptomic RNA sequencing with Illumina HiSeq2000®. The generated reads were aligned to the reference genome (Homo Sapiens Hg19) with TopHat2. Cufflinks/Cuffdiff pipeline was used to calculate gene expression levels, reported as fragments per kilobase per million reads (FPKM). Paired t-test was performed to evaluate significance of the change in FPKM. Differential expression analysis revealed that 2 genes may play an important role in BP response to HCTZ: *RHOB* and *CDC42EP2* (FDR adjusted p-value < 0.05). RHOB is a small GTP binding protein, which is known to modulate Ca^{2+}sensitization of vascular smooth muscle cells (VSMCs) and is thought to act by inhibiting myosin phosphatase activity, sustaining muscle contraction. CDC42EP2 is a CDC42 Effector Protein 2. CDC42 is also a GTP binding protein, member of the Rho protein superfamily. It activates serine/threonine kinase MRCK, promoting myosin phosphorylation. Previous studies highlight the convergence of Rho and CDC42 signaling in VSMCs, corroborating our RNA-Seq findings. These results suggest that Rho/Rho-kinase and CDC42/CDC42EP2 pathways may be involved in vascular resistance hypertension and possible reflect in BP response variability associated with thiazide diuretics anti-HTN treatment.

733S

Contribution of rare protein-coding variants to anti-TNF treatment response in rheumatoid arthritis patients. *D. Diogo[1–4], J. Cui[1], R.S. Fulton[5], J.D Greenberg[6,7], D.A. Pappas[7,8], J.M. Kremer[9], A. Barton[10], M.J.H. Coenen[11], B. Franke[11], L.A. Kiemeney[12], X. Mariette[13,14], C. Richard-Miceli[13,14], H. Canhão[15,16], J.E. Fonseca[15,16], N. de Vries[17], P.P. Tak[17], J.B.A. Crusius[18], M.T. Nurmohamed[19], F. Kurreeman[20], T.W.J. Huizinga[20], Y. Okada[1–3], E.A. Stahl[1–3], L. Klareskog[21], L. Padyukov[21], E.R. Mardis[5], P.K. Gregersen[22], R.M. Plenge[1–3], S. Raychaudhuri[1–4,10].* 1) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's, Hospital, Harvard Medical School, Boston, MA; 2) Division of Genetics, Brigham and Women's, Hospital, Harvard Medical School, Boston, MA; 3) Medical and Population Genetics Program, Broad Institute, Cambridge, MA; 4) Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA; 5) The Genome Institute, Washington University School of Medicine, St. Louis, MO; 6) New York University Hospital for Joint Diseases, New York, NY; 7) Corrona, LLC, Southborough, MA; 8) Columbia University, College of Physicians and Surgeons, New York, NY; 9) The Albany Medical College and The Center for Rheumatology, Albany, NY; 10) Arthritis Research UK Epidemiology Unit, University of Manchester, Manchester Academic Health Sciences Centre, Manchester, UK; 11) Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 12) Department for Health Evidence, Radboud University Medical Center, Nijmegen, The Netherlands; 13) Université Paris-Sud, Orsay, France; 14) APHP-Hôpital Bicêtre, INSERM U1012, Le Kremlin Bicêtre, Paris, France; 15) Rheumatology Research Unit, Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Lisbon, Portugal; 16) Rheumatology Department, Santa Maria Hospital-CHLN, Lisbon, Portugal; 17) Amsterdam Rheumatology and immunology Center, Department of Clinical Immunology & Rheumatology, Academic Medical Center / University of Amsterdam, The Netherlands; 18) Laboratory of Immunogenetics, Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, The Netherlan; 19) Amsterdam Rheumatology and immunology Center, Department of Rheumatology, Reade, Amsterdam, The Netherlands; 20) Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands; 21) Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden; 22) The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, NY, USA.

There are several effective drugs that treat systemic inflammation in rheumatoid arthritis (RA), including the widely used anti-tumor necrosis factor alpha (anti-TNF) biologic therapies. However, a significant number of RA patients fail to enter remission with the drugs currently available. Despite multiple efforts, no confirmed genetic predictor of anti-TNF treatment response in RA patients has been identified from association studies of common variants. Here, we targeted 750 genes for exon-sequencing and investigated the aggregate contribution of rare protein-coding variants to anti-TNF treatment response (improvement in disease activity score) in 1,131 RA patients of European ancestry. Gene-based association tests resulted in a study-wide significant association at *TNF*, driven by two rare missense mutations in three patients predicting good treatment response (Beta = +3.04, $P=4\times10^{-5}$). Interestingly, we observed several additional genes involved in the TNF signaling pathway among the genes with the strongest signals of association ($P<0.05$). We performed a gene set enrichment analysis based on association Pvalues ranking, and observed a nominal enrichment of association at genes involved in the TNF pathway ($P_{enrichment}=0.03$, based on 1,000 phenotype permutations). Our results, if validated in independent collections, would suggest that rare protein-coding variants in genes from the TNF signaling pathway contribute to treatment response in RA patients.

734M

Assessing the clinical utility of massively parallel sequencing for pharmacogenomics research in the ClinSeq® study. *D. Ng[1], L.N. Singh[1], C.S. Hong[1], K.L. Lewis[1], J.C. Mullikin[2,3], L.G. Biesecker[1,2], NIH Intramural Sequencing Center, National Human Genome Research Institute, NIH.* 1) Medical Genomics and Metabolic Genetics Branch, MGMGB/NHGRI/NIH, Bethesda, MD; 2) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Comparative Genomics and Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

ClinSeq® is a medical sequencing project designed to study the clinical application of massively parallel sequencing (MPS) in a cohort of mostly healthy volunteers. One of the goals of personalized medicine is to optimize the efficacy of pharmacologic treatment and to minimize adverse drug reactions. Pilot projects studying the utility of pharmacogenetic (PGx) guided treatment have used Sanger sequencing or genotyping chips to identify a finite number of important PGx variants. To study the utility of MPS for identifying clinically important PGx variants, we compared the accuracy and coverage of whole exome (WES) and genome sequencing (WGS) at 1,934 variant positions interrogated by the Affymetrix Drug Metabolism and Transport (DMET) Plus chip. ClinSeq® cohort is comprised of 973 volunteers age 45–65 from the Washington DC area. Participants received baseline clinical evaluations, WES, consent for return of genetic results and iterative phenotyping to study incidental variants. Five participants had both WES and WGS. Sequence reads were aligned to NCBI assembly GRCh37 with Novoalign v3.02.00 (Novocraft technologies). Genotypes were determined with GATK v3.0-0. Coverage of 1,934 variant positions in 231 DMET genes were analyzed in 973 exomes. High quality reads were defined as GATK score =99 and read depth (RD) ≥10. A variant position had high coverage (HC) if ≥80% (n=779–973) of participants had high quality reads, low coverage (LC) if <80% (1–778) and no coverage (NC) if no one had a high quality read at that variant position. WES had the poorest coverage for intergenic (0/31 HC), promoter (3/91 HC) and 5'UTR variants (10/39 HC). Unsurprisingly, WES coverage was the highest in exonic variants (1,239/1,332 HC, 92/1,332 LC, 1/1,332 NC). WES RD can be used to identify copy number variants (CNVs) affecting drug metabolism (i.e., CYP2D6) with XHMM software. WGS had excellent coverage of coding and noncoding (average RD ≥10) DMET Plus variants ranging from 92% for 5'UTR to 100% for promoter, 3'UTR and intergenic variants. WGS has several advantages over the DMET Plus array for preemptive PGx testing. Any array or other panel test has to be updated as new PGx variants are discovered, necessitating periodic retesting. WGS captures both known and yet to be discovered coding, noncoding and CNV PGx variants in one assay. As the cost of MPS decreases and clinical interpretation improves, WGS will likely be the genomic test of choice for personalized medicine.

735S

Development of a multiplex genotyping method for CYP2C19 specialized to Korean using the single base extension technique. *D. Seo, H. Jeong, J. Seo, J. Ha, I. Hwang, J. Choi, H. Na, J. Kim.* Clinical Research Division, National Institute of Food and Drug Safety Evaluation, MFDS, Cheongju, South Korea.

Most of drugs are xenobiotics for human body and are metabolized mainly in the liver. Drug metabolism was exerted by phase I and II enzymes in the liver and highly dependant on the activity of enzymes. Cytochrome P450s(CYPs) are known to play essential roles in phase I drug metabolism in the liver. Among the CYPs, CYP2D6, CYP2C9 and CYP2C19 are known to play important roles in drug metabolism. According to US FDA's list of pharmacogenomic biomarkers in drug labeling, CYP2C19 is the most common valid biomarker involved in metabolisms of 14 different drugs and a number of genotyping kits are already available in the market after US FDA's approval. National Institute of Food and Drug Safety Evaluation(NIFDS) has put its effort to conduct pharmacogenomic studies since early 2000s for harnessing the new paradigms of drug utilization such as personalized medicine in Korea. As a part of the effort, we developed a genotyping method for CYP2C19 variants using a single base extension technique(SNaPShot) containing Korean specific minor alleles. The kit is capable of genotyping 15 variants of CYP2C19 including common variants, like *2 [A], *3 [A], *4 [G], *5A [T], *8 [C], *17 [T] and also a couples of variants specific to Korean population such as rs11568732, rs3758580, rs3758581, rs4417205 and rs4986894. The validity of the kit was evaluated by triple repetition test on 96 Korean DNA samples. The accuracy of the genotyping was also validated by sequencing of each target variants. The accuracy and precision of the kit ranged from 81.6% to 100% and from 94.8% to 100% on common CYP2C19 alleles, respectively. The accuracy and precision on Korean specific alleles ranged from 80.2% to 90.6% and from 98.3% to 100%, respectively. The overall accuracy and precision of the kit turned out as 85.8% and 99.1% respectively. The results suggested that the newly developed kit was capable of genotyping both common and Korean-specific CYP2C19 alleles with effectiveness. We are expecting that this kit provides a cost-effect mean for genotyping pharmacogenetically important CYP2C19 alleles in Korean population. Acknowledgement : This research was supported by a grant (13181MFDS705) from Ministry of Food And Drug Safety in 2013. And The Korean DNA was supplied by Center for Genome Science National Institute of Health, KCDC. Korean DNA was originated by Ansung-Ansan cohort data.

736M

NAT2 polymorphisms in a Brazilian indigenous group. *V.M. Zembrzuski[1], R.L.F. Teixeira[2], P.H. Cabello[1,2].* 1) UNIGRANRIO, Duque de Caxias, RJ, Brazil; 2) IOC-Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, RJ, Brazil.

Tuberculosis is an important disease to public health in Brazil and around the world, especially among indigenous people from Amazonia and Central Brazil. Despite treatment to be effective, adverse drug reactions such as hepatotoxicity can occur, decreasing adherence to medication therapy and may lead to drug resistance. Isoniazid (INH) is the major drug associated with drug-induced hepatotoxicity. INH is metabolized via hepatic N-acetyltransferase2 (NAT2). In subjects with low NAT2 activity there is a higher risk of developing hepatic disorder. On other hand, the high activity shows a correlation between low plasma concentrations of INH and therapeutic failure. The NAT2 protein is encoded by NAT2 gene, and the acetylator status, i.e., related to activity of this protein, is determined by single nucleotide polymorphisms (SNPs) in this gene. Variant NAT2 alleles or haplotypes possessing combinations of SNPs are segregated into clusters possessing a signature SNP either alone or in combination with others. Based in a trimodal classification, individuals homozygous for rapid NAT2 acetylator alleles are deduced as rapid acetylators, individuals homozygous for slow acetylator NAT2 alleles are deduced as slow acetylators, and individuals possessing one rapid and one slow NAT2 allele are deduced as intermediate acetylators. In this study, our objective was to determine the frequency of mutations at the polymorphic gene coding for arylamin N-acetyltransferase 2 (NAT2) and its genotypes associated with acetylation in 94 individuals of a Xavante population, Native American group of Central Brazil. Sequencing analysis showed five polymorphisms already described in other studies, and most frequent in different ethnic groups: 282 C>T, 341 T>C, 481C>T, 803 A>G and 857G>A. It was found three haplotypes: NAT2*4 (wild type), which is a functional allele, was the most frequent (62.2%); NAT2*7B and NAT2*5B accounted for 22.4 and 15.4%, respectively. 55.3% of all individuals had fast acetylation genotypes, while intermediate and slow acetylators comprised 13.9 and 30.8%, respectively. In conclusion, an understanding of genetic diversity of NAT2 has medical relevance for pharmacogenetics and epidemiological applications. Due to major differences in allele frequencies between populations, individuals from different ethnic or geographic origins may respond differently to drugs.

737S

Pharmacogenomic Analysis of the Ashkenazi Jewish Population by Whole-Genome Sequencing. *S.A. Scott[1], K.L. Ayers[2], Y. Yang[1], S. Carmi[3], K.Y. Hui[4], D. Ben-Avraham[5], N. Barzilai[5], A. Darvasi[6], K. Offit[7], S. Bressman[8], L.J. Ozelius[1], J.H. Cho[1], H. Ostrer[9], G. Atzmon[5], L.N. Clark[10], T. Lencz[11,12], I. Pe'er[3,13], B. Reva[1,2], I. Peter[1].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 3) Department of Computer Science, Columbia University, New York, NY 10027; 4) Department of Internal Medicine, Genetics & Pediatrics, Yale School of Medicine, New Haven, CT 06519; 5) Department of Genetics and Department of Medicine, Albert Einstein College of Medicine, Bronx, NY 10461; 6) Department of Genetics, The Institute of Life Sciences, The Hebrew University of Jerusalem, Givat Ram, Jerusalem, Israel, 91904; 7) Cancer Biology and Genetics Program and Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065; 8) Department of Neurology, Beth Israel Medical Center, New York, NY 10003; 9) Department of Genetics and Department of Pathology, Albert Einstein College of Medicine, Bronx, NY 10461; 10) Department of Pathology and Cell Biology and Taub Institute for Research of Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, NY 10032; 11) Center for Psychiatric Neuroscience, The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, NY 11030; 12) Department of Psychiatry, Division of Research, The Zucker Hillside Hospital Division of the North Shore-Long Island Jewish Health System, Glen Oaks, NY 11004; 13) Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032.

The Ashkenazi Jewish (AJ) population has a history of bottlenecks, expansions and geographical isolation, which has resulted in a unique genetic architecture that includes a high prevalence of founder mutations for autosomal recessive diseases and higher frequencies of some common disease risk alleles. Although recent candidate gene studies also have identified some unique pharmacogenetic variants in the AJ, no pharmacogenomic studies have been reported in this population. To identify novel variants potentially implicated in drug response variability and to characterize the pharmacogenomic profile of the AJ population, 128 whole-genomes of healthy AJ individuals sequenced to high-depth were interrogated across a panel of 320 selected autosomal pharmacogenetic genes representing ~16.9 Mb of the human genome. Of the 73,683 total variants detected, 1,228 (1.7%) were non-synonymous, and 16,449 (22.4%) were previously unreported with 5,547 (33.7%) of these observed in two or more individuals. Of the non-synonymous variants, 373 (30.4%) were novel, with 371 (99.5%) having minor allele frequencies (MAF)<5% and 158 (42.4%) predicted *in-silico* to be functional. AJ individuals had a significantly lower burden of predicted functional variants (mean=23) compared to the European (EUR; 28), African (AFR; 38), and Asian (ASN; 29) populations in the 1000 Genomes Project (p<2e-22), consistent with an isolated population. Over 6% of the AJ and EUR predicted functional variants had significant differences in MAF (p<0.0001), including several predicted to have high functional impact (e.g., *FMO2* p.S195L, *UGT2B11* p.C156R). In addition, after randomly sampling 128 individuals (100X) from the 1000 Genomes Project populations, 149 (non-singleton) non-synonymous variants were unique to the AJ, compared to the EUR (63), AFR (457), and ASN (202) populations. Analysis of clinically actionable pharmacogenetic variants defined by PharmGKB indicated that the AJ harbor most of these variants with comparable frequencies to the EUR; however, ~16% of AJ harbor unique (non-singleton) variants in these clinically relevant pharmacogenes compared to only ~3% of the EUR, the majority predicted to be functional. These data indicate that in addition to common pharmacogenomic variants found in other populations, the AJ harbor many pharmacogenomic variants that are uncommon or absent in other populations and which may play an important role in drug response variability in this unique population.

738M

CYP2D6 allele specific copy number analysis using TaqMan® SNP Genotyping Assays and digital PCR. *T. Hartshorne, S. Patel, D. Keys.* Genetic Analysis, Thermo Fisher Scientific, South San Francisco, CA.

The major drug metabolizing enzyme, CYP2D6, is encoded by a highly polymorphic gene. Over 100 star allele haplotypes are known, which can contain SNP, InDel, and copy number variants (CNVs) and which fall into 3 main functional categories (full, reduced, or none). The diploid star allele content is predictive of CYP2D6 drug metabolizer phenotype (ultrarapid, extensive, intermediate, or poor). We previously described a workflow whereby sample SNP genotype and CNV results, generated using TaqMan® SNP assays and TaqMan® Copy Number assays, respectively, can be translated to star allele diplotypes using AlleleTyper™ software. However, for samples that carry CYP2D6 duplications and are heterozygous for key SNPs, the specific allele that is duplicated cannot always be identified. A phenotype can be assigned if all 3 alleles are of the same functional category, but if alleles are from different functional categories there may be 2 possible phenotypes. To address this issue, we developed a method to detect allele-specific copy number by digital PCR using the QuantStudio™ 3D Digital PCR System. Sample DNA is loaded onto nanofluidic chips at concentrations to give one or no copies of target per dPCR. A count of reactions with and without amplification can be used for target quantification purposes. For the allele-specific dPCR application, TaqMan® SNP assays to CYP2D6 variants that are associated with specific duplicated alleles were run in dPCRs on samples of known SNP genotype and CNV status. Samples were first digested with a restriction enzyme to separate tandem duplicated CYP2D6 alleles and enable their partitioning into distinct dPCRs. Sample input amounts and thermal cycling conditions were optimized to best amplify and resolve each allele in cluster plot analysis. Reactions positive for each allele, detected by allele-specific VIC® or FAM™ dye-labeled probes, were counted and the allele ratios determined. For samples heterozygous for target SNPs, 2-copy samples gave close to 1:1 allele ratios, whereas 3-copy samples gave close to 1:2 ratios and the duplicated allele was readily identified. We have thus shown that allele-specific copy number analysis using dPCR and TaqMan® SNP assays is a simple and effective method for identifying specific duplicated alleles in heterozygous samples. This method facilitates accurate CYP2D6 allele genotyping and better prediction of drug metabolizer phenotype.

739S

Analysis of CYP1A2 gene non-coding region polymorphisms in Roma and Hungarian population samples. *B. Melegh[1,2], P. Matyas[1], B.I. Melegh[1], L. Magyari[1,2], J. Bene[1], B. Duga[1,2], Zs. Banfai[1], A. Szabo[1], R. Szalai[1,2].* 1) Department of Medical Genetics, University of Pecs, Pecs, Hungary; 2) Szentagothai Research Centre, University of Pecs, Pecs, Hungary.

Purpose: CYP1A2 enzyme contributes to biotransformation of wide variety of therapeutically important drugs, including caffeine, clopidogrel, clozapine, warfarin, procarcinogens, and some endogenous substrates. The aim of this study was to determine the pharmacogenetic profile and interethnic differences of variants of CYP1A2 gene in Roma (Gypsy) and Hungarian population.

Methods: A total of 404 Roma and 396 Hungarian healthy subject's biobanked DNA were genotyped for two non-coding variants of CYP1A2: -163C>A (*1F) and -3860G>A (*1C). Polymerase chain reaction - restriction fragment length polymorphism technique was applied.

Results: The minor allele frequency for CYP1A2*1C variant was 2.02% in Hungarians, while was not detectable in the Roma samples; the AA homozygous genotype was also not detectable. For CYP1A2*1F polymorphism we found a marked differences in AA genotype in Roma population compared with the Hungarians (31.9% vs. 49.5%, p<0.001) and in minor allele frequency (56.9% vs. 68.6%, p=0.025). The following CYP1A2 genotypes were identified in Roma and Hungarian samples, respectively: *1A/*1A (18.1% vs. 12.4%), *1A/*1F (50% vs. 36.9%), *1F/*1F (31.9% vs. 46.7%). In Hungarian population we found the *1C/*1F genotype (4.04%), but it was not present in Roma subjects.

Conclusions: Analysis of distribution of CYP1A2 gene variants revealed further pharmacogenetic differences between Roma and Hungarian population samples. Theoretically, as a further consequence, the Hungarians have higher chance for rapid metabolism of CYP1A2 substrates, intensified procarcinogen activation, and perhaps thereby elevated risk for cancers.

740M

Evaluation of DNA extracted from up to 16 years old *post-mortem* blood FTA cards using Quantifiler Human Plus Quantification Kit. *AL. Rahikainen, JU. Palo, A. Sajantila.* Department of Forensic Medicine, Hjelt Institute, P.O.Box 40, FI-00014 University of Helsinki, Helsinki, Finland.

Introduction: The field of pharmacogenetics may significantly benefit from *post-mortem* (PM) studies which often harvest the extreme cases from the population. When combining genetic information to other autopsy data, especially from toxicological analyses, genetic outliers may be revealed. A potential source of DNA is FTA™ Gene Cards (Whatman™, GE Healthcare). In our department, blood samples from all autopsies have been preserved routinely since June 1998 on FTA cards. However, due to the PM changes it is expected that the starting condition of the samples vary prior to sampling and therefore methods used for PM DNA analysis must be compatible with degraded or inhibited DNA samples. The aim of this study was to determine the quality and quantity of DNA on the FTA cards sampled during 1998–2013 and to evaluate their usability in pharmacogenetic studies. *Materials and methods:* Four random samples from eight time points covering sampling years 1998 to 2013 resulting 32 DNA samples in total were extracted in triplicates. Four punches (2.0 mm Harris Micro-Punch™) were used in each extraction. DNA extraction was performed with Automate Express™ System (Applied Biosystems (AB)) according to manufacturer's protocol using PrepFiler Express™ Forensic DNA Extraction Kit (AB). Extracted samples were quantified and qualified by using Quantifiler HP Quantification Kit (AB) and analyzed with 7500 Real-Time PCR system (AB). *Results:* According to the quantification results all samples contained amplifiable DNA, but both quality and quantity were expectedly higher in the more recent samples. There was variation in the DNA quantity between the samples, despite equal amount of input material. Also internal variation was seen among the triplicates. All samples were free of inhibition. *Conclusions:* Blood samples on FTA cards are rather stable over 16 years although the older samples showed more degradation and lower yields compared to the recent ones. This study showed that also in the oldest samples from 1998 the mean DNA yield was 4.9 ng/μl (small target), and 3.4 ng/μl (large target) making them usable in pharmacogenetic studies, but the higher degradation level must be taken account. In Finland the full toxicological analyses are conducted from approximately 70% of the forensic autopsy cases. Therefore the DNA preserved on FTA cards combined with the results of toxicological analyses offers unique opportunity for conducting pharmacogenetic studies.

741S

Utilizing Pharmacometric Modeling in Pharmacogenetic Association Tests to Increase Study Power. *H. Zhou, Z. liu, S. Nayak.* Pfizer, Cambridge, MA.

In parmacogenetic studies, association tests are routinely performed to identify novel genetic markers predictive of treatment response. Traditionally, three approaches can be used to identify genetic markers in clinical trials that are designed to investigate multiple drug doses: (1) to only use data from the placebo arm and the most efficacious arm and test for genotype by treatment interaction in regression models; (2) to use data from all arms and perform an omnibus test for interactions between genotype and treatment with multiple degrees of freedom in regression models; (3) to use data from all arms and test for interaction between genotype and administered dose in regression models. However, all these approaches have drawbacks: ignoring other treatment arms in approach (1) might lead to decreased power due to reduced sample size and loss of information; approach (2) might not be powerful due to the extra degrees of freedom applied to the omnibus test; administered dose in approach (3) might not be the dose at the target site and therefore less correlated with clinical efficacy. In this study we propose an approach that utilizes data from all available treatment arms and incorporates biological mechanisms into genetic association tests in clinical trials. Drug-target binding or dose-response is estimated using pharmacometric modeling and interaction between genotype and estimated drug-target biding or dose-response is tested in regression models. Genetic and treatment-response data are simulated to mimic dose-ranging Phase II trials. Scenarios with a wide range of genetic effect sizes, mutation frequencies, rate constants for drug-target complex formation, and correlations between drug-target complex and efficacy are considered. Power of the proposed approach and conventional approaches to identify genetic markers for treatment response is compared using simulated data. The proposed approach has great potential to increase the power and probability of success of genetic studies to identify biomarkers at early clinical trial stages.

742S

Exome sequencing of multiplex oral clefts families detects recurrent shared rare variants in 9 genes. *E.R. Holzinger[1], Q. Li[1], M. Parker[2], J.B. Hetmanski[2], M.L. Marazita[3], E. Mangold[4], M.M. Nothen[4], J.C. Murray[5], A. Scott[2], T.H. Beatty[2], J.E. Bailey-Wilson[1].* 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes Health, Baltimore, MD., Select a Country; 2) Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; 3) Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine; Department of Human Genetics, Graduate School of Public Health; and Clinical and Translational Science Institute and Department of Psychiatry, School of Me; 4) Institute of Human Genetics, Life and Brain Center, University of Bonn, Bonn, Germany; 5) Departments of Biology, Epidemiology, and Pediatrics, University of Iowa, Iowa City.

Non-syndromic oral cleft phenotypes, including cleft lip with or without cleft palate (CL/P), are complex disorders with some known genetic effects. Previous GWAS have identified multiple loci with small effects on risk of CL/P and recently we have identified a novel, potentially damaging variant in CDH1 in one multiplex CL/P family. For this analysis, we used whole exome sequence (WES) data in families with 2 or 3 distantly related (2nd or 3rd degree relationship) affected individuals to identify genes containing within-family shared rare variants. Fifty-five families containing 114 individuals, 4 duplicate subject controls and 2 unrelated CEPH HAPMAP controls were sequenced on the Illumina Hi-Seq 2500 and processed through the GATK pipeline ... The families were from several different populations including 12 of Indian, 11 of Filipino, 19 of German, 10 of Syrian, 1 of European-American and 2 of Asian descent. Ingenuity 'Variant Analysis' was used to identify variants that were shared under a recessive model by all sequenced affected individuals in a family and where such sharing was observed in the same gene in at least two separate multiplex families and thus potentially associated with oral clefts. After filtering based on variant quality and frequency (MAF<0.05 in HapMap), we identified nine genes with variants that are homozygous in all affected individuals in at least two families. The variants were sometimes different across families, allowing for allelic heterogeneity. These rare variants are not present in either of the HapMap controls sequenced here. The genes are: ARHGEF12, CCT4, HSD3B7, MAN1B1, RREB1, SNRPC, STARD9, ZDHHC11, and ZNF835. Some of these genes have known biological functions which may be related to oral cleft. For example, variants in HSD3B7 have been associated with Hardikar syndrome, which sometimes presents with cleft lip and/or palate. Follow-up will include validation of these genes using Sanger sequencing and genotyping of these variants in other individuals in these same families.

743M

Candidate gene analysis of non-syndromic tooth agenesis in Japanese. *J. Machida[1], T. Tatematsu[2], A. Shibata[2], M. Kimura[3], S. Yamaguchi[4], Y. Abe[1], K. Yasui[1], A. Kitagawa[1], S. Makino[1], H. Miyachi[2], K. Shimozato[2], T. Nishiyama[5], Y. Tokita[6].* 1) Oral and Maxillofacial Surgery, Toyota Memorial Hospital, Toyota city, Aichi, Japan; 2) Maxillofacial Surgery, Aichi-Gakuin University, Nagoya, Japan; 3) Oral and Maxillofacial Surgery, Ogaki Municipal Hospital, Ogaki, Japan; 4) Oral and Maxillofacial Surgery, Aichi Children's Health and Medical Center, Obu, Japan; 5) Department of Public Health, Aichi Medical University, Nagakute, Japan; 6) Perinatology, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan.

Tooth agenesis is one of the most common congenital anomalies in humans and it is characterized by developmental absence of teeth. According to the number of missing permanent tooth excluding the third molars, tooth agenesis is classified into two categories as follows. Hypodontia is generally defined as agenesis with absence of less than six teeth, and oligodontia is a condition in which six or more teeth are missing. It is well known that both genes including ethnic background and environment play a role in common birth defects. For example, oral cleft is ubiquitous disorder world wide, although its incidence varies widely from highest in Asian populations to lowest in those of African descent. Previously we reported the prevalence of permanent tooth agenesis in Japanese was 7.0%, which is similar to that in other population-based studies.Recent studies revealed that mutations in WNT10A gene are presented in 28 to 56% of tooth agenesis in Caucasian population. In this study, we investigated mutation analysis to determine the contribution of WNT10A and other candidate genes such as MSX1 and PAX9 variants. In total, 38 patients (21 male (55%) and 17 female (45%)) were classified as non-syndromic and included in this study. A mean of 8.4 (range: 5 to 20) teeth were missing. The mean age of these patients was 19.4 years old. In 17 patients (44.7%), there was a positive family history (third degree or more closely related) for tooth agenesis. WNT10A mutations were identified in 16.7% of cases with hypodontia. MSX1 and PAX9 mutations were presented in 21.2%, and 36.4% of cases, respectively. Our findings on the intra oral distribution of agenesis of permanent teeth in children may help us better understand the etiology of agenesis.

744T

Exome sequencing reveals novel genetic cause of hereditary motor Neuropathy. *S. Poornima[1], B. Harika[1], S. Pradeep[2], B. Ravishankar[2], Q. Hasan[1].* 1) Department of Genetics & Molecular Medicine, Kamineni Hospitals, Hyderabad, India; 2) Department of Microbiology, Kamineni Hospitals, Hyderabad, India.

Charcot -Marie -Tooth (CMT), is a genetically and clinically heterozygous group of inherited disorders of the peripheral nervous system characterized by progressive loss of muscle tissue and touch sensation across various parts of the body. A twenty one year old male with hereditary motor neuropathy was referred by an orthopedic surgeon for genetic evaluation. The proband was an only son born to a consanguineous couple. He was full term baby born vaginally, who cried at birth and had a good APGAR score and had no development delay achieving all his milestones at appropriate age. He suffered with high fever at the age of 8 years and developed chicken pox. He subsequently developed gradual weakness of lower limbs followed by wasting of muscles both in the lower and upper limbs. Cytogenetic analysis showed a 46 XY normal karyotype. To rule out that it was not an autoimmune disorder due to the Varicella zoster virus Ig G and IgM antibodies were assessed in the serum and were found to be negative ruling out that the clinical symptoms were part of Vz complications. Elevated levels of CPK and ENMG gave the impression of hereditary motor neuropathy most likely CMT. Detailed pedigree analysis did not indicate any other family member with similar neurological symptoms. NGS analysis for whole genome was carried out from peripheral blood DNA and a panel of 586 neurological targeted genes were evaluated by in silico analysis, which included the genes of the CMT panel. A homozygous missense variation in exon 3 of VRK1 gene (chr14:97321690;G>A) which resulted in the amino acid substitution of Methionine by Valine at codon 236 (p.V236M; NM_003384) was detected. Variations in VRK1 gene were reported as disease causing, in patients with Ponto cerebellar hypoplasia type 1A but this patient doesnot conform to that clinical picture and appears to be unique.

745S

Whole Exome Sequencing Analysis of Severe, Early-Onset COPD in Extended Pedigrees. *D. Qiao[1,2], C. Lange[2], T.H. Beaty[5], J.D. Crapo[4], E.K. Silverman[1,3], M.H. Cho[1,3].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 3) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 4) National Jewish Medical and Research Center; Department of Medicine, Division of Pulmonary, Critical Care and Sleep Medicine, University of Colorado Denver, CO; 5) Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.

Genomic regions identified by genome-wide association studies explain only a small fraction of heritability to chronic obstructive pulmonary disease (COPD). Alpha-1 antitrypsin deficiency shows that rare coding variants of large effect also influence COPD susceptibility. We hypothesized that exome sequencing in families identified through a proband with severe, early-onset COPD would identify additional rare genetic determinants of large effect.

A total of 351 subjects in 49 extended pedigrees from the Boston Early-Onset COPD Study were ascertained through a proband with severe, early-onset COPD (age < 53, FEV1 < 0.4 predicted, without alpha-1 antitrypsin deficiency). These 351 subjects underwent whole exome sequencing and 347 subjects were analyzed after quality control. 61,695 of 124,288 total variants were defined to be non-synonymous, stop, or splice variants using SNPEff. In addition, exome sequencing data on 200 unrelated subjects with severe COPD and 200 normal controls from the COPDGene study were available as supportive evidence of association. We used two different analytical approaches to identify rare genetic determinants of COPD: 1) association methods for analyzing family-based sequencing data and 2) filtering approaches within individual pedigrees. Because the optimal test for family-based association is not clear, we applied famSKAT, PedGene, FBAT for rare variants, and EMMAX. For the filtering approach, we excluded variants with a minor allele frequency > 0.001 in public datasets and identified a remaining set of functional variants, predicted to be deleterious using Condel and CADD, that segregated with severe COPD within each pedigree.

From the association analysis, all methods except FBAT demonstrated inflation in the association p-values. No genes showed genome-wide significance over all the methods, but suggestive associations were found for *E2F2, MIP, ASAP3*, and *DNAAF2* using famSKAT. With the filtering-based approach, one variant in *ADAMTS1* segregated in two families and had a marginally significant result using SKAT-O test in the case-control dataset. We identified 70 genes containing variants segregated in more than one pedigree, and 5 of these genes showed evidence of association using the case-control dataset, including *BCAR3, ALCAM, EP300, PLD1*, and *POLR1A*. Further network analysis and functional studies are needed to validate the candidate genes discovered.

† Trainee Award Finalist

746M

Whole Exome Sequencing in Severe Chronic Obstructive Pulmonary Disease. *J. Xing[1,*], S. Bruse[2,*], M. Moreau[1,*], Y. Bromberg[3,*], J. Jang[2,4], N. Wang[1], M. Picchi[2], J. Klensney-Tait[5], J. Zabner[5], J. Mao[4], S. Belinsky[2], T. Nyunoya[2,4].* 1) Department of Genetics, Rutgers University, Piscataway, NJ; 2) COPD Program, Lovelace Respiratory Research Institute, Albuquerque, NM; 3) Department of Biochemistry and Microbiology, Rutgers University, Piscataway, NJ; 4) Department of Internal Medicine, University of New Mexico and New Mexico VA Health Care System, Albuquerque, NM; 5) Department of Medicine, University of Iowa, Iowa City, IA.

Chronic obstructive pulmonary disease (COPD) is characterized by an irreversible airflow limitation in response to inhalation of noxious stimuli, such as cigarette smoke. However, only 15-20% smokers manifest COPD, suggesting a role of genetic predisposition. Although genome-wide association studies have identified common genetic variants that are associated with the susceptibility to COPD, the odds ratios of these variants are much lower than those of familial form of emphysema, such as alpha 1 antitrypsin deficiency. We thus hypothesize that rare genetic variants contribute to the susceptibility to COPD. To test this hypothesis, we performed whole exome sequencing on 62 susceptible smokers with COPD (GOLD stage 3 or 4) and 30 resistant smokers with normal spirometries and without significant comorbidities. Overall we found a significantly higher mutation load in the genic regions of the susceptible smokers than the resistant smokers. However, the difference was mainly accounted for by non-coding regions and the number of loss-of-function mutations is not significantly different between the two groups. We then used the Variant Annotation Analysis Selection Tool (VAAST) to prioritize candidate disease-causing genes in susceptible smokers. We identified several hundred candidate genes that show higher prevalence of deleterious mutations in susceptible smokers. Among these candidates, we selected ~100 candidate genes with relatively high expression in human airway epithelial cells for function analysis using siRNA knockdown experiment. We identified several candidate genes that augment cigarette smoke extract-induced cytotoxicity in vitro. These potentially deleterious mutations of candidate genes may contribute to cigarette smoke-induced cytotoxicity, and potentially COPD.

747T

Autism spectrum disorders and dystrophinopathy in three non-identical twins. *D.P. Moreira[1], M. Lazar[1], K.M. Rocha[1], M. Aguena[1], G. Yamamoto[1], G.S. Kobayashi[1], M.S. Naslavsky[1], M.L. Lebrao[2], Y. Duarte[2], R.C.M. Pavanello[1], M. Zatz[1], M.R. Passos-Bueno[1].* 1) HUG-CELL, Dept of Genetics and Evolutionary Biology, Institute of Biosciences - University of Sao Paulo, São Paulo, SP., Brazil; 2) School of Public Health - University of Sao Paulo, Sao Paulo, SP, Brazil.

Autism spectrum disorders (ASD) are a genetically complex group of neurodevelopmental disorders that have been reported either as an isolated condition or in association with some syndromes. Among them, it has been shown that about 4% of patients with Duchenne/Becker muscular dystrophy (DMD/BMD) have ASD. Here we report three ASD-affected non-identical twins, of which two were also diagnosed with DMD. The ASD diagnosis was made according the DMS-IV-TR criteria. The two ASD-DMD patients harbor a 22-base-pair deletion involving exon 2 (c.64_69del [p.Ser22-Lys23nfs]) in the dystrophin gene, which was maternally inherited. Exome sequencing was performed in the two ASD-DMD patients in order to verify if they harbor other pathogenic mutations that could contribute to the ASD phenotype. To filter for the possible pathogenic variants, we considered only stop codon and frameshift mutations in heterozygosis or homozygosis, with a minor allele frequency (MAF) ≤0.01 in the Exome Variant Server (http://evs.gs.washington.edu/EVS) and 1000 Genomes (http://www.1000genomes.org) databases, and present in ≤5% of a cohort of 136 individuals whose exomes were sequenced at the same laboratory. This filtering led to 10 potential pathogenic mutations. Out of these variants, we highlight two mutations which, in addition to the dystrophin mutation, might contribute to the ASD phenotype: a heterozygous stop codon mutation in *OPALIN* and a frameshift mutation in *DPYSL4*, both of which have not been found in 600 Brazilian controls. *OPALIN* codes for a transmembrane glycoprotein present in oligodendrocytes and is related to myelination. *DPYSL4* plays regulatory roles in neuronal differentiation and death and in neurite outgrowth. Dystrophin is a cytoskeleton protein essential for neuron survival, and has already been associated with neurodevepmental disorders. However, the role of *OPALIN* and *DPYSL4* in these disorders has not been established. These preliminary findings suggest that the ASD phenotype in these boys might depend on at least two hits. We are currently validating these findings in the ASD non-DMD brother as well as in their parents. As the three sibs are twins, we cannot rule out the action of environmental factors contributing to ASD.

748S

Rare variants in high-risk pancreatic cancer susceptibility genes may increase risk for pancreatic cancer in some patients with and without *CDKN2A* mutations. *A.M. Goldstein[1], P. Ghiorzo[2], N. Gruis[3], M. Rotunno[1], M. Malasky[1,4], J. He[1,4], X. Zhang[1,4], L. Pastorino[2], G. Bianchi-Scarra[2], W. Bruno[2], L. Burdett[1,4], B. Hicks[1,4], K. Jones[1,4], S. Bass[1,4], S.J. Chanock[1,4], J. Boland[1,4], A. Hutchinson[1,4], M. Yeager[1,4], X.R. Yang[1], M.A. Tucker[1], Genoa Pancreatic Cancer Study Group.* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA; 2) University of Genoa, Genoa, Italy; 3) Department of Dermatology, Leiden University Medical Center, Leiden, Netherlands; 4) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA.

Germline mutations in *CDKN2A*, the major known high-risk melanoma susceptibility gene, have been described in about 20% of familial melanoma kindreds. Several variables have been shown to be associated with an increased frequency of *CDKN2A* mutations, most notably pancreatic cancer (PC) in a family. The distribution of PC differs according to the specific *CDKN2A* mutation suggesting the importance of other factors in this association. However, the precise relationship between *CDKN2A*, melanoma and PC remains unknown. Only a small subset of individuals with *CDKN2A* mutations develops PC, thus, next-generation sequencing (NGS) may present a critical strategy for uncovering genes related to PC development in these families. Since PC patients with *CDKN2A* mutations have not been systematically assessed for mutations in other known high-risk PC genes, we conducted NGS to examine the known high-risk PC and/or hereditary pancreatitis susceptibility genes (*ATM, BRCA1, BRCA2, CDKN2A, PALB2, PRSS1, STK11, EPCAM, TP53, MLH1, MSH2, MSH6, PMS2, APC, VHL, XRCC2, MEN1, PALLD, FANCC, CTRC, CPA1, CFTR, CASR, SPINK1*). DNA was available for NGS on 29 PC patients from 21 American, Dutch, and Italian families with *CDKN2A* mutations and 12 Italian PC patients without *CDKN2A* mutations who had a personal or family history of melanoma. We interrogated exome data for rare deleterious (co-segregating) mutations in the 24 PC-related genes. Deleterious variants were defined as loss-of function (frameshift, nonsense, indel) variants or substitutions predicted to be damaging by most *in silico* algorithms examined. Rare was defined as frequency <0.01 in publicly available exome databases. As quality control, we confirmed all *CDKN2A* mutations in the 29 *CDKN2A*+ PC patients plus a known *BRCA2* frameshift mutation in a *CDKN2A*- PC patient. Rare, potentially deleterious variants were found in *ATM* (n=3), *EPCAM* (1), *MSH6* (1), *PMS2* (2), *CPA1* (1), and *CFTR* (1). Six of the 9 variants were in *CDKN2A*- PC patients (p=0.028). All *ATM* variants including a novel frameshift were found in *CDKN2A*- PC patients. The variants in the *CDKN2A*+ PC patients included a novel *PMS2* frameshift variant, a *MSH6* substitution (frequency=0.0001) somatically mutated in colon cancer, and a *CFTR* substitution (frequency=0.0003). The findings suggest that rare variants in known high-risk PC-related genes may contribute to PC risk in a subset of patients from melanoma-prone families with and without *CDKN2A* mutations.

749M

Whole Exome Sequencing of 75 Hereditary Prostate Cancer Families. *E. Ostrander[1], D. Karyadi[1], E. Karlins[1], B. Decker[1], L. McIntosh[2], S. McDonnell[3], S. Middha[3], D. Schaid[3], S. Thibodeau[4], J. Stanford[2].* 1) Cancer Genetics Branch, NHGRI, NIH, Bethesda, MD; 2) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 4) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

Rare inherited mutations are predicted to account for 5% to 10% of all prostate cancer (PCa) cases. However, few causative mutations have been identified. Whole-exome sequencing (WES) of large hereditary PCa (HPC) families offers a new strategy to address genetic heterogeneity in PCa in order to identify rare PCa susceptibility mutations with moderate to high penetrance. Only a small number of families segregating the same rare mutation are needed to highlight genes of potential importance in PCa susceptibility. We performed WES in 75 HPC families from the *PROGRESS* study. In total, 160 affected men were sequenced with one to six affected men sequenced per family. Median read depth was 64.5. GATK UnifiedGenotyper was utilized for SNV and indel detection. All samples were coordinately genotyped and genotype quality was set to 99.9% concordance with the SNP chip. After quality filters, 453,977 variants were identified and genotyped in the 75 HPC families. We searched for rare variants that segregate with prostate cancer in multiple families and are predicted to damage the encoded protein. Rare variants were defined as those with frequencies ≤ 2% in all populations from multiple databases: NHLBI ESP European American and African American, and the five 1000 Genome populations. Protein damaging variants were defined as either high-impact variants (stop gain/loss, start loss, frameshift or splice site acceptor/donor) or missense variants that are SIFT deleterious or PolyPhen probably/possibly damaging. After population frequency and protein impact filters, 22,139 variants including 2,002 high impact variants remained. In order to determine the segregation with PCa, we linked the exome sequencing data to haplotype data from the Omni Express 700K SNP Chip. This approach allowed us to determine how many affected men in each family potentially carried the identified variants and was especially valuable in the 44 families with only one affected sequenced. After comparing the frequency in our dataset to the population frequencies, 861 variants were enriched in our set and present in three or more families, 110 of those variants were in six or more families and 89 were predicted to segregate in at least 50% of the cases on average. We have selected over 350 of the 861 variants for follow up validation and analysis in our population-based, case-control study of approximately 1,300 cases and 1,150 controls.

750T

IDENTIFICATION OF GENETIC VARIANTS IN A CONSANGUINEOUS FAMILY WITH PSYCHOTIC CASES USING AUTOZYGOSITY MAPPING AND WHOLE-EXOME NEXT GENERATION SEQUENCING. *A. Al Amri[1], J. Ivorra[1], M. Ali[2], C. Logan[2], A. Cardno[3], J. Mullins[4], T. Mahmood[5], S. Khan[2], C. Johnson[2], S. Clapcote[1], C. Inglehearn[2].* 1) school of Biological Sciences, St. James's University Hospital, Leeds, West Yorkshire, United Kingdom; 2) Leeds Institute of Biomedical & Clinical Sciences, University of Leeds (UK); 3) Leeds Institute of Health Sciences, University of Leeds (UK); 4) College of Medicine, Swansea University (UK; 5) Leeds and York Partnership NHS Foundation Trust (UK).

Psychosis is a condition in which an individual losses contact with reality. It covers profound forms of psychiatric disorders including schizophrenia, bipolar and schizoaffective disorder. These are complex disorders but have been known to run in families. The biological changes taking place in psychosis are not yet well understood but there is clear evidence of the involvement of both environmental and genetic factors. Here we describe the genetic analysis of a first-cousin consanguineous family of eight offspring, with two psychosis-affected cases, that was recruited from the West Yorkshire Pakistani population. Autozygosity mapping was combined with next generation sequencing to identify putative causative mutations. By using Affymetrix Genome-Wide SNP 6.0 arrays, two homozygous regions shared by the two affected siblings were identified on chromosomes 5q14.3-5q14.5 and 9q22.33-9q33.3. Whole-exome capture was carried out on one affected sibling using the SureSelect All Exon V4 reagent and sequenced on an Illumina HiSeq 2500 platform. The sequence reads were aligned against the human reference genome (hg19) and processed in SAM/BAM format using Picard and GATK. Annovar was used to annotate the variants before filtering according to depth of coverage, minor allele frequency, zygosity, mapping information and pathogenicity profile. Four variants, located in the genes DFNB31, MUSK1, OR1J1 and NUP188, passed the filtering criteria and were identified as the best candidates in this family. Work to assess the implications of these mutations in psychosis is currently ongoing. Keywords: Psychosis, Autozygosity mapping, Schizophrenia.

751S

Haploinsufficiency for DLX4 is Associated With Abnormal Craniofacial Development and Upregulated BMP4. *A.L. Choi[1], S. Mandal[1], J. Talbot[2], D. Wu[1], M. Prochazkova[3], H. Perry[3], V.L. Gil-da-Silva-Lopes[4], R. Lao[5], E. Wan[5], P. Tang[5], P-Y. Kwok[5], O. Klein[3], B. Zhuan[6], A. Slavotinek[1].* 1) Dept Pediatrics, Division Genetics, Univ California, San Francisco, San Francisco, CA; 2) Department of Molecular Genetics, Ohio State University; 3) Department of Orofacial Sciences, University of California, San Francisco, San Francisco; 4) Department of Medical Genetics, University of Campinas, São Paulo Paulo, Brazil; 5) Cardiovascular Research Institute, University of California, San Francisco, San Francisco; 6) School of Stomatology, Wuhan University, China.

Cleft lip and/or palate (CL/P) is found in 1 in 600 to 1 in 900 live births and is the commonest structural birth defect in humans. We used exome sequencing in a female and her son who had bilateral CL/P, euryblepharon and lagophthalmos, but no other features of blepharocheilodontic syndrome (BCDS). We identified a single nucleotide deletion in the mother, c.546_546delG, predicting p.Gln183Argfs*57 and replacement of the terminal 57 amino acids of the protein in the Distal-less 4 (DLX4) gene. The sequence variant was absent from public databases, predicted to be deleterious and was verified in mother and son by Sanger sequencing. There are three Dlx homeobox clusters with closely located gene pairs (Dlx1/Dlx2, Dlx3/Dlx4, Dlx5/Dlx6) and DLX4 has not been previously associated with disease. Both wildtype and mutant DLX4 localized to the cell nucleus and transfection studies with Western blotting did not show that the mutant DNA was significantly degraded. We used antisense morpholinos (MOs) to reduce expression of the orthologous Danio rerio genes, Dlx4a and Dlx4b, and found that morphant larvae injected with Dlx4b alone and Dlx4b together with Dlx4a had severe jaw malformations and cyclopia that were dose dependent. We also injected human wildtype DLX4 mRNA and did not observe a phenotype, implying that haploinsufficiency, rather than gain of function, was likely. We used siRNA to knockdown DLX4 in human cells and found a consistent increase in BMP4 expression. Co-injection of Dlx4b MO with an antisense MO targeting Bmp4 was able to prevent, but not rescue the Dlx4b phenotype, as there was substantial toxicity from the dual injections. We sequenced DLX4 in three unrelated patients with BCDS, but could not demonstrate any sequence variants. We then sequenced DLX4 in 155 patients of Chinese ethnicity with non-syndromic CL/P and CP, but no sequence variants (not even synonymous) were observed. From the published literature, Dlx1/Dlx2 double homozygous null mice and Dlx5 homozygous null mice both have a cleft of the secondary palate and a girl with cleft palate had a de novo chromosome deletion with the proximal breakpoint just 88 kb downstream from DLX5, leading to the hypothesis that dysregulation of DLX5 caused the cleft. Despite the fact that we have only one family, we consider that the clefts are likely to be due to haploinsufficiency for DLX4 and our findings thus have significance for craniofacial morphogenesis and for human CL/P.

752M

EXOME SEQUENCING REVEALS A NOVEL *CUBILIN* MISSENSE VARIANT ASSOCIATED WITH ALBUMINURIA IN AMERICAN INDIANS. *N. Franceschini[1], P.E. Melton[2], K. Haack[3], C. Bizon[4], E.T. Lee[5], L.G. Best[6], L.A. Almasy[3], J.A. Umans[7], S.A. Cole[3].* 1) Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Centre for Genetic Origins of Health and Disease, University of Western Australia, Crawley, Australia; 3) Texas Biomedical Research Institute, San Antonio, TX; 4) Renaissance Computing Institute, University of North Carolina, Chapel Hill, NC; 5) University of Oklahoma Health Sciences Center, Oklahoma City, OK; 6) Missouri Breaks Research Industries, Inc, Timber Lake, SD; 7) MedStar Health Research Institute, Hyattsville, MD.

Cubilin mediates the uptake of albumin in kidney proximal tubules, and a missense variant in the gene *CUBN* was associated with albuminuria in individuals of European ancestry. We previously were unable to replicate these associations (rs1801239, minor allele frequency [MAF]=0.17) with urine albumin-to-creatinine ratio (UACR) in American Indians. To identify coding variants in *CUBN* associated with UACR, we sequenced it's coding regions in 94 unrelated participants of the Strong Heart Family Study (SHFS), a large family study of American Indians recruited without phenotype ascertainment. We tested associations of log-UACR with each variant annotated to *CUBN* using linear regression methods, and additive genetic models adjusted for age, sex and its interactions. In a secondary analysis, we additionally adjusted for diabetes, and performed associations with clinically defined albuminuria (UACR≥ 30 mg/g) using logistic regression methods. Linkage disequilibrium (LD) was estimated from the data. The mean participant age was 57 years, 43% were men, and 55% had diabetes. The median UACR was 13 mg/g and 37% had UACR≥ 30 mg/g. Whole exome sequencing (mean call rate=0.98, Ti/Tv =2.50 and average variant read depth = 37x) identified 73 *CUBN* variants, 15 not previously described in publicly available databases, and 6 missense variants. In analysis of variants with MAF ≥ 0.03, a missense variant (rs1276712, MAF= 0.09 in American Indians) was significantly associated with UACR (beta 0.81, SE= 0.28, p=0.0048) when accounting for multiple testing. This variant explained 8.8% of the inter-individual variance of the trait, and it was associated with clinical albuminuria (odds ratio 4.4, p=0.02). The association was attenuated when adjusting for diabetes (p=0.02). rs1276712 is not in LD with the European ancestry rs1801239 (r^2=0.003 in American Indians), which was not significant in our data (p=0.55). rs1276712 C allele results in an amino-acid change (C2162Y) at cubilin domain 15, a site of direct interaction of cubilin with megalin, and required for albumin update at the proximal tubule. In conclusion, this study identified a novel and independent association of a *CUBN* missense variant with UACR and increased albuminuria in American Indians, suggesting allelic heterogeneity at this locus. Future studies will examine the impact of this variant in changes in albuminuria over time, and other clinically relevant chronic kidney disease outcomes.

753T

Examining Associations Between Multiple Sclerosis Cognitive Impairment and Genes Previously Associated with Cognitive Decline in Other Disease. *C. Holingue[1], M. George[1], C. Schaefer[2], A. Bernstein[3], R. Whitmer[2], L. Barcellos[1,2].* 1) Division of Epidemiology, Genetic Epidemiology and Genomics Laboratory, School of Public Health, University of California, Berkeley, CA, USA; 2) Kaiser Permanente Division of Research, Oakland, CA, USA; 3) Palm Drive Hospital, Sebastopol, CA, USA.

Cognitive impairment due to multiple sclerosis (MS) affects over half of individuals with the disease. Cognitive deficits can range from impairments in memory to diminished visual/spatial processing. These symptoms can affect social/emotional function, employment status, and quality of life. Yet, there is great variability in the severity and type of impairment that individuals with MS experience. The current evidence regarding the risk factors for cognitive impairment is limited. A recent finding shows evidence for association between the ε4 variant of the *APOE* gene and exacerbated cognitive decline in MS cases (Shi et al. 2011), which is also implicated in Alzheimer's Disease. This suggests there may be associations between MS cognitive decline and polymorphisms in genes previously associated with neurodegeneration. The current study investigated the association between cognitive score in MS cases and polymorphisms in genes related to cognitive impairment. Cases were identified from Kaiser Permanente, Northern California Region (KPNC). A cognitive score was calculated using the TICS-M (Modified Telephone Interview for Cognitive Status) at study entry. All participants provided biospecimens for DNA extraction; DNA samples were genotyped using Illumina's Human 660K BeadChip with imputation based on the 1000 Genome Reference and utilizing the IMPUTE2 software. To choose candidate genes, the NCBI gene database was used to search for genes associated with "cognitive impairment" in humans and then genes associated with Alzheimer Disease, Parkinson's Disease, Depression, Learning and Memory were extracted from the "Genes to Cognition Online" website (www.g2conline.org). This resulted in a total of 51 candidate genes. A sample of 954 individuals were analyzed using PLINK. Linear regression was performed with each SNP in the one of the above genes as the independent variable and cognitive score as the dependent variable. The top two principal components, derived from EIGENSTRAT SMARTPCA were included as covariates in the model, in order to account for possible confounding by ancestry. No evidence for association was observed for SNPs in these genes. This is the first study to investigate cognitive impairment in MS in relation to a broad set of genes previously associated with neurodegeneration or depression.

754S

Genetic contribution to cerebral palsy. *G. McMichael[1], M.N. Bainbridge[2], E. Haan[3], M. Corbett[1,4], A. Gardner[1,4], S. Thompson[5], B.W.M. van Bon[6], C.L. van Eyk[1], J. Broadbent[1], C. Reynolds[1], M.E. O'Callaghan[1], L.S. Nguyen[4], D.L. Adelson[7], R. Russo[8], S. Jhangiani[2], H. Doddapanein[2], D.M. Muzny[2], R.A. Gibbs[2], A.H. MacLennan[1], J. Gecz[1,4].* 1) The Robinson Institute, University of Adelaide, Adelaide, Australia; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, USA; 3) South Australian Clinical Genetics Service, SA Pathology (at Women's and Children's Hospital), North Adelaide, Australia; 4) School of Pediatrics and Reproductive Health, The University of Adelaide, Adelaide, Australia; 5) Department of Pediatric Neurology, Women's and Children's Hospital, North Adelaide, Australia; 6) Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands; 7) School of Molecular and Biomedical Science, The University of Adelaide, Adelaide, Australia; 8) Department of Pediatric Rehabilitation, Women's and Children's Hospital, North Adelaide, Australia.

Cerebral palsy (CP) is the most frequent cause of physical disability in childhood with a prevalence of 2-2.5 per 100 live births. It is largely sporadic and causally heterogeneous. Evidence of intrapartum fetal compromise is found at birth in <10% of cases. Several known epidemiological risk factors have been identified for CP including preterm delivery, intrauterine growth restriction and intrauterine infection. The contribution of genetic causes to CP is frequently overlooked. We performed whole exome sequencing of 183 cases with CP and their parents when available (98 trio and 67 duo cases). We used multiple prioritization criteria to identify potentially causative variants including the type of mutation (eg. protein truncating), involvement of a known disease gene, haploinsufficiency index, brain expression, Residual Variation Intolerance Score and Combined Annotation-Dependent Depletion. Based on these strict prioritization criteria, 10 de novo mutations were predicted to be causative for CP; three in known disease genes: TUBA1A (n=2), SCN8A (n=1) and KDM5C (n=1) and six in novel candidate CP genes: AGAP1, JHDM1D, MAST1, NAA35, RFX2 and WIPI2. Additionally, we found four inherited from an unaffected mother to affected son X-chromosome variants, in two known disease genes L1CAM and PAK3 and two novel candidate CP genes CD99L2 and ODZ1 respectively. In summary, 14% of CP cases have a predicted pathogenic disease variant with 8 occurring in novel candidates for CP. The genetic heterogeneity of CP revealed by these results reflects the complex and clinically variable nature of this disorder. These results highlight a previously unappreciated role for the contribution of genetics to CP causation, which will influence future management, especially genetic counselling.

755M

Testing the effect of compound heterozygosity on anthropometric traits in the general population. *S. Lessard[1], P.L. Auer[2], A. Giri[3], C. Schurmann[4], T. Karaderi[5], H. Yaghootkar[7], C. Lindgren[5,6], T. Edwards[3], T. Frayling[7], R.J.F. Loos[4], G. Lettre[1], GIANT ExomeChip Consortium.* 1) Montreal Heart Institute and Université de Montréal, Montreal, Quebec, Canada; 2) University of Wisconsin, Milwaukee, Milwaukee, USA; 3) Vanderbilt University, Nashville, USA; 4) Icahn School of Medicine at Mount Sinai Hospital, New York, USA; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; 6) Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge 02142, MA, USA; 7) Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK.

Background: In Mendelian genetics, a recessive disease can result when two different recessive alleles in the same gene are inherited in trans, that is on each copy of the gene. This genetic mechanism - termed compound heterozygosity - is observed for several monogenic disorders, such as phenylketonuria and sickle cell disease. However, the role of compound heterozygosity in human complex diseases and traits has not been explored extensively. **Method:** We analyzed height, body mass index (BMI) and waist-to-hip ratio (WHR) in up to 68,000 individuals from 9 studies included in the GIANT ExomeChip Consortium. We applied inverse normal transformation on phenotypes after adjusting for sex and age. All participants were genotyped using the Illumina ExomeChip. We defined compound heterozygotes as individuals that carry at least two loss-of-function variants in the same gene, assuming that such variants are rare and not inherited in cis. We compared the phenotypic mean in compound heterozygotes to the mean in the remaining individuals, and assess statistical significance using phenotype permutations. We combined results across studies using a weighted Z-score meta-analysis method, where the weight is the sample size of the corresponding study. **Results:** By considering 5,624 nonsense and 10,578 splice site variants genotyped on the ExomeChip, we identified at least one compound heterozygote for 2,911 genes for height, 2,664 genes for BMI and 1,795 genes for WHR. No genes reached a Bonferroni-corrected threshold for statistical significance ($P<6\times10^{-6}$). However, there was a strong signal for the cyclin D1 gene with height (*CCND1* $P=1.9\times10^{-5}$) and the parathyroid hormone-like hormone gene with WHR (*PTHLH*, $P=9.9\times10^{-5}$). In both cases, however, the signal was due to a common variant and reflects a "true" recessive signal. The strongest compound heterozygosity signal observed is for BMI with the taste receptor gene *TAS1R1* ($P=5.1\times10^{-4}$, five nonsense variants included). The BMI of the 137 *TAS1R1* compound heterozygotes is on average 0.74 kg/m2 higher than the BMI of the remaining 40,210 participants. **Conclusion:** The role of compound heterozygosity has rarely been tested in the context of complex human diseases and traits. Here, we successfully developed a method of analysing this model in large meta-analysis data. As with other studies of predominantly low-frequency and rare variants, our data suggest that very large sample sizes will be needed to detect robust associations.

756T

Locating new genes which may be involved in the development of Primary Congenital Glaucoma. *D. Bercovich[1], A. Wolf[2], O. Geyer[2].* 1) Human Molec Gen & Pharm, Tel Hai College, Galile Elyon, Israel; 2) Department of Ophthalmology, Carmel Medical Center, Israel.

Since there are Primary Congenital Glaucoma patients (PCG) who do not contain mutations in genes known to be involved in this disease (like the CYP1B1, MYOC & FOXC1 genes), prenatal screening and diagnosis in fetuses suspected with this disease is not viable. Furthermore, after birth, it is difficult to accurately identify primary Congenital Glaucoma, in many cases it is miss-diagnosed as another syndrome associated with eye diseases. The main problem is the Jewish population, in which individuals with significant mutations in these genes is rare. In order to locate the gene (or genes), which may be involved in this disease, Full sequencing of the whole Exom was preform by next generation sequencing (NGS) method to identify new loci with point mutations or possible copy number variations. 6 PCG Jewish patients from different ethnic groups (Ashkenazi, Sfaradi, Balkan and Ethiopia) and two parents pairs, to combine this data with their patients NGS (a total of 10 DNAs) were screened. 87 different genes were found be possible candidates in the recessive model. All DNA alterations were found to be in the same way in the parents as well which exclude them as been the pathogenic DNA alterations. One locus with CMA was found to be varied among Congenital Glaucoma patients and family members. This indicated that copy number around this gene could contribute somehow to the Congenital Glaucoma phenotype and more investigation may be done and this possible connection of this gene.

757S

Genome-wide Copy Number Variants in Chronic Obstructive Pulmonary Disease (COPD). *F. Begum[1], I. Ruczinski[2], S. Li[3], M. Parker[1], J. Hetmanski[1], T. Beaty[1], E. Silverman[4], J. Crapo[5], COPDGene Investigators.* 1) Department of Epidemiology, Johns Hopkins University, Baltimore, MA., USA; 2) Department of Biostatistics, Johns Hopkins University, Baltimore, MA, USA; 3) Cancer Genomics Research Laboratory (CGR), Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA; 4) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; 5) Department of Medicine, National Jewish Health, Denver, USA.

Chronic Obstructive Pulmonary Disease (COPD), a common lung disease, is the third leading cause of death in the United States. This chronic and progressive disease has both environmental (mainly cigarette smoking) and genetic risk factors. Several genome-wide association studies (GWAS) have identified multiple genes influencing risk to COPD including *CHRNA3*, *FAM13A*, *HHIP*, *RIN3*, *MMP12*, and *TBFB2*. To properly understand the genetic etiology of COPD, it is also important to explore the role of copy number variants (CNVs) since the presence of CNVs can alter gene expression and may be causal for disease. We delineated CNVs using PennCNV on 9076 COPDGene study subjects using genome-wide marker data generated using Illumina's Omni-Express array. COPDGene subjects are comprised of one-third African-American and two-thirds Non-Hispanic white adult smokers, with or without COPD After employing rigorous quality control procedures to reduce the false positive CNV calls, we tested for association between CNV components (defined as disjoint intervals of copy number regions within race) and several COPD related phenotypes. We detected hemizygous deletions that achieved genome-wide significance on chromosome 5q35.2, near the gene *FAM153B*, in tests of association with total lung capacity assessed by chest CT among African-Americans. This region includes multiple reported CNVs, and we are currently following up on this signal.

758M

Investigating the role of salivary amylase copy number in obesity using low-pass whole genome sequencing. *M.A. Tuke[1], A.R. Wood[1], L. Harries[6], L. Boquete-Vilarino[1], M. Nalls[2], D. Hernandez[2,3], S. Bandinelli[4,5], A. Singleton[2], D. Melzer[6], L. Ferrucci[7], M.N. Weedon[1], T.M. Frayling[1].* 1) Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK; 2) Laboratory of Neurogenetics, National Institute of Aging, Bethesda, Maryland, USA; 3) Department of Molecular Neuroscience and Reta Lila Laboratories, Institute of Neurology, UCL, London, United Kingdom; 4) Tuscany Regional Health Agency, Florence, Italy, I.O.T. and Department of Medical and Surgical Critical Care, University of Florence, Florence, Italy; 5) Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy; 6) Institute of Biomedical and Clinical Sciences, University of Exeter Medical School, Barrack Road, Exeter, UK; 7) Longitudinal Studies Section, Clinical Research Branch, Gerontology Research Center, National Institute on Aging, Baltimore, Maryland, USA.

Copy number variation at the salivary amylase gene (AMY1) has recently been strongly associated with body mass index (BMI) and obesity (Falchi et al. Nature Genetics, 2014). Each additional copy of the AMY1 gene was associated with an odds ratio of 1.19 for obesity compared to normal weight individuals and the variant was reported as explaining a larger proportion of the variance in BMI than the variant at the FTO/IRX3 locus. Here, we aimed to further replicate and characterise this association in the InCHIANTI study. We performed whole genome sequencing on 657 samples to an average depth of 7x. We aligned the reads to a repeat masked GRC build 37 reference genome using the mrsFAST alignment algorithm that can align single reads to multiple positions in the genome and so is optimal for regions of variable copy number. We derived an absolute copy number value for the full amylase region using mrCaNaVaR. We then regressed the resulting copy number against body mass index corrected for age and sex. We had between 88% to >99.9 power to detect association when assuming the published variance explained of between 1.73% and 7.94%. The preliminary analysis did not provide any evidence of an association between salivary Amylase copy number and BMI (P = 0.457). However, our work is ongoing and we are validating the association using quantitative PCR and expanding the analysis to several thousand individuals. In conclusion, our preliminary results do not support the association between BMI and salivary amylase gene copy number.

759T

Novel Missense Mutations in ABCC8 and Type 2 Diabetes in Pima Indians. *L.J. Baier, Y.L. Muller, K. Huang, A. Nair, B. Gene, A. Stacy, V. Ossowski, C. Wiedrich, E.J. Weil, R.G. Nelson, W.C Knowler, P.H. Bennett, R.L. Hanson, C. Bogardus.* PECRB, NIDDK/NIH, Phoenix, AZ.

To identify novel mutations that affect risk for type 2 diabetes in American Indians, we obtained whole genome sequence data on 335 Pima Indians. One initial analysis focused on identifying novel variants predicted to be damaging in genes previously implicated with type 2 diabetes. These variants are currently being genotyped in additional samples (N=7,355) for association analysis with type 2 diabetes. Using this approach, we identified 6 novel missense variants and one known variant (rs757110; predicts an Ala1369Ser) in the ABCC8 gene that encodes the sulfonylurea receptor 1 protein (SUR1). SUR1, together with inward-rectifier potassium ion channel encoded by the KCNJ11 gene, regulate insulin secretion by ATP/ADP-sensing. Rare activating mutations in SUR1 cause neonatal diabetes while rare inactivating mutations cause hyperinsulinemia of infancy. The common SNP rs757110 has been associated with type 2 diabetes in some studies, but in our sample had no association (minor allele frequency 0.39; odds ratio (OR)= 0.99 [95% confidence interval 0.88-1.11]; P= 0.9 adjusted for age, sex, birth year, fraction of Pima heritage). Of the 6 novel variants, 4 have been genotyped in follow-up samples to date (Arg1420His, Gly1316Gln, Met801Ile, and Ser165Leu) while 2 are pending genotyping (Asp691Glu and Lys1565Glu). Among the variants with follow-up genotypic data, Arg1420His provided the most compelling evidence for having a role in type 2 diabetes. Individuals with a His allele (243 heterozygotes and 1 homozygote; overall frequency = 0.02 in Pima Indians) were at higher risk for type 2 diabetes (OR= 1.96, 95%CI=1.4 -2.7; P= 6×10^{-5} adjusted as above) despite being leaner (BMI = 33.4 vs. 35.3 kg/m^2, respectively for Arg/His vs Arg/Arg; P= 0.002 adjusted for sex, age, birth year, and fraction of Pima). The one individual homozygous for the risk allele (His/His) had been diagnosed with hyperinsulemia and hypoglycemia at 4 months of age, and was diagnosed with type 2 diabetes at 3.5 years of age, suggesting a very profound phenotype for this variant in homozygosity. Arg1420 is located in the second nucleotide binding domain of SUR1 and therefore a substitution at this position is predicted to have a functional impact. In conclusion, whole genome sequencing identified potentially functional variation in ABCC8 that appears to contribute to type 2 diabetes in Pima Indians.

760S

Targeted sequencing of genes associated with type 2 diabetes in 6800 individuals. *V. Bansal[1,2], J. Gassenhuber[5], T. Phillips[1], G. Oliviera[1], N. Villarasa[1], R. Tisch[1], E.J. Topol[1], B.O. Boehm[3,4].* 1) Scripps Translational Science Institute, La Jolla, CA; 2) Department of Pediatrics, University of California San Diego, La Jolla, CA; 3) Department of Internal Medicine, Division of Endocrinology and Diabetes, University Medical Centre Ulm, Ulm, Germany; 4) Lee Kong Chian (LKC) School of Medicine, Nanyang Technological University, Singapore and Imperial College London, London, UK; 5) Sanofi-Aventis Germany GmbH, Frankfurt am Main, Germany.

Advances in high-throughput sequencing technologies present unprecedented opportunities for understanding the role of rare variants in complex diseases. For type 2 diabetes, more than 70 common genetic variants have been identified by extensive genome-wide association studies. We utilized a candidate gene approach to search for rare variants associated with type 2 diabetes and focused on genes within loci identified by GWAS for T2D and related traits, genes involved in monogenic forms of diabetes and additional genes linked to diabetes. The coding and 5' and 3' untranslated regions of 130 genes (1.2 megabases in total) were sequenced in 1880 cases (including 600 patients with age of onset < 40 years) and 1840 controls. To increase efficiency, sequencing was done in pools (20 individuals per pool) and library preparation for each pool was done using the Agilent SureSelect target enrichment method. To process the pooled sequence data generated in this study, we developed computational methods for variant calling, assessment of population stratification, and rare variant association analysis. Single variant association analysis and gene-based association analysis of rare variants using a number of statistical methods (CAST, Calpha, and SKAT) did not identify any new statistically significant associations. To increase power, we sequenced an additional 2100 type 2 diabetes cases and 1050 controls using the same approach. Joint analysis of rare variants in the two cohorts identified 10 variants (minor allele frequency in the 0.1-0.5% range) that were moderately associated with T2D with effects on the phenotype in the same direction. This list included a rare non-synonymous variant in the MTNR1B gene that was recently associated with T2D (Bonnefond et al, Nature Genetics 2012). In addition, we observed a significant excess of rare non-synonymous variants in the glucokinase (GCK) gene in T2D patients with an early age of onset. Our analysis suggests that rare variants that have a moderate to strong effect on the risk for type 2 diabetes are likely to have very low minor allele frequencies and sequencing of tens of thousands of individuals is needed to identify such variants. Ongoing work includes (1) genotyping of select rare variants in additional cohorts and (2) further phenotype analysis of individuals with specific variants of interest.

761M

Characterizing variation under linkage peaks in families. *K.L. Edwards[1], J.Y. Wan[1], C. Johnson[1], S. Santorico[2].* 1) Epidemiology, University of California Irvine, Irvine, CA; 2) University of Colorado, Denver Denver, CO.

Background: The Metabolic Syndrome (MetS) is a complex condition characterized by a cluster of CVD risk factors, including obesity, lipid abnormalities, hypertension and glucose intolerance. In our previous work using extended pedigrees with MetS and linkage analysis we identified three chromosomal regions with evidence for linkage to multivariate traits defined by clusters of MetS risk factors. Methods: Whole exome deep sequencing was performed (50x) using the library and exome capture with 62Mb target from Roche/Nimblegen SeqCap EZ v2.0 (~300,000 exons and flanking regions), the HiSeq sequencer, and GATK read processing. Whole exome sequencing was performed in a subset of linked families of European, Mexican, African and Japanese-American descent. Results: We provide a descriptive summary of rare and common variation found under one of the linkage peaks on chromosome 2. Results are based on sequencing 5 European American families consisting of two families (n=11 individuals) with linkage to the region and three families (n=17 individuals) with no evidence of linkage to the same region. Among the 267 biallelic variants in the linked families, 34 (12.7%) were possibly damaging (PolyPhen>0.80) and 72 had a CADD score >10. In the unlinked families 332 biallelic variants were identified, with 48 variants (14.5%) having a PolyPhen >0.80 and 92 variants with a CADD >10. The minor allele frequency of deleterious variants appears to be higher in our linked families, with 25% having an alternate allele frequency (AF) of 0.127 or below compared to 25% with an AF of 0.078 or below in the unlinked families. Among the linked families 6 variants (2.2%) were private and were not found in the unlinked families or 1000 Genomes. We also describe the distribution of variants in this same linkage region in Mexican, Japanese and African American families sequenced under the same project and compared to 1000 Genomes. Summary: Overall it appears that the frequency of rare and private mutations is higher in regions under linkage peaks.

762T

An exome-wide sequencing study for type 2 diabetes-associated kidney disease in African Americans. *M. Guan[1,2], P. Mudgal[2], J.G. Wilson[3], B.I. Freedman[4], D.W. Bowden[2,5,6], M.C.Y. Ng[2,5].* 1) Integrative Physiology and Pharmacology Program, Wake Forest University, Winston-Salem, NC., USA; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA; 3) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA; 4) Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA; 5) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA; 6) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA.

End-stage kidney disease (ESKD) is a significant worldwide public health concern; diabetes accounts for 44% of incident ESKD cases in the U.S. Previous GWAS from our group have identified several strongly associated SNPs for type 2 diabetic (T2D)-ESKD. However, the common SNPs identified in GWAS have modest effects and cumulatively explain only a small proportion of disease variance. Therefore, we evaluated the contribution of low frequency coding variants in T2D patients on dialysis or with high risk for T2D-ESKD (eGFR<30 ml/min/1.73 m^2) since these variants may harbor stronger effects. Exome sequencing data in 1,565 African Americans including 537 T2D-ESKD cases and 1,028 non-T2D, non-ESKD controls from the T2D-GENES Consortium exome sequencing study were examined. Single variant association was performed for an additive effect with a mixed linear model that corrects for sample structure and hidden relatedness (EMMAX). Sequencing Kernel Association Test (SKAT) was used for gene-based analyses with stronger weighting for rare SNPs. Nine rare variants from nine different genes were significantly associated with T2D-ESKD at exome-wide significance level (p<5E-7). The top signal was located in MRPL12 (rs62077220, p=2.25E-18, MAF=0.037) which encodes a 39S subunit protein for mammalian mitochondrial ribosomal proteins. Only two of nine T2D-ESKD associated genes from single variant analyses demonstrated modest significance in gene-based analyses. Interestingly, several regions achieved nominal significance in both tests. One of the novel genes was VPS33A, with single variant association at p=1.41E-5 (rs2271411, MAF=0.3) and gene-based p-value at 4.4E-4, respectively. VPS33A plays a vital role in segregation of intracellular molecules into distinct organelles. A mutation in VPS33B, which is an important paralog of VPS33A, was reported to be associated with arc syndrome (Arthrogryposis-renal dysfunction-cholestasis syndrome). Overall, these findings may partially explain the genetic predisposition to T2D-ESKD in African Americans. Replication in independent African American populations remains necessary.

763S

Exome chip meta-analysis identifies novel loci and low-frequency variants contributing to central body fat distribution. *A.E. Justice[1], H.M. Highland[2], K.L. Young[1], M. Graff[1], T. Karaderi[3], N.L. Heard-Costa[4,5], D. Pasko[6], V. Turcot[7], Y. Lu[8], L. Southam[9], L.A. Cupples[10], C.T. Liu[10], C.S. Fox[4], T.W. Winkler[11], N. Grarup[12], R.A. Scott[13], M.M. McCarthy[14], K. Mohlke[15], R.J.F. Loos[8], I. Borecki[16], K.E. North[1], C.M. Lindgren[17] For the BBMRI, the GOT2D, the CHARGE, and the GIANT Consortia.* 1) Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA; 2) Human Genetics Center, The University of Texas Graduate School of Biomedical Sciences at Houston, The University of Texas Health Science Center at Houston, School of Public Health, Houston, TX, USA; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 4) National Heart, Lung, and Blood Institute, the Framingham Heart Study, Framingham, MA, USA;; 5) Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 6) Genetics of Complex Traits, University of Exeter, UK; 7) Montreal Heart Institute, University of Montreal, Canada; 8) The Genetics of Obesity and Related Metabolic Traits Program, The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 9) The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; 10) Boston University School of Public Health, Boston, MA; 11) Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany; 12) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark; 13) MRC Epidemiology Unit, University of Cambridge, UK; 14) Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, UK; 15) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 16) Department of Genetics Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO, USA; 17) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA.

Increased central fat is a leading predictor of greater cardiometabolic risk. Recent work shows that genetic factors contribute to the distribution of central fat, measured here as waist to hip ratio adjusted for BMI (WHR^a), which is correlated with visceral adiposity (r=0.56). Also, among the loci with known association with WHR^a, more than half show marked sexual dimorphism. The genetic underpinnings of WHR^a may include rare and protein-coding variants with high penetrance and large effect sizes; however, the majority of genetic studies focus on common variants in European-descent populations. Thus, we set out to identify coding and low-frequency variants (LFVs) (MAF<5%) associated with central fat distribution across the human exome using exome array data from 90,224 women and 75,759 men of European (90%), African (8%), and Asian (2%) ancestries. We performed a two-stage meta-analysis using an inverse variance weighted approach: 1) ancestry-specific meta-analysis of the study-specific summary results for WHR^a; and 2) meta-analysis of ethnic-specific results. Analyses included up to 236,047 exome array single nucleotide variants (SNVs) (201,126 with MAF<5%). Four SNVs reached genome-wide significance (GWS) (P<5E-8) in men only, with one novel SNP in *RREB1* (MAF=40%, P=4.4E-9, β=0.03). A total of 16 SNVs reached GWS for women, including two novel loci in *RAPGEF3* (MAF=1.8%, P=7.6E-9, β=0.139) and *DNAJA3* (MAF=29%, P=3.4E-8, β=0.03). The non-synonymous LFV within *RAPGEF3* exhibited a much greater effect on WHR^a compared to common variants (β=0.026 to 0.062). Similar to previously identified WHR^a loci, *RAPGEF3* is involved in angiogenesis and the regulation of insulin secretion. Interestingly, RAPGEF3 activity is decreased in presence of progesterone, an ovulatory sex steroid present in high levels in women of reproductive age, potentially helping to explain the GWS for this locus in women, but not men. These observations highlight the importance of large-scale genomic studies for identifying central fat distribution susceptibility LFVs. We plan to extend our analyses to include a greater number of individuals from diverse ancestries, and to conduct gene-based association and pathway analyses to identify genes and pathways influencing central fat distribution. Such efforts will provide insights into the etiology of central fat distribution and may highlight population-specific variants that pose increased susceptibility among individuals at greater risk.

764M

Whole genome sequence based analysis of thyroid function. *NJ. Timpson[1], PN. Taylor[2], E. Porcu[3,4,5], S. Chew[6], PJ. Campbell[6], M. Traglia[7], SJ. Brown[6], BH. Mullin[6,8], HA. Shihab[1], J. Min[1], K. Walter[9], Y. Memari[9], J. Huang[9], MR. Barnes[10], JP. Beilby[11,12], JRB. Perry[13,14], P. Danecek[9], D. Muddyman[9], V. Panicker[6], D. Toniolo[7,15], CM. Dayan[2], S. Naitza[3], JP. Walsh[6,8], TD. Spector[14], G. Davey-Smith[1], R. Durbin[9], B. Richards[14,16,17], S. Sanna[3], N. Soranzo[9], SG. Wilson[6,8,14], The UK10K Consortium.* 1) MRC/UoB IEU/SSCM, Bristol Univ, Bristol, United KingdomMRC Integrative Epidemiology Unit at the University of Bristol, University of Bristol, Bristol, United Kingdom; 2) Thyroid Research Group, Institute of Molecular & Experimental Medicine, Cardiff University School of Medicine, Cardiff University, Cardiff, United Kingdom; 3) Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, c/o Cittadella Universitaria di Monserrato, Monserrato, Cagliari, Italy; 4) Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy; 5) Center for Statistical Genetics, Biostatistics Department, University of Michigan, Ann Arbor, MI, USA; 6) Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia; 7) Division of Genetics and Cell Biology, San Raffaele Research Institute, Milano, Italy; 8) School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia, Australia; 9) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, United Kingdom; 10) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom; 11) Pathwest Laboratory Medicine WA, Nedlands, Western Australia, Australia; 12) School of Pathology and Laboratory Medicine, University of Western Australia, Crawley, Western Australia, Australia; 13) MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Box 285, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, United Kingdom; 14) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 15) Institute of Molecular Genetics-CNR, Pavia, Italy; 16) Department of Medicine, Jewish General Hospital, McGill University, Montréal Québec, Canada; 17) Departments of Human Genetics, Epidemiology, and Biostatistics, Jewish General Hospital, Lady Davis Institute, McGill University, Montréal Québec, Canada.

Thyroid hormones are essential for health and maintenance of metabolic balance. Work here aimed to identify genetic variants, including rare variants, associated with the heritable traits: thyrotropin (TSH) and free thyroxine (FT4) using whole genome sequencing (WGS) data from the UK10K Consortium. TSH and FT4 levels were standardized adjusting for age, age2 and sex. We analyzed data from the two UK10K cohorts (ALSPAC and TwinsUK N=2,287) and used an additional collection with WGS data (SardiNIA) and deeply imputed datasets (imputed to a joint 1000g and UK10K reference panel) to perform a meta-analysis for common variants (MAF >1%) associated with TSH and FT4 (N=16,335). We then undertook analysis of rare variants (MAF <1%) using sequence kernel association testing (SKAT) in 40 candidate genes and performed genome-wide complex trait analyses (GCTA) to explore the extent that common SNPs (MAF >1%) explained the variance in TSH and FT4. For TSH we report a novel variant at 3p25 (MAF= 23.5%, P=6.15×10-09) and a new independent variant in established locus (MAF=10.4%, P=5.94×10-14). Expression quantitative trait locus analysis revealed our variant at 3p25 modulates gene transcription in adipose, skin and whole blood cells. Methylation profiles also revealed evidence for methylation quantitative trait locus effects for our novel variant in an established locus (P=4.38×10-07). For FT4 we identified a low frequency variant in 18q11 (MAF=3.2%, P=1.27×10-09) tagging a rare functional variant (MAF= 0.4%, P=2.14×10-11) known to be associated with substantially raised FT4 levels through altered thyroxine binding in transthyretin. SKAT analysis also revealed a novel association with FT4 in chromosome 8p12 (P=2.53×10-06). GCTA analysis estimated that common SNPs (MAF>1%) explained 24% (95%CI 19, 29) and 20% (95%CI 14, 26) of TSH and FT4 variance, respectively (P=<0.0001). Our results demonstrate that the increased coverage in WGS population association studies allows novel association detection of both common and rare variants in thyroid function. Common variants collectively account for over 20% of the variance in TSH and FT4; a substantial advance on estimates from earlier genome-wide association studies.

765T

Gene-variants associated with familial mesial temporal lobe epilepsy identified by whole exome sequencing. *R. Secolin[1], P.A.O. Ribeiro[1], F.R. Torres[1], M.G. Borges[1], A.C. Coan[2], M.E. Morita[2], C.V. Soler[1], M.L. Santos[1], C.V. Maurer-Morelli[1], B.S. Carvalho[1], F. Cendes[2], I. Lopes-Cendes[1].* 1) Department of Medical Genetics, University of Campinas - UNICAMP, Campinas, SP, Brazil; 2) Department of Neurology, University of Campinas - UNICAMP, Campinas, SP, Brazil.

Background: Epilepsy is a common chronic neurological disorder that affects approximately 1% of the population worldwide. Familial mesial temporal lobe epilepsy (FMTLE) is a clinically well characterized syndrome with an autosomal dominant inheritance. **Objective:** To identify genes associated with FMTLE in two large families. **Methods:** Whole exome sequencing was performed in eight individuals from two FMTLE families (F-10 and F-26), including six patients and two unaffected parents. Exome was targeted with Nextera Rapid Capture Expanded Exome kit (Illumina™) and sequenced in a high-performance Hiseq Illumina 2500 sequencing machine (Illumina™) to obtain more than 50X average coverage per sample. A bioinformatics analysis was performed using the GATK software package. Sequences were aligned using BWA algorithm. Variant calling and functional prediction was performed using VariantAnnotator and SnpEff tools. We prioritized nonsynonymous, frameshift, splicing, and indel variants according to novelty, quality score, and putative pathogenicity. **Results:** We found a total 1,955,506 and 2,134,863 variants in F-10 and F-26 families, respectively. After bioinformatics processing, we identified 184 functional variants in F-10 and 193 in F-26, which are shared only by patients and absent in unaffected individuals. Among them, we observed three genes as potential candidates for FMTLE in the families studied. **Conclusions:** Putative roles related to formation of axon connections, protein-protein interaction, phosphorylation-dependent ubiquitination pathway make the three candidate genes identified relevant for FMTLE. **Supported by:** CEPID-BRAINN FAPESP, São Paulo, Brazil.

766S

Mutations in Human Capicua Gene Found in Patients with CFD and NTDs. *Y. Lei[1], H. Zhu[1], M. Parker[1], W. Yang[3], G. Shaw[3], R. Finnell[1,2].* 1) Dell Pediatric Research Institute, Department of Nutritional Sciences, The University of Texas at Austin, Austin, Texas; 2) Department of Chemistry, College of Natural Sciences, The University of Texas at Austin, Austin, Texas; 3) Department of Pediatrics, Division of Neonatology, Stanford University School of Medicine, Stanford, CA.

Cerebral folate deficiency (CFD) syndrome is characterized by very low concentration of 5-methyltetrahydrofolate (5-MTHF) in cerebrospinal fluid, while folate levels in plasma and red blood cells are normal. Previously, mutations in several folate pathway genes, including hFRα (folate receptor alpha), DHFR (dihydrofolate reductase), and PCFT (proton coupled folate transporter) have been identified in CFD patients. In an effort to identify causal mutations for CFD, we performed whole exome sequencing analysis of DNA samples collected from a CFD patient, her healthy sibling, and her biological parents. A de novo mutation in human Capicua gene (CIC), c.1057C>T (p.R353X), was identified in the patient. The results were confirmed using Sanger sequencing. In addition, a missense mutation predicted to be damaging, c.1738G>GT (p.G580GC) was identified in another CFD patient. The CIC protein is a HMG-box transcriptional repressor. The DNA binding domain located at amino acid residues 200-268 binds the octomer sequence T(G/C)AATG(A/G)A. The mutation identified in the CFD patient, p.R353X, yields a truncated protein which still contains the DNA binding domain (HMG box), therefore it is still able to bind to its targets. CIC target binding octomer sequence has been found in the promoter regions of folate transport genes FOLR1, PCFT, RFC1, and DHFR, which is involved in folate metabolism. In the patient's induced pluripotent stem (iPS) cell, the p.R353X mutation down regulated FOLR1, PCFT and RFC1 gene expression compared with H9 stem cells and an iPS cell line from an individual with wildtype CIC. Chromatin immunoprecipitation assays demonstrated that CIC bound to the FOLR1, PCFT and RFC1 promoter in vitro. In dual-luciferase assay, the CIC protein repressed FOLR1 promoter transcription. CIC coding regions were also sequenced in 190 patients with spina bifida, a common form of neural tube defects (NTDs), from California. Four missense mutations (p.S68P, p.M379L, p.P484S, p.P896A) were identified. Three of the mutations (p.M379L, p.P484S, p.P896A) were predicted to be damaging by MutationTaster and FATHMM. When compared to the rare (<1%) missense mutation frequency (38/6464) based on NHBLI exome sequencing database, there is significant difference between the spina bifida cohort and the NHBLI cohort. Our findings suggest that human Capicua gene regulates folate transporter gene expression, and may play a role in the etiology of CFD and NTDs.

767M

Combining linkage analysis and whole-exome sequencing for the identification of novel ADHD-related variants in multi-generation pedigrees. *J. Corominas-Galbany[1], M. Klein[1], M. Onnink[1,2], S. Gross-Lesch[3], C. Jacob[3], A. Reif[3], M. Romanos[4], B. Franke[1,2], K.P. Lesch[3].* 1) Department of Human Genetics, Radboud university medical center, Nijmegen, The Netherlands; 2) Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud university medical center, Nijmegen, The Netherlands; 3) Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany; 4) Department of Child Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany.

Attention-Deficit/Hyperactivity Disorder (ADHD) is a highly heritable and multifactorial disorder characterized by inattention, and/or motor hyperactivity and impulsivity. Susceptibility loci for ADHD have been investigated in the recent years and valuable candidate genes have been suggested. Nevertheless, the understanding of the genetic basis of ADHD is still a challenge. Large multi-generation pedigrees with affected individuals allowed the identification of haplotype blocks shared among all ADHD-affected family members. The characterization of these region is essential for the identification of genes and variants that may be related to the ADHD phenotype. Hence, the main goal of this study is to combine linkage analysis with whole-exome sequencing (WES) in order to identify novel genetic ADHD risk factors. This study was performed with nine ADHD families of German descent, for which WES data was obtained from 3 or more affected individuals to detect rare and common genetic variants shared among all sequenced members of a family. WES was carried out using Agilent SureSelect All Exon 50Mb Target Enrichment kit and single-end sequencing on the 5500xl SOLiDTM System. All individuals were genotyped on GeneChip Human Mapping 50K Array Hind240 and SNP data was used to fine-map the ADHD-associated loci previously described. Rare and common candidate variants to be associated with ADHD were validated using Sanger sequencing. ADHD phenotype and haplotype co-segregation in large pedigrees points towards a dominant inheritance pattern of the disease. An initial genome-wide approach focused on one family allowed the identification of 7 rare variants, but none of them segregated with phenotype through the family. This result made us move to the inclusion of linkage analysis data, because segregation of ADHD does not have to be due to rare variants only. In fact the combination of several common and rare variants, each with small or intermediate effect, may underlie co-segregating haplotype blocks. Linkage analyses performed showed an haplotype co-segregating with ADHD-affected individuals at 9q31-9q32 that lead to the detection of several common and rare variants of interest in distinct genes. Validation of these variants and segregation analysis through the family are ongoing. This study uses a complementary strategy in addition to GWAS approaches to detect variants that may shed new light on our understanding of ADHD's genetic basis.

768T

Targeted exome sequencing in extended pedigrees with type 2 diabetes identifies a novel diabetic nephropathy susceptibility gene. *M.G. Pezzolesi[1,2], A.M. Smiles[1], J. Skupien[1], J.C. Mychaleckyj[3], S.S. Rich[3], J.H. Warram[1], A.S. Krolewski[1,2].* 1) Section on Genetics and Epidemiology, Joslin Diabetes Center, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA; 3) Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA.

Genome-wide linkage analysis in extended pedigrees from the Joslin Study of Genetics of Nephropathy in Type 2 Diabetes Family Collection (Joslin T2DN Family collection) previously identified several genomic regions with evidence of linkage for urinary albumin excretion levels or variation in renal function in individuals with type 2 diabetes (T2D). Several in independent studies have reported evidence of linkage to many of these same regions, however, to date no gene located within or near these loci that contributes to these diabetic nephropathy (DN) sub-phenotypes has been established. To advance these efforts, we recently performed targeted exome sequencing of all protein-coding genes across 6 loci (f4 genomic regions with evidence for linkage with urinary albumin excretion levels: chromosomes 5q, 7q, 21p, and 22q; and 2 genomic regions linked to variation in renal function: chromosomes 2q and 7p) using a custom target enrichment library followed by Next-Generation sequencing. A total of 662 members from 49 extended pedigrees in the Joslin T2DN Family collection were resequenced for the coding region of the 361 genes across the 6 linkage regions. Following quality control analyses, multi-sample variant calling was performed for 603 samples that exceeded 35X on-target read depth and with greater than 15-fold sequence coverage across 70% of the target regions. In total, more than 6,000 non-reference variants were identified, including > 1,200 missense and nonsense variants. Using data from the NHLBI's Exome Sequencing Project reference panel, rare functional variants were selected from among these variants and segregation analysis was performed in families contributing evidence of linkage to each of the 6 linked genomic regions. Analyses performed to date have identified 3 rare variants in the proline rich 14-like (*PRR14L*) gene that segregate in DN cases from 4 families with evidence of linkage on chromosome 22. These data suggest that rare variants in *PRR14L* account for the linkage peak identified at this locus and that this gene contributes to variation in urinary albumin excretion in individuals with type 2 diabetes.

769S

Mis-matches between adiposity and metabolic traits: A replicated genomewide association study for metabolic disparity. *L.J. Corbin[1], K. Burrows[1], M. Mangino[2], SM. Ring[3], NJ. Timpson[1].* 1) MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol, United Kingdom; 2) DTR Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 3) School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom.

Whilst increased Body Mass Index (BMI) is correlated with a range of cardio-metabolic risk factors two sub-phenotypes of obesity have been characterised that contradict this. Metabolically obese normal-weight (MNOW) individuals present with a normal BMI but have a poor metabolic profile. In contrast, metabolically healthy obese (MHO) individuals have high BMI, but do not suffer the metabolic disturbances expected. By developing a novel phenotypic outcome acting as an indicator of metabolic disparity, we explore the biological pathways underlying these sub-phenotypes using genomewide association analyses in two independent collections. Analyses were performed in a collection from the Avon Longitudinal Study of Parents and Children (ALSPAC, mean age 15.5yrs) and in a replication cohort of adults from the UK Adult Twin Registry (TwinsUK, mean age 50.9yrs). A compound phenotype was generated by subtracting standardised BMI from standardised fasting plasma glucose to yield a continuously distributed variable where positive values were indicative of "healthier" adiposity status relative to glucose, zero was indicative of matching adiposity/glucose status and negative values were indicative of "healthier" glucose status relative to adiposity. Genomewide analysis of this trait in ALSPAC (n=2584) gave evidence of association at the G6PC2 locus where each additional T allele at the well-known variant rs560887 was related to a -0.25(SE 0.04, p=3.9e-10) change in standardised disparity score. This was replicated in TwinsUK (n=2599)(-0.11(SE 0.04, p=0.003)) and a meta-analysis gave a joint p-value 6.1e-11. This signal was also seen when using DXA derived fat mass in place of BMI (meta-analysis p=4.4e-10) and, as anticipated, with an inverted beta coefficient when replacing glucose status with HOMA_b. This work illustrates how derived phenotypes can be used as indicators of disparity between adiposity and glycaemic traits. Using a standardised disparity score allows a powerful analysis of metabolic features avoiding the costly procedure of defining MONW and MHO by threshold. Results are in line with the carriage of additional T alleles at G6PC2(rs560887) being associated with improved glycaemic profile. Work here did not yield signals of variants showing independent contributions to adiposity; this likely the result of genetic architecture and the causal effect that BMI has on many downstream, causal effects.

770M

Association analysis of exome chip data of Polycystic Ovary Syndrome in Estonian Biobank. *R. Magi[1], A.P. Morris[1,2,3], T. Karaderi[2], T. Laisk-Podar[4], T. Tammiste[4], A. Metspalu[1], A. Salumets[4,5], C.M. Lindgren[2,6].* 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Department of Biostatistics, University of Liverpool, Liverpool, UK; 4) Department of Obstetrics and Gynaecology, University of Tartu, Tartu, Estonia; 5) Institute of Bio- and Translational Medicine, University of Tartu, Tartu, Estonia; 6) Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA.

Polycystic ovary syndrome (PCOS) is a common multifactorial disease affecting up to 10% of women of reproductive age, therefore being their most frequent endocrine problem. It is one of the leading causes of female subfertility as ~40% of women with PCOS are infertile. Genome-wide association studies (GWAS) have revealed several candidate genes, but most of the heritability of PCOS is unexplained. To investigate the contribution of potentially causal coding variants to PCOS, we have genotyped 167 cases and 711 population controls (363 females) from the Estonian Biobank with the Illumina exome array.

We conducted single variant and burden tests of association using SKAT-O within genes for (i) loss of function (LOF) and (ii) rare non-synonomous (NS) variants with minor allele frequency (MAF) <1%). The association analyses were adjusted for first two principal components to account for the population stratification. In the autosomal analysis, both male and female samples were used in the control group but in the X chromosome analysis, only female samples were used.

Altogether 55,345 polymorphic variants were successfully tested in single variant analysis. It revealed one missense variant which was showing exome-wide evidence of association ($p<5\times10^{-7}$, Bonferroni correction for 100,000 variants): exm233350 in the nebulin coding *NEB* gene ($p=4.9 \times 10^{-9}$, MAF=0.05%). Mutations in *NEB* have previously been associated with myopathy and muscle structure. None of the associations were statistically significant in the gene-based tests after multiple testing correction for 20,000 genes ($p<2.5\times10^{-6}$). The strongest associations came from aggregating non-synonymous rare variants within *POLK* ($p=4.3 \times 10^{-5}$) and *PELI3* (7.3×10^{-5}) gene, which are DNA replication and immune response related genes.

Our study suggests that rare variants can contribute to the genetic component of PCOS, but cannot explain previously reported association signals in established GWAS loci.

771T

Large-scale exome chip genotyping reveals novel coding variation associated with endometriosis. *A.P. Morris[1,2,3], R. Mägi[2], N. Rahmioglu[3], A. Mahajan[3], N. Robertson[3], M. Peters[4], M. Saare[4], A. Salumets[4,5], K.T. Zondervan[3], UK Exome Chip Consortium.* 1) Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom; 2) Estonian Genome Centre, University of Tartu, Tartu, Estonia; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 4) Department of Obstetrics and Gynaecology, University of Tartu, Tartu, Estonia; 5) Institute of Bio- and Translational Medicine, University of Tartu, Tartu, Estonia.

Genome-wide association studies (GWAS) have identified nine loci harbouring common variants associated with endometriosis. However, these variants typically map to non-coding genomic regions, and together explain only ~3% of the estimated 52% heritability of the condition. To investigate the contribution of coding variation to endometriosis pathogenesis, we undertook genotyping with the Illumina Exome Chip of two studies of European ancestry: (i) 910 cases from the Oxford Endometriosis Gene (OXEGENE) study and 13,334 population controls (6,828 females) from the UK Exome Chip Consortium; and (ii) 326 cases and 711 population controls (363 females) from the Estonian Biobank. Within each study, we evaluated the association of endometriosis with: (i) individual coding variants; and (ii) burden/over-dispersion of loss of function (all frequencies) and rare non-synonymous (minor allele frequency [MAF] less than 1%) variants within genes using SKAT-O. Analyses were adjusted for principal components to account for population structure. Association summary statistics were combined across studies by meta-analysis (fixed-effect sample size weighted *Z*-score method for single variants and Fisher's method for gene-based tests). We conducted pathway analysis on the basis of single variant meta-analysis summary statistics using the most up to date curated pathway gene-sets from the molecular signatures database as implemented in MAGENTA. No individual coding variants achieved exome-wide significant evidence of association ($p<5\times10^{-7}$, Bonferroni correction for 100,000 variants). The strongest signals include deleterious missense variants in *TAF1L* (D141N, $p=1.5\times10^{-5}$, MAF= 0.077%) and *BMP3* (Y67N, $p=3.2\times10^{-5}$, MAF=2.7%). We observed exome-wide significant evidence of association ($p<2.5\times10^{-6}$, Bonferroni correction for 20,000 genes) with burden/over-dispersion of loss of function variants in *C16orf89* ($p=1.1\times10^{-6)}$ and rare non-synonymous changes in *NECAB3* ($p=1.7\times10^{-7}$), *ZNF485* ($p=1.1\times10^{-6}$), and *RSAD2* ($p=2.1\times10^{-6}$). MAGENTA analyses highlighted potential involvement of cell adhesion/structure, immune function and cancer-related pathways in endometriosis. Our study provides preliminary novel insight into the contribution of coding variation to the genetic component of endometriosis. None of the identified genes from these analyses map to established endometriosis loci, providing no support for the hypothesis that rare coding variation can explain common GWAS association signals.

772S

Association of Rare Variants with Cerebral Palsy by Whole Exome Sequencing. *JJ. Connolly[1], Y. Guo[1], D. Abrams[1], M. Oskui[4], M. Shevell[4], L. Tian[1], H. Hakonarson[1,2,3].* 1) Ctr Applied Genomics, Children's Hosp Philadelphia, Philadelphia, PA, USA; 2) Division of Human Genetics Dept of Pediatrics, Perelman School of Medicine, Philadelphia, PA, USA; 3) Pulmonary Medicine, CHOP, Philadelphia, PA, USA; 4) Departments of Pediatrics and Neurology/Neurosurgery, McGill University, Montreal, Canada.

Despite the fact that cerebral palsy (CP) is the most common motor disability of childhood, with a prevalence of 0.2-0.3% of live births, the basis of the majority of CP disorders is still unknown. Arguably the most widely-researched cause of CP is birth asphyxia, resulting primarily from adverse events during labor. However, several recent studies have suggested that this phenomenon only accounts for ~10% of CP cases. The pediatric biorepository at Center for Applied Genomics (CAG) includes 814 individuals diagnosed with CP. Of these, we selected 65 patients for whom evidence of a genetic cause is strongest (or evidence of environmental causes are weakest). Relevant exclusion factors include premature birth (<37 weeks), complications of pregnancy, complications of labor, and co-morbidity with known Mendelian disorders. Here, we present results from all 65 cases, who underwent whole exome sequencing (WES) at an average coverage of 65X, which was compared against a control set from the 1000 Genomes Project. Due to excessive rare variants in WES results, a burden test was applied. The total number of rare variants across a gene was tabulated in each individual and these totals were compared between cases and controls. Top associations (per SNP-set (Sequence) Kernel Association Test SKAT) included signals detected in *KANK3*, *RNF135*, *ADAP2*, and *SMURF1*.

773M

Whole Genome Sequencing for Discovery of Variants associated with Neuromyelitis Optica. *A. Day-Williams[1], K. Estrada[2,3], F. Zhao[2,3], T. Harris[1], B. Greenberg[4,5], D. MacArthur[2,3], J. Carulli[1].* 1) Computational Biology and Genomics, Translational Sciences and Technology, Biogen Idec, Cambridge, MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, USA; 3) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 4) Department of Neurology and Neurotherapeutics, University of Texas Southwestern Medical Center, Dallas, TX; 5) Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX.

Neuromyelitis optica (NMO) is an autoimmune disorder of the central nervous system that affects the optic nerves and spinal cord. Auto-antibodies against Aquaporin 4 (AQP4) are found in ≈70 percent of NMO patients yet no association to AQP4 has been identified in small genome-wide association studies (GWAS), but associations to the HLA have been reported. We have performed the first whole-genome sequencing analysis of NMO by deep sequencing (>30X) 143 NMO patients and 658 population controls. We have performed joint variant calling of single-nucleotide variants (SNVs) and insertions/deletions (indels) using the GATK HaplotypeCaller yielding 42 million SNVs and 6.3 million indels. In addition, large deletions have been identified in the same cohort using GenomeSTRiP and to further investigate the role of the HLA we have also called classical HLA alleles from the sequencing data using the Omixon method. Quality control of the SNVs and indels includes concordance of non-reference alleles with whole-genome array genotypes, mean depth of coverage, Ti/Tv ratio, call rate, number of alternative alleles, number of singletons, number of indels, allelic balance, heterozygosity and heterozygote/homozygote ratio. We have performed single-variant association analysis using the Efficient Mixed-Model Association eXpedited (EMMAX) software which adjusts for ancestry and relatedness via a variance-component matrix and performs a 2-sided test. Gene-level association analysis was performed using the Sequence Kernel Association Test (SKAT) and collapsing methods to test for an aggregation or burden of rare variants per gene (or group of genes) based on different minor allele frequency and functional characterizations. We will present the results of the association analyses from the first whole-genome sequencing analysis of NMO as well as the HLA analysis and a detailed investigation of AQP4.

774T

Whole exome sequencing in a patient with multiple miscarriages identifies a novel candidate gene. *C. DEMETRIOU[1,3], N. SETO-SALVIA[1], A.C. THOMAS[1], E. CHANUDET[2], H. WILLIAMS[2], R. SHAHNI[1], M. BITNER-GLINDZICZ[1], P. STANIER[1], L. REGAN[3], G.E. MOORE[1].* 1) Genetics and Genomic Medicine, Institute of Child Health - University College London, London, UK; 2) Centre for Translational Genomics-GOSgene, UCL Institute of Child Health, London, UK; 3) Department of Obstetrics and Gynaecology, St. Mary's Campus, Imperial College London, London, UK.

Miscarriage is the most common pregnancy complication with up to 50% of women experiencing one or more early miscarriages (<10 weeks). Recurrent miscarriage (RM) is defined as 3 or more consecutive pregnancy losses and it affects about 1-2% of couples trying to conceive. The etiology of RM has been attributed to genetic defects, uterine anomalies, thrombophilia and endocrine defects.

This study involves a family of Bangladeshi origin in which the female proband has suffered a total of 29 early miscarriages and no successful pregnancies (17 occurred with her first partner, 5 with her second partner and 7 with her current partner). Her mother had 3 successful pregnancies but also 3 late miscarriages, and her two brothers both have two children. The patient was negative for all known RM risk factors tested.

Whole exome sequencing was performed on available samples including the patient, current partner and both brothers. Based on variant changes that were present in the patient but not present in the other family members or 4 Asian female controls, a list of genes was prioritized using the Ingenuity variant analysis program.

Four candidate genes were identified, *FKBP4*, *SERPINB2*, *MIRLET7D* and *CD46*, all with previously reported data linking the genes to infertility or other processes involved in pregnancy. We confirmed the variant changes found in these 4 genes in the family by Sanger sequencing and then sequenced these genes in our control and RM cohorts. These cohorts consisted of 120 White-European and 100 Asian female patients that had RMs and no live births as well as 100 Bangladeshi controls.

Four novel or rare missense variants (not reported in 1000 Genomes, dbSNP or exome variant server) were identified in one of these genes, which were present only in Asian patients but not those of white-European origin. All four were predicted to be damaging by several pathogenicity prediction programs. Functional analysis is being performed to investigate protein localisation and protein-protein interactions using cell transfection with constructs containing the 4 variants and 3 common SNPs (as controls) created by site-directed mutagenesis.

775S

The use of exome sequencing to identify single nucleotide variants associated with necrotizing enterocolitis in premature infants. *J.M. Devaney, A.L. Franklin, M. Said, Z. Tatari-Calderone, S. Vukmanovic, K. Rais-Bahrami, N.L.C. Luban, H. Gordish-Dressman, A.D. Sandler.* Children's National Medical Center, Washington, DC, United States.

Necrotizing enterocolitis (NEC) and spontaneous intestinal perforation are gastrointestinal emergencies in premature neonates. NEC is a common in neonates affecting approximately 11% of premature infants with a mortality rate of from 20% to 40%. Recent studies have suggested that genetic polymorphisms, specifically variants in genes affecting inflammation, may contribute to susceptibility for neonatal bowel inflammation and NEC. We utilized whole exome sequencing (WES) to search for variants that may increase the risk of developing NEC. Buccal swabs were collected for DNA extraction from infants ≤ 32 weeks gestation with and without a diagnosis of NEC. Infants with congenital heart disease, congenital anomalies, and inherited blood/metabolic were excluded. We performed exome sequencing using an Illumina TruSeq Exome Enrichment Kit on an Illumina HiScanSQ system for 54 samples (n = 27 without NEC; n = 27 diagnosed with NEC). We aligned the resulting reads to the hg19 reference genome with BWA, applied GATK base quality score recalibration, indel realignment, duplicate removal, and performed SNP and INDEL discovery. Variants were annotated using ANNOVAR. We analyzed 11 genes (*IL17A*, *IL17B*, *IL17D*, *IL17F*, *NFKB1*, *NFKB2*, *IL1A*, *IL1B*, *IL6*, *PTAFR*, and *TLR4*) for non-synonymous or missense variants generated by WES and selected 26 variants that we tested for an association with development of NEC. None of the variants tested were associated with the development of NEC. However, we uncovered 26 candidate variants in inflammatory genes. None of the discovered variants were associated with the development of NEC. In this limited, candidate gene approach to the analysis of WES data, we did not discover a causal variant for NEC; however, we will be expanding the search for variants to the entire exome.

776M

Whole-exome imputation of sequence variants identified two novel alleles associated with adult body height in African Americans. *M. Du[1,10], P.L. Auer[2], S. Jiao[1], J. Haessler[1], D. Altshuler[3], E. Boerwinkle[4], C.S. Carlson[1], C.L. Carty[1], Y.I. Chen[5], K. Curtis[1], N. Franceschini[6], L. Hsu[1], R. Jackson[7], L.A. Lange[8], G. Lettre[9], K.L. Monda[6], D.A. NIckerson[10], A.P. Reiner[1], S.S. Rich[11], S.A. Rosse[1], J.I. Rotter[3], C.J. Willer[12], J.G. Wilson[13], K. North[6], C. Kooperberg[1], N. Heard-Costa[14], U. Peters[1], National Heart, Lung, and Blood Institute (NHLBI) Go Exome Sequencing Project.* 1) Fred Hutchinson Cancer Research Center, Seattle, WA., USA; 2) University of Wisconsin-Milwaukee Joseph J. Zilber School of Public Health, Milwaukee, WI; 3) Broad Institute, Cambridge, MA; 4) The University of Texas Health Science Center at Houston, Houston, TX; 5) LABioMed at Harbor-UCLA Medical Center, Torrance, CA; 6) University of North Carolina Gillings School of Global Public Health, Chapel Hill, NC; 7) The Ohio State University Wexner Medical Center, Columbus, OH; 8) University of North Carolina School of Medicine, Chapel Hill, NC; 9) Montreal Heart Institute and Université de Montréal, Montreal, Quebec; 10) University of Washington, Seattle, WA; 11) University of Virginia School of Medicine, Charlottesville, VA; 12) University of Michigan Medical School, Ann Arbor, MI; 13) University of Mississippi Medical Center, Jackson, MS; 14) Boston University School of Medicine, Boston, MA.

Adult human body height is a quantitative trait for which genome-wide association studies (GWAS) have identified numerous loci, primarily in European populations. These loci, composed of common variants, explain less than 10 percent of the phenotypic variance in height. Here, we searched for novel associations between height and common (minor allele frequency, MAF≥5 percent) or less frequent (0.5 percent<MAF<5 percent) variants across the exome in African Americans. Using a reference panel of 1,692 African Americans and 471 Europeans from the National Heart, Lung, and Blood Institute's (NHLBI) Exome Sequencing Project (ESP), we imputed whole-exome sequence data into 13,719 African Americans with existing array-based GWAS data (discovery set). Variants achieving a height-association threshold of P<5E-06 in the imputed dataset were followed up in an independent sample of 1,989 African Americans with whole-exome sequence data (replication set). We used a Bonferroni-corrected P<2.5E-07 (=0.05/196,779 whole-exome sequence variants) to define statistically significant associations in meta-analyses combining the discovery and replication sets (N=15,708 participants). We discovered and replicated 3 independent loci for association following Bonferroni-correction: 5p13.3/*C5orf22*/rs17410035 (MAF=0.10, β=0.64 cm, P=8.3E-08), 13q14.2/*SPRYD7*/rs114089985 (MAF=0.03, β=1.46 cm, P=4.8E-10), and 17q23.3/*GH2*/rs2006123 (MAF=0.30; β=0.47 cm; P=4.7E-09). Conditional analyses revealed that 5p13.3 (*C5orf22*/rs17410035) and 13q14.2 (*SPRYD7*/rs114089985) harbored novel height alleles independent of previous GWAS-identified variants in all populations (r2<0.01 with GWAS loci); in contrast, 17q23.3/*GH2*/rs2006123 was correlated with GWAS-identified variants in European and African populations. Notably, 13q14.2/rs114089985 is uncommon in African Americans (MAF=0.03) and is monomorphic in European and Asian populations, suggesting it is an African American-specific height allele. In summary, our findings demonstrate that whole-exome imputation of sequence variants can identify low frequency variants as well as discover novel variants in non-European populations.

777T

Next steps for whole exome sequenced cases: imputing non-coding regions and incorporating whole genome sequenced controls. *A.E. Hendricks[1,2], S.A. McCarthy[2], I.S. Farooqi[2], E. Zeggini[2], I. Barroso[2], UK10K Obesity Group.* 1) Mathematical and Statistical Sciences University of Colorado -- Denver, CO; 2) Wellcome Trust Sanger Instititute, Cambridge, CB10 1HH.

The number of whole-exome sequenced (WES) case, or case-control, sets has recently increased substantially. After researchers complete an initial family or case-control exome based analysis, one might ask what else can be done. Two possible ways to proceed are to impute non-sequenced regions or to use external data sets to increase the number of samples in the control set and subsequently the power to detect association. Here, we do both.

Others have shown that imputation of non-coding regions and association tests of simulated quantitative phenotypes is possible given the <1x off target reads produced by high-depth WES. Here, we investigate the imputation and association of non-coding regions using real data from the UK10K project (www.uk10k.org). Specifically, we use high depth (50x) WES case (N=926) and control (N=1233) samples as well as low depth (6x) whole-genome sequenced (WGS) control samples (N=3621). The case patients are from the Severe Childhood Onset Obesity Project consisting of UK white patients, with an age of onset below 10, and with a body mass index (BMI) Standard Deviation Score (SDS) > 3, a design that is enriched for rare, highly penetrant causes of obesity.

We focus on case-control analysis, which is particularly susceptible to bias, and add WGS controls to an existing WES control set as well as use the WGS separately. Using IMPUTE2, we impute the WES cases and controls genome-wide with both the 1000Genomes only and 1000Genomes plus the UK10K WGS as the imputing backbone. As a true positive, we look for the FTO signal, a signal that is not in high LD with variants in the coding region and was thus notably missing when we performed our case-control analysis on the WES coding regions. We also work to show that, after sufficient quality control and filtering, we attain results without inflated test statistics despite imputation into non-coding regions and the use of WGS controls.

We believe this work is vital to getting the most information and benefit out of WES data. Successful research in this area will enable groups to increase power to replicate signals and find new genetic associations using WES case or case-control data regardless of whether the association is inside or outside of a coding region.

778S

Whole-exome DNA sequencing to find new variants associated with fetal hemoglobin levels. *K.S. Lo[1], G. Lettre[1,2].* 1) Montreal Heart Institute, Montreal, Quebec, Canada; 2) Faculté de Médecine, Université de Montréal, Montreal, Quebec, Canada.

The most promising strategy to treat patients with the commonest Mendelian diseases in the World - sickle cell disease (SCD) and beta-thalassemia - is to increase the endogenous production of fetal hemoglobin (HbF). HbF is a highly heritable trait (h2~0.6-0.8) and genetic studies have identified SNPs at the *BCL11A*, *HBS1L-MYB* and *beta-globin* loci that are associated with its levels. To find new HbF-associated loci, and test the role of rare coding genetic variation, we sequenced the exome of 189 African-American patients with SCD selected from the lower and upper 10% tails of the HbF distribution. We achieved a mean coverage of 76X and identified 164,038 novel variants. Because a subset of these patients (N=164) is also genotyped on the Illumina ExomeChip, we estimated that 67% of the coding genetic variation in this African-American population is missed by the ExomeChip. More specifically, we identified 795 nonsense, 261 splice site and 45,966 missense variants that are not genotyped on the ExomeChip. As expected, these variants are mostly rare in SCD patients (mean minor allele frequency (MAF) =4.6%). Most of the variants polymorphic in the ExomeChip dataset were also identified by whole-exome sequencing (WES) (96%), confirming the high-sensitivity of our experiment. We carried out single variant- and gene-based analyses but did not identify new significant association with HbF level variation. The strongest novel association is in *LINC00908* (P= 2.3×10-5, MAF=5.8%), a long intergenic non-protein gene coding RNA. To increase statistical power, we imputed the coding variants identified by WES into 1279 SCD patients genotyped on the Illumina 610-quad array. We imputed successfully (RSQR>0.5) 906, 5,478 and 28,091 variants with MAF<1%, 1%≤MAF<5% and MAF≥5%, respectively. In testing this imputed dataset for association with HbF levels, we identified a strong signal with several missense variants in olfactory receptor genes on chromosome 11. However, these coding variants are in linkage disequilibrium with the HbF-associated *XmnI* polymorphism at the *beta-globin* locus. No other novel variants reached statistical significance in these analyses. In conclusion, although the initial 3 HbF-associated loci were identified in modest size populations, the remaining common and rare variants associated with HbF have small phenotypic effect sizes and we will need large cohorts to find them.

779M

Whole-Exome Sequencing Identifies Rare, Functional CFH Variants in Families with Macular Degeneration. *J. Seddon[1], Y. Yu[1], M. Triebwasser[2], E. Wong[3], E. Schramm[2], B. Thomas[4], E. Mardis[2], J. Atkinson[2], M. Daly[4], S. Raychaudhuri[5], D. Kavanagh[3].* 1) Tufts Medical Center, Department of Ophthalmology and Ophthalmic Epidemiology and Genetics Service, 800 Washington St. #450, Boston, MA 02111; 2) Washington University School of Medicine, St. Louis, MO; 3) Newcastle University, UK; 4) Mass. General Hospital, Boston, MA; 5) Brigham and Women's Hospital, Boston, MA.

We sequenced the whole exome of 35 cases and 7 controls from 9 age-related macular degeneration (AMD) families in whom known common genetic risk alleles could not explain their high disease burden and/or their early-onset advanced disease. Two families harbored novel rare mutations in CFH (R53C and D90G). R53C segregates perfectly with AMD in 11 cases (heterozygous) and 1 elderly control (reference allele) (LOD = 5.07, P= 6.7×10-7). In an independent cohort, 4 out of 1,676 cases but none of the 745 examined controls or 4300 NHBLI Exome Sequencing Project (ESP) samples carried the R53C mutation (P=0.0039). In another family of 6 siblings, D90G similarly segregated with AMD in 5 cases and 1 control (LOD= 1.22, P=0.009). No other sample in our large cohort or the ESP had this mutation. Functional studies demonstrated that R53C decreased the ability of FH to perform decay accelerating activity. D90G exhibited a decrease in cofactor-mediated inactivation. Both of these changes would lead to a loss of regulatory activity, resulting in excessive alternative pathway activation. This study represents an initial application of the whole-exome strategy to families with early-onset AMD. It successfully identified high impact alleles leading to clearer functional insight into AMD etiopathogenesis.

780T

Whole genome sequencing of 3,514 individuals from the founder population of Sardinia. *C. Sidore[1,2,3], F. Busonero[1,2,4], A. Maschio[1,2,4], M. Zoledziewska[1], A. Mulas[1,3], E. Porcu[1,2,3], G. Pistis[1,2,3], M. Steri[1], F. Danjou[1], A. Kwong[2], C.W. Chiang[7], R. Lyons[4], A. Angius[1,5], H.M. Kang[2], J. Novembre[6], S. Sanna[1], D. Schlessinger[8], F. Cucca[1,3], G. Abecasis[2].* 1) Istituto di Ricerca Genetica e Biomedica, CNR, Monserrato, Cagliari, Italy; 2) Center for Statistical Genetics, Ann Arbor, University of Michigan, MI, USA; 3) Università degli Studi di Sassari, Sassari, Italy; 4) University of Michigan, DNA Sequencing Core, Ann Arbor, MI, USA;; 5) Center for Advanced Studies, Research, and Development in Sardinia (CRS4), AGCT Program, Parco Scientifico e tecnologico della Sardegna, Pula,Italy; 6) Department of Human Genetics, University of Chicago, IL, USA; 7) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA; 8) Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA.

Recently, the combination of genome wide association studies with reference panels based on next generation sequencing has profoundly improved the resolution of genome analysis and the ability to identify causal variants associated with complex traits or diseases. The commonly used sequencing-based reference panels are usually based on diverse populations and do not allow imputation of population-specific variants, particularly those present in isolated populations. Here we substantiate the increased informativeness that is recoverable by further analysis of population-specific variation present in genetically distant populations, and especially in isolated founder populations whose genetic distance from more cosmopolitan groups is most marked. To investigate the Sardinian specific genetic variation we used whole genome sequencing (coverage ~4x) of 3,514 Sardinian individuals, identified >23M single nucleotide polymorphisms (SNPs) and generated a reference panel for imputation. Strikingly, for Sardinian individuals, imputation using our Sardinian reference panel was much improved for low frequency variants (MAF 0.5-5%) compared to imputation based on 1000 Genomes Project haplotypes (increasing r2 with directly measured genotypes from 0.59 to 0.90). We will show how relative isolation affects haplotype length, enrichment of deleterious variants, and genetic differentiation with Europeans. Next, we used this reference panel to study the genetics of LDL-cholesterol in the island by first imputing missing variants in 6,602 individuals from an isolated population in the Lanusei Valley genotyped with >800,000 SNPs. Variants detected by our sequencing explained >90% of the genetic heritability of LDL-cholesterol and point to many interesting association signals. Particularly interesting is the Q39X mutation in the HBB gene, common in Sardinia (MAF>5%) and very rare in Europe, not previously identified by any GWAS analysis. The variant results in a decrease of 14.4 mg/dl in blood among carriers (p < 10-21). Our results illustrate the benefits of large-scale sequencing efforts in founder populations and their ability to uncover functionally relevant variants that may be very rare elsewhere.

781S

Insights into the genetic architecture of anthropometric traits using whole genome sequence data. *I. Tachmazidou[1], G. Ritchie[1,2], J. Min[3], K. Walter[1], J. Huang[1], J. Perry[4], T. Keane[1], S. McCarthy[1], Y. Memari[1], UK10K consortium.* 1) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, United Kingdom; 3) MRC Integrative Epidemiology Unit, University of Bristol; 4) MRC Epidemiology Unit, University of Cambridge.

Body weight and fat distribution measures are associated with increased risk of cardiometabolic disease. As part of the UK10K study, we have investigated the genetic architecture of 12 anthropometric traits in 3,538 individuals with ~7x whole genome sequence (WGS) data from the ALSPAC and TwinsUK cohorts. Variants discovered through WGS, along with those from the 1000 Genomes Project (1KG), were imputed into additional individuals from the ALSPAC and TwinsUK cohorts with GWAS data, increasing the total sample size to 11,178. We investigated association between anthropometric traits and ~9 million variants with MAF≥0.01 and ~5 million variants with MAF 0.001-0.01. *In silico* replication was sought in 16 external cohorts for a total sample size of 15,000-40,000 depending on trait. To determine novelty of signals, we performed conditional analyses on all established variants within a 1MB window. We observe a significant excess of independent (r2<0.2) previously not reported variants with MAF>0.01 and p<10-5 in the UK10K discovery set in all anthropometric traits. We find significant enrichment of variants associated with BMI in UK10K and established monogenic obesity genes (binomial p=4.8×10-5). Further replication is ongoing, but interim analyses identify replicating signals, for example, variant chr5:105105444 (EAF 0.0084; UK10K p=4.69×10-5; replication p=2.53×10-4; overall p=5.53×10-8, beta -0.37, SE 0.07, sample size=27,687) is a novel signal associated with waist circumference adjusted for BMI, a measure of central adiposity. Waist to hip ratio is associated with variants at chr9:23016057 (EAF 0.003; UK10K p=6.11×10-5; replication p=2.92×10-04; overall p=5.98×10-8, beta 0.55, SE 0.10, sample size=25,373) and chr14:100931682 (EAF 0.001; UK10K p=5.18×10-2; replication p=2.88×10-07; overall p=8.11×10-8, beta 0.97, SE 0.18, sample size=18,212). These replicating signals are at variants with MAF<0.01, have modest effect sizes and are not present in HapMap. Imputation based on the combined UK10K and 1KG panel has helped identify these loci. Larger sample sizes are required for the identification and replication of further rare variant associations with anthropometric traits.

782M

Next generation association studies in isolated populations. *E. Zeggini[1], L. Southam[1,2], K. Panoutsopoulou[1], K. Hatzikotoulas[1], G.R.S. Ritchie[1], A-E. Farmaki[3], I. Tachmazidou[1], A. Matchan[1], N.W. Rayner[1,2], J. Schwartzentruber[1], I. Ntalla[3], E. Tsafantakis[4], M. Karaleftheri[5], G. Dedoussis[3], A. Gilly[1].* 1) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 3) Harokopio University Athens, Athens, Greece; 4) Anogia Medical Centre, Anogia, Greece; 5) Echinos Medical Centre, Echinos, Greece.

Isolated populations have unique characteristics that can be leveraged to increase power in genetic association studies. In founder populations genetic drift can drive trait-associated alleles to higher frequency and thus enable the identification of rare variant associations with smaller discovery sets. We have collected samples from two isolated populations in Greece (HELlenic Isolated Cohorts study): the Pomak villages (HELIC-Pomak) in the North of Greece; and the Mylopotamos villages (HELIC-MANOLIS) on Crete. All samples (n~3000) have information on a wide array of anthropometric, cardiometabolic, biochemical, haematological and diet-related traits, genotypes from the Illumina OmniExpress and exome-chip platforms, and are being whole-genome sequenced at low depth. Using 1x WGS data from 995 (HELIC-MANOLIS) individuals, we demonstrate that over 80% of true low-frequency (0.01<MAF<0.05) variants are found, compared to an average 60% for 0.001<MAF<0.01 and 40% for MAF<0.001. Genotype concordance reaches >95% and minor allele concordance >90% across the whole MAF spectrum. We replicate known association hits, thereby providing a proof of concept for a robust processing pipeline for low-depth WGS variant calls. Using genotype data, we find that 80% of subjects have at least one "surrogate parent" in the isolates, compared to 1% in the outbred Greek population. In the MANOLIS cohort we observe an enrichment of missense variants amongst the variants that have drifted up in frequency by >5 fold. We have previously reported a lipid traits association with a functional variant in the APOC3 gene in 1267 individuals in MANOLIS. The equivalent sample size needed to detect this in the general European population would be 67,000. In the Pomak cohort we find novel associations at variants on chr11p15.4 showing large allele frequency increases (from 0.2% in the general Greek population to 4.6% in the isolate) with haematological traits, for example with mean corpuscular volume (at rs11035019, beta=-1.249, p= 3.45×10^{-29}). Their detection in cosmopolitan populations would necessitate thirteen times as many samples. We demonstrate the significant power gains that can be afforded by studying well-characterised founder populations.

783T

When nature meets science: longevity blueprint. *D. Ben-Avraham[1], S. Carmi[2], Y. Freudenberg-Hua[3,4], J. Freudenberg[5], K.Y. Hui[6], E. Kochav[2], X. Liu[7], J. Xue[2], F. Grady[2], S. Guha[8,9,10], A. Darvasi[11], K. Offit[12], L.J. Ozelius[10], I. Peter[10], J.H. Cho[6], H. Ostrer[13,14], L.N. Clark[7,15], T. Lencz[8,9], N. Barzilai[1], P. Davies[3], I. Pe'er[2,16], G. Atzmon[1].* 1) Institute for Aging Research Departments of Medicine and Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461; 2) Department of Computer Science, Columbia University, 500 W 120th St, New York, NY, 10027; 3) The Litwin-Zucker Research Center for the Study of Alzheimer's Disease and Memory Disorders, The Feinstein Institute for Medical Research, North Shore-LIJ, Manhasset, NY. 11030; 4) Division of Geriatric Psychiatry, Zucker Hillside Hospital, North Shore-LIJ, Glen Oaks, NY 11040; 5) Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, North Shore-LIJ, Manhasset, NY. 11030; 6) Department of Internal Medicine, Genetics & Pediatrics, Yale School of Medicine, 300 Cedar St, New Haven, CT, 06519; 7) Department of Pathology and Cell Biology, Columbia University Medical Center, 1150 St Nicholas Ave, New York, NY, 10032; 8) Center for Psychiatric Neuroscience, The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, NY, 11030; 9) Department of Psychiatry, Division of Research, The Zucker Hillside Hospital Division of the North Shore-Long Island Jewish Health System, Glen Oaks, NY, 11004; 10) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, One Gustave L. Levy Pl, New York, NY, 10029; 11) Department of Genetics, The Institute of Life Sciences, The Hebrew University of Jerusalem, Givat Ram, Jerusalem, Israel, 91904; 12) Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, 1275 York Ave, New York, NY, 10065; 13) Department of Genetics, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY, 10461; 14) Department of Pathology, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY, 10461; 15) Taub Institute for Research of Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, 1150 St Nicholas Ave, New York, NY, 10032; 16) Center for Computational Biology and Bioinformatics, Columbia University, 1130 St Nicholas Ave, New York, NY, 10032.

Longevity has emerged as a proof of concept for a healthy lifespan. While life expectancy almost doubled during the last century, suggesting tremendous progress in medical treatment and life style improvement, the outcome (i.e. living longer) had resulted with increase prevalence of age-associated diseases, complex morbidities and frailty. Understanding the physiology of aging had placed its mark on a pallet of treatments resulting in partially enhancement of quality of life, setting the stage for genetic studies. However, a decade of research and numerous efforts attempting to decipher the genetic component of longevity have resulted in limited success. Here, we report of an effort to detect the genetic signature of longevity by using the most advance technology (Whole genome sequencing-WGS) as well as a population well fitted for such a genetic study (Ashkenazi Jews-AJ). We obtained whole genome sequences of 148 AJs (74 centenarian (By Illumina) and 74 elderly unrelated controls (By Complete Genomics-CGI)). Following structured filtration pipeline, we merged the two data sets guided by 5 samples that were sequenced on both platforms. Our genome-wide analysis identified ~6M SNVs that were concordant between both platforms and were further used for merging. We also applied additional quality control filters, resulting in filtering out 30% of the selected SNVs and leaving ~4.2M SNVs that were used for the case control analysis in the entire cohort. We evaluated the association with longevity using linear regression model of the ~4.2M SNVs in parallel (PLINK). Variants in 4 loci (rs78563403, rs67505584, rs8010307and rs59010232) reached p<10-7. Those loci were validated in imputed Human OmniExpress genotyping in 639 AJ cases and 473 AJ control. We further applied meta-analysis of the two datasets, revealing 2 segments on chromosomes 1 (rs1765725, rs1698975, rs1765724) and 13 (rs2182054, rs7319950, rs9550998) and 2 SNPs on chromosome12 (rs2710278, rs7961171) with p<10-7. This study highlights the utility of the Ashkenazi Jewish population in genetic research, where a small sample size was required to detect significant variant association in a complex polygenic trait.

784S

Quantitative trail loci for plasma proteins in current and former smokers with and without chronic obstructive pulmonary disease (COPD). *R.P. Bowler[1], W. O'Neal[2], K. Kechris[3], W. Sun[2], S. Jacobson[1], T.H. Chen[2], M.B. Drummond[4], R.E. Kanner[5], P.G. Woodruff[6], D. Couper[2], G.A. Hawkins[7], COPDGene and SPIROMICS.* 1) Department of Medicine, National Jewish Health, Denver, CO; 2) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Biostatistics and Informatics, School of Public Health, University of Colorado Denver, CO; 4) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 5) Department of Medicine, University of Utah School of Medicine, Salt Lake City, Utah; 6) 5Department of Medicine and Cardiovascular Research Institute, University of California San Francisco, CA; 7) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston Salem, NC.

Background: Chronic obstructive pulmonary disease (COPD) occurs most commonly in current and former smokers and is the third leading cause of death in the United States. Genetic association studies have identified multiple genetic loci associated with COPD phenotypes such as airflow obstruction and emphysema. Similarly protein biomarker studies have identified blood biomarkers associated with these COPD phenotypes. The relationship between genetic markers and protein levels has not been explored in at risk subjects. **Methods**: Using two similar cohorts of older smokers (COPDGene and SPIROMICS), we explored the relationship between genome wide genetic markers and 96 candidate blood biomarkers for COPD in 750 (SPIROMICS) and 602 (COPDGene) non-Hispanic Whites. Genotyping was performed on Illumina platforms. Candidate COPD blood biomarkers were assessed in fresh frozen serum and plasma using 13 Myriad-RBM single and multiplex panels. *cis* and *trans* protein quantitative trait loci (pQTL) were identified using an additive regression model adjusted for covariates (age, gender, BMI, smoking status, pack-years), as well as genetic principal components, and correction for multiple comparisons. The Stouffer-Liptak test was used for the meta-analysis to combine the p-values from the two studies. **Results**: We identified 539 significant pQTL SNPs (69% *cis*; 31% *trans*) in 36 proteins. The top *cis*- association was for GC (rs7041; Vitamin D binding protein; P= 5 X 10^{-394}). This SNP leads to an Asp to Glu amino acid change at position 432 of the protein and is associated with more severe COPD. The top trans association was rs507666 (chromosome 9) for SELE (chromosome 1; E - selectin; P = 7 X 10^{-103}). This SNP is an intron variant in ABO (alpha 1-3-N-acetylgalactosaminyltransferase) and has also been associated with blood group and altered glycosyltransferase activities. Other notable pQTLs include rs2070600, a non-synonymous *cis*-SNP (P = 4 X 10^{-22}) located in exon 3 of AGER (advanced glycation end product receptor), which leads to a glycine to serine amino acid change at position 82 and is associated with emphysema. **Conclusions**: This is the largest integration of genetics and protein biomarkers in two highly phenotyped cohorts. It demonstrates how genetic variation influences intermediate phenotypes of disease (e.g. protein biomarkers) and may allow us to begin to unravel the mechanisms through which genotype influences disease risk.

785M

Low-frequency coding variants associated with female reproductive ageing. *K.S. Ruth, ReproGen consortium (www.reprogen.org).* Genetics of Complex Traits, University of Exeter Medical School, Exeter, United Kingdom.

Background: Genome-wide association studies (GWAS) have identified over 100 loci associated with age at menarche and 18 for age at natural menopause. The majority of these variants are common (minor allele frequency (MAF) > 5%), have small effect sizes, and in aggregate explain a small fraction of the population variation in these reproductive ageing traits. It is predicted that functional genetic variants with lower MAF, but larger effects on phenotype, may contribute substantially to the genetic variance underlying these traits.

Aim: To identify low-frequency coding variants contributing to the genetic variation in age at menarche and age at natural menopause.

Methods: Exome chip genotyping data were collected from 17 studies of European populations. Single marker and burden testing meta-analyses were carried out in 67,628 women for age at menarche and 35,605 women for age at natural menopause.

Results: New low frequency signals were identified for both age at menarche and age at natural menopause. For age at menarche, a nonsynonymous variant with MAF 1.1% was identified in *ALMS1* (2p13) ($p=4.95\times10^{-9}$), which increased age at menarche by approximately 3 months per allele. Burden testing identified the gene *PRKAG1* (12q13.12, containing 4 variants with MAF <3.5%) as associated with age at menarche ($p=2.2\times10^{-6}$). For age at natural menopause, two low frequency nonsynonymous variants were identified in *HELB* (12q14.3) (MAF 2.5% $p=3.87\times10^{-28}$ and MAF 3.5% $p=1.37\times10^{-29}$); these variants are 49 bp apart and are in linkage disequilibrium ($r^2=0.7$), each increasing age at natural menopause by approximately 1 year. Two-directional burden testing identified the gene *SLCO4A1* (20q13.33, containing 18 variants) as associated with age at natural menopause ($p=4.54\times10^{-7}$).

Conclusions: We report the first large-scale exome chip genotyping project for reproductive ageing. Two genes containing low frequency or rare variants were associated with age at menarche and a further two genes were associated with age at natural menopause. None of these genes have been previously implicated in genetic associations for reproductive traits in humans and the effect sizes were large compared to GWAS findings.

786T

Redefining the contiguous gene syndrome in the era of high-throughput sequencing. *L. COLLEAUX[1], M. LANGOUET[1], K. SIQUIER-PERNET[1], S. SANQUER[3], C. BOLE-FEYSOT[4], P. NITSCHKE[5], A. MUNNICH[1], R. BAROUKI[3], J. AMIEL[1,2].* 1) Imagine Institute, Necker-enfants malades hospital, Paris, France; 2) Genetic Service, Necker-enfants malades hospital, AP-HP, Paris, France; 3) Metabolic and proteomic biochemistry service, Necker-enfants malades hospital, AP-HP, Paris, France; 4) Genomic platform, Imagine Institute, Necker-enfants malades hospital, Paris, France; 5) Bioinformatic platform, Imagine Institute, Necker-enfants malades hospital, Paris, France.

We ascertained an Algerian consanguineous family in which two sibs presented with psychomotor delay, progressive microcephaly, spasticity, thin corpus callosum, and severe and early onset obesity. Exome sequencing identified two homozygous substitutions cosegregating with the phenotype and locating 170 kb apart on 7q22.1: a c.1137+1G>T splice mutation in *AP4M1* previously described in a Moroccan family and a c.595A>T misense variation in *AZGP1* which encodes zinc-alpha2-glycoprotein (ZAG). Haplotyping analysis indicated that the AP4M1 mutation was a founder mutation shared between both families, whereas the *AZGP1* mutation occurs secondarily and is unique in our family. Mutations in *AP4M1* cause AP4-deficiency syndrome, a condition characterized by severe intellectual disability, progressive microcephaly and spasticity. Notably, none of the 25 previously reported cases with AP4-deficiency syndrome exhibited obesity. On the other hand, ZAG is an adipokine stimulating lipolysis in adipocytes; ZAG likely regulates body weight since administration of human ZAG to ob/ob mice resulted in progressive weight loss. We propose that the phenotype of our patients resulted from the additional effects of the two mutations in *AP4M1* and *AZGP1* accounting for the neurological signs and the precocious morbid obesity, respectively. The contiguous gene syndrome was proposed in 1986 to explain the association of multiple and unrelated clinical features due to the deletion of multiple adjacent genes: the phenotype results from the combination of the endophenotypes of each contiguous gene sensitive to haploinsufficiency. Today, high-throughput sequencing allows us to enlarge this concept to describe simultaneous transmission of independent mutations that are genetically linked.

787S

Sequencing of genes expressed in podocytes uncovers FSGS risk alleles in European-American population. *M. Artomov[1,2,3], H. Yu[4], C. Winkler[6], J. Kopp[5], A.S. Shaw[4], M.J. Daly[2,3].* 1) Chemistry and Chemical Biology Dept, Harvard University, Cambridge, MA; 2) Broad Institute, Cambridge, MA; 3) Analytic and Translation Genetics Unit, Massachusetts General Hospital, Boston, MA; 4) Department of Pathology and Immunology, HHMI, Washington University School of Medicine, St. Louis, MO; 5) Kidney Disease Section, NIDDK, National Institutes of Health, Bethesda, MD; 6) Molecular Genetic Epidemiology Studies Section, NCI, Frederick, MD.

Focal segmental glomerulosclerosis (FSGS [MIM 607832, 603965, 613237, 614131, 603278, 612551]) is one of the major causes of nephrotic syndrome in adults and children. Over the last 10 years, various genetic approaches have identified at least eight genes: ARHGAP24, TRPC6, MYH9, INF2, APOL1, NPHS2, ACTN4, and CD2AP as FSGS susceptibility genes. Since mutations in these genes explain only a small fraction of FSGS cases, it seems very likely that other genes await discovery. The FSGS phenotype is known to be specific to podocytes, thus we focused only on sequencing 2500 genes that are highly expressed in the podocyte. Here we present our analysis of 214 European FSGS cases aiming to search for novel germline mutations in Europeans predisposing to the disease. We used expression studies in both human and mouse to assemble a list of approximately 2500 genes representing about 7 Mb and sequenced these genes in about 200 patients and we have used 378 full exome sequences of European ancestry controls for the case/control study. Most of the genes so far reported to be associated with FSGS have been found in studies involving individuals of African-American descent. Here we focused on genetic factors driving disease in European individuals. We found damaging rare mutations in several genes that have previously been reported as risk genes in Africans. We then performed unique to cases mutations anlaysis and found that top hits are ARHGAP24, which is a known risk gene in Africans, APOL3, which is in a close linkage with known APOL1 risk gene and KANK1 which is associated with renal cell carcinoma, MPRIP, which directly interacts with MYH9 at the protein level and potentially could be a new risk gene for FSGS. We have used allelic association test to search for associated variants. Our two top hits were the known FSGS allele - G1 APOL1 variant and a stop codon mutation in APOL3. Interestingly, we saw a perfect co-inheritance of these SNPs within our cases. Our results suggest that the genetic nature of FSGS in Europeans is similar to Africans and mostly is driven by the same alleles and risk genes. We find several new genes that seem to increase risk of FSGS in Europeans and were not previously reported.

788M

Evaluating rare variants in *ZNF469* and other GWAS identified candidate genes in keratoconus. *K.P. Burdon[1, 2], S.E.M. Lucas[1], R.A. Mills[2], E. Souzeau[2], J.E. Craig[2].* 1) Menzies Research Institute Tasmania, University of Tasmania, Hobart, TAS, Australia; 2) Department of Ophthalmology, Flinders University, Adelaide, SA, Australia.

Keratoconus [OMIM148300] is a complex disease characterised by progressive thinning and conical protrusion of the cornea, resulting in major visual impairment. Rare mutations in *VSX1* and *MIR184* have been reported. Recent genome-wide association studies have implicated genetic variation near the *HGF* and *RAB3GAP1* genes and GWAS for central corneal thickness followed by analysis in multiple keratoconus cohorts have highlighted *FOXO1* and *ZNF469* as candidate genes. The role of *ZNF469* is controversial with the risk allele for keratoconus being associated with thicker corneas. To evaluate coding variants in these genes in keratoconus, following ethics approval and informed consent of the participants, we undertook exome sequencing in 51 patients with severe keratoconus and 51 examined normal controls with the SureSelect V4 exome capture (Agilent), sequenced on HiSeq2000 (Illumina). Reads were aligned to human genome hg19, variants called with Samtools and annotated with Annovar. The rate of rare (MAF<0.01 in 1000genomes) and novel (not present in dbSNP135) nonsynonymous variants was compared between cases and controls. The potential functionality of variants was assessed with SIFT. Variants are undergoing validation by Sanger sequencing. In *RAB3GAP1* (NM_012233), 3 variants were found in cases; p.Q860H, p.I305V (rs116775947) and p.R336C (rs150478342) and 2 in controls; p.N718S and p.R530C. All five are predicted to be tolerated. In *HGF* (NM_000601) p.R328H was detected in 1 case and p.A46V (rs150267054) in 2 cases. Both are predicted to be tolerated. In *FOXO1* (NM_002015), 1 variant, p.S390L, was detected in a single case. This variant is predicted to be deleterious by SIFT and may play a role in disease susceptibility. In *ZNF469* (NM_001127464), 9 novel and 7 rare nonsynonymous variants were detected in 18 cases including 9 deleterious in 12 cases. In controls 15 novel and 5 rare variants were detected in 20 samples including 11 controls with deleterious or deletion mutations. The proportion of samples with novel or rare nonsynonymous variants was not different between cases and controls (χ^2 p=0.838) nor was the rate of predicted deleterious nonsynonymous variants (p=0.810). Despite strong association signals at these loci, rare coding variants in these genes do not account for disease, although a larger cohort may be required to test *FOXO1*. *ZNF469* is a highly variable gene and the burden of rare variants does not appear to contribute to keratoconus.

789T

Whole genome sequencing of two trios identifies mutations in *ADNP2* and *ZNFP2*, candidate genetic modifiers for 22q11DS cognitive and cardiac phenotypes. *J. Chung[1], J. Cai[2], B.G. Suskin[1,3], K. Coleman[4], B.E. Morrow[1].* 1) Genetics, Albert Einstein College of Medicine, Bronx, NY; 2) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Obstetrics & Gynecology and Women's Health, Montefiore Medical Center, Bronx, NY; 4) Childrens Healthcare of Atlanta at Egleston, Atlanta, GA.

The 22q11.2 Deletion Syndrome (22q11DS) is characterized by highly variable expressivity, ranging from mild craniofacial defects to severe congenital heart disease and/or psychiatric illness. Genetic modifiers may account for variable expressivity through a multi-hit model where the 22q11.2 deletion acts as the primary hit and genetic variants act as second hits to modify the phenotype. To identify candidate genetic modifiers for physical and behavioral phenotypes, we used whole genome sequencing (WGS) of two trios with normal parents and probands (C1 and C2) with 22q11DS, who have similar *de novo* 3 Mb deletions but discordant phenotypes. On the remaining 22q11.2 allele in each case, we found 30 coding variants in each case and 1,038 and 940 noncoding variants in C1 and C2 respectively. For C1, we found 349,978 rare (MAF < 0.05), 57,319 novel, and 48 *de novo* variants elsewhere in the genome. For C2, we found 371,778 rare, 61,367 novel, and 35 *de novo* variants. C1, a male, recruited at 14 years old, had tetralogy of Fallot (TOF), cognitive deficits, and attention deficit hyperactivity disorder. He was later diagnosed with schizophrenia (SCZD) at age 28. C2, a female, age 14, had juvenile rheumatoid arthritis that has resolved, no heart anomalies, and no major cognitive deficits or psychiatric illness. Our first goal was to determine if there was an enrichment of potentially pathogenic DNA variants in candidate genes for TOF or cognitive/SCZD in C1 vs C2. We identified 62 genes in C1 and 65 genes in C2 with potentially pathogenic novel, rare compound heterozygous, or *de novo* variants using ANNOVAR. We used ToppGene to prioritize these target genes against known TOF and cognitive/SCZD genes from OMIM. In C1, we did not identify an enrichment of candidate genes for TOF, but we did find enrichment of candidate genes for cognitive/SCZD, compared to C2 (Fisher's p-value = 9.72E-05). Among the interesting variants in C1, were common variants in *COMT* (rs4680) and *PRODH* (rs450046) on the remaining 22q11.2 allele, and a heterozygous *de novo* mutation in *ADNP2* (c.2243G>C), encoding a protein that may be neuroprotective. In C2, we identified a novel nonsynonymous mutation in *ZFPM2* (*FOG2*; c.1576C>T), a known TOF disease gene, which may act as a protective variant when combined with haploinsufficiency of *TBX1* on 22q11.2. Overall, this analysis reveals the utility and challenges of WGS for identifying potential modifier genes (*ZFPM2* and *ADNP2*) for this complex disorder.

790S

Whole exome sequencing of Cold Medicine-Related Stevens-Johnson Syndrome/Toxic Epidermal Necrolysis (CM-SJS/TEN) with Severe Mucosal Involvement. *Y. Hitomi[1], S. Khor[1], M. Ueta[2,3], C. Sotozono[2], S. Kinoshita[2], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan; 3) Research Center for Inflammation and Regenerative Medicine, Faculty of Life and Medical Sciences, Doshisha University, Kyoto, Japan.

Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are acute inflammatory vesiculobullous reactions of the skin and mucous membranes, and these reactions are reported to be caused by inciting drugs, viral infections, or malignant tumor. Although the occurance of SJS/TEN is rare at about 1-6 cases per million, mortality rates are higher than other drug rush (SJS: 3%, TEN: 27%). In addition, Quality-of-life (QOL) of the most survivors is often low, as severe ocular surface complications are often developed as the aftereffect. Over 1,000 drugs have been reported as the so-called "causal medicine" of SJS/TEN, and susceptibility to SJS/TEN caused by some of them the medicines have been reported to be associated with *HLA* genes. As examples, *HLA-A*31:01* and *HLA-B*15:02* have been reported to be associated with carbamazepine-induced severe cutaneous adverse reactions (SCARs). Besides, *HLA-B*58:01* have been reported as susceptible allele of Allopurinol related SJS/TEN. Recently, *HLA-A* and other genes have been reported to be susceptible for Cold Medicine-Related Stevens-Johnson Syndrome/Toxic Epidermal Necrolysis (CM-SJS/TEN) in Asian populations by genome-wide association study (GWAS). However, remaining genetic susceptibility of CM-SJS/TEN including rare variants and structural variants remains to be discovered. To identify the functional variants for CM-SJS/TEN with severe mucosal involvement, whole-exome sequencing was performed in about 50 Japanese CM-SJS/TEN patients with severe mucosal involvement using next generation sequencer (NGS) - ion Proton (Thermo-Fisher Scientific). Whole-exome sequencing data were then analyzed using CLC Genomics Workbench (CLC bio) software. Here we will show the provisional results of Whole-exome sequencing for CM-SJS/TEN. This study illustrates novel diagnostic and therapeutic methods for CM-SJS/TEN. (Hitomi Y. and Khor S. are contributed equally to this work.).

791M

Identification of putatively causative variants in three anorexia nervosa multiplex families by whole exome sequencing. *D. Li[1], R.D. Cone[2], B. Li[3], H. Hakonarson[1,4,5].* 1) The Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 2) Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, TN 37270, USA; 3) Center for Quantitative Sciences, Vanderbilt University, Nashville, TN 37270, USA; 4) Division of Human Genetics Department of Pediatrics, The Children's Hospital of Philadelphia and The Perelman School of Medicine, Philadelphia, PA 19104, USA; 5) Division of Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

Anorexia nervosa (AN) is a perplexing illness characterized by low body weight and persistent fear of weight gain during period of growth, resulting in extreme emaciation. Family studies have consistently demonstrated that AN occurs in families and twin studies have revealed the contribution of additive genetic factors to the observed familial aggregation, but so far just few genetic factors have been found to be specific to AN and no single gene has been shown to be necessary or sufficient to express the phenotype. While GWAS relies on proxy association of genetic variants with unknown disease causality, in which it only focuses on one affected child from multiplex families, the published studies don't specifically analyze inherited mutations despite the fact that AN is highly heritable. Familial AN with young age of onset, extremely low body weight, extended course of illness or male affected constitutes a unique and extreme form of the disease. Here we report the discovery of 5 ultra rare (MAF=0.000077 in ESP6500SI) or novel variants in *COMT, DRD4, KCNN3, GRPR and ESR1*, in 3 AN multiplex kindreds as strong biological candidates. Polymorphisms in these genes have been associated with several psychological disorders, including but not limited to AN, Attention Deficit Hyperactivity Disorder (ADHD), schizophrenia, depression and bipolar disorder. We are currently recruiting additional family members to further validate the cosegregation pattern and conducting follow-up functional assessment to elucidate the roles these genes play in AN and potentially other eating disorders. The resulting variants will be presented together with detailed phenotypic characterization of the families examined.

792T

Whole exome sequencing implicates novel rare genetic variants in susceptibility to Legionella pneumonia. *A. Ndungu[1], KS. Elliott[1], TC. Mills[1], A. Rautanen[1], P. Hutton[2], C. Garrard[3], AC. Gordon[4], CJ. Hinds[5], AVS. Hill[1], SJ. Chapman[1,6], GAinS Investigators.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) John Radcliffe Hospital, Oxford, UK; 3) Intensive Care Unit, John Radcliffe Hospital, Oxford, UK; 4) Section of Anaesthetics, Pain Medicine & Intensive Care, Imperial College, London, UK; 5) William Harvey Research Institute, Barts & The London School of Medicine, London UK; 6) Respiratory Medicine, Churchill Hospital Site, Oxford Radcliffe Hospital, Oxford, UK.

Community acquired pneumonia (CAP) is associated with considerable mortality and morbidity worldwide. 2-9 % of all CAP cases are attributable to the flagellated bacteria *Legionella pneumophila*. Exposure through contaminated water supplies may be relatively common, yet there is great interindividual variation in susceptibility to disease. Host genetic variation is increasingly recognised as playing an important role in determining susceptibility to infectious agents, however, the role of host genetics in susceptibility to Legionella pneumonia has not been fully characterised. Few candidate gene studies have mainly focused on genes in the Toll-like receptor pathway of the innate immune system. The role of these genes in pathogenesis has further been studied in mice with *NAIP, TLR4, 5* and *6* being implicated in susceptibility to Legionella. We use a whole exome sequencing approach in an attempt to identify novel genes and large effect genetic variants associating with Legionnaires' disease in humans. We performed whole exome sequencing of 16 Caucasian individuals from the UK presenting to hospital with microbiologically proven CAP due to *Legionella pneumophila*. Individuals were recruited as part of the Genetics of Sepsis and Septic Shock (GenOSept) study. We obtained whole genome sequence data of 1,927 healthy individuals from the UK10K ALSPAC cohort for use as controls. Over 63,000 mutations were identified as missense, nonsense, splice site variants or variants occurring in untranslated regions. We performed Fisher's exact test to compare frequencies of non-synonymous variants between cases and controls. Two rare single nucleotide variants in *LAMA5* and *SLC22A18* genes were identified with suggestive evidence of association at $P = 1.43 \times 10^{-5}$ and $P = 5.40 \times 10^{-5}$ respectively. We further performed gene-based testing by collapsing variants with allele frequency less than 5% predicted to be deleterious to protein function. Two genes *FAM71E2* and *RFX8* had suggestive evidence of association at $P = 2.99 \times 10^{-5}$ and $P = 5.38 \times 10^{-5}$ respectively. Plans to sequence additional samples are under way. Legionella infection has a mortality rate of between 16-30% if untreated. A comprehensive understanding of genetic predisposition is greatly needed for better prevention, diagnostic and treatment strategies.

793S

Exome sequencing of multiplex pedigrees for the identification of novel rare susceptibility variants for CD. *B.-S. Petersen[1], S. Nikolaus[2], S. Schreiber[2], A. Franke[1].* 1) Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Germany; 2) Department of Internal Medicine, University Hospital Schleswig Holstein, Kiel, Germany.

Crohn's disease (CD) is a complex, chronic inflammatory bowel disease (IBD). A variety of genetic and environmental factors likely play a role in causing CD. Genome-wide association studies (GWAS) and meta-analyses have so far identified 163 genetic susceptibility loci for IBD. These, however, explain less than 30% of the heritability so far. Apart from the mostly common variants included in classical GWAS, further sources of heritability include rare and novel variants with possibly higher penetrance which can be identified through systematic resequencing studies. We have therefore carried out exome sequencing for 50 affected individuals from 17 pedigrees with at least three individuals affected by CD for finding rare high penetrance variants and novel candidate genes involved in disease etiology. The results were analyzed pedigree-wise and filtered for rare and novel coding variants shared by the patients. In several cases this led to the identification of candidates likely to play an important role for disease development in the underlying pedigree. A total of 300 variants were selected and are currently being genotyped in a large patient control cohort to test for recurrence and for proving that the identified variants are not common in healthy individuals. Our preliminary results have revealed novel rare candidate variants in known and novel CD genes shared by the affected individuals of a pedigree which may have an important impact on the development of the disease, also in other patients.

794M

Mutations in LRP2 and NUP205 in Patients with Non-Syndromic Autosomal Recessive Intellectual Disability Using Exome Sequencing. *N. Vasli[1], I. Ahmed[1,2], K. Mittal[1], M. Ohadi[1], A. Mikhailov[1], MA. Rafiq[1], A. Bhatti[2], M. Carter[3], DM. Andrade[4,5], M. Ayub[6], P. John[2], JB. Vincent[1,7,8].* 1) Molecular Neuropsychiatry and Development Lab, The Campbell Family Brain Research Institute, The Centre for Addiction & Mental Health (CAMH), Toronto, Ontario, Canada; 2) Atta-ur-Rehman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad-Pakistan; 3) Department of Pediatrics, Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto ON Canada; 4) Division of Neurology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada; 5) Krembil Neuroscience Centre, Toronto Western Research Institute, Toronto, Canada; 6) Division of Developmental Disabilities, Dept. of Psychiatry, Queen's University, Kingston, ON, Canada; 7) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 8) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada.

Intellectual disability (ID) is a genetically heterogeneous disorder with more than 50 mutated genes to date. ID is characterized by deficits in memory skills and language development with difficulty in learning and problem solving. It affects ~2% of population with difference in severity based on the Intelligence quotient (IQ) scores from mild forms with IQ range from 50 to 70, to profound forms with IQ below 20-25. For detection of disease-causing mutation in such a heterogeneous disorder, exome sequencing is a powerful tool as almost all known gene can be assessed simultaneously in a high-throughput manner.

In this study, we performed exome-sequencing for several families with non-syndromic autosomal recessive ID (NS-ARID). After genotyping and exome sequencing, several interesting genes were detected. For example, a homozygous missense mutation was detected in *LRP2* gene in a Pakistani patient with mild ID from a marriage between second cousins. The variation was absent in 400 ethnically matched healthy control chromosomes and is not listed in SNP databases. The *LRP2* mutation identified here is located in one of the LDL-receptor class Adomains which is a cysteine-rich repeat that plays a central role in mammalian cholesterol metabolism suggesting that alteration of cholesterol processing pathway can lead to intellectual disability. In another Pakistani family, a homozygous missense mutation in *NUP205* gene was detected in a female patient with mild ID. Sanger sequencing analysis showed complete segregation within the family. Functional studies are ongoing to verify the pathogenicity of variants detected.

Here we describe the use of autozygosity mapping and whole exome sequencing to identify new genes for NS-ARID in several families. Due to the widely use of exome sequencing where all the genes are assessed at once, more genes for ID are identified leading to identification of underlying pathways.

795T

Low-frequency coding variation in PRF1 and GALC mediate multiple sclerosis risk. *C. Cotsapas[1,2], M. Mitrovic[1], International MS Genetics Consortium.* 1) Neurology, Yale School of Medicine, New Haven, CT; 2) Broad Institute of MIT and Harvard, Boston MA.

Multiple sclerosis (MS) is an autoimmune disease in which the myelin sheath surrounding brain neurons is destroyed by the immune system, leading to progressive physical and cognitive morbidity, brain atrophy and death. Over the last decade we have conducted a series of genome-wide association studies showing that the genetic architecture of MS is complex and likely involves hundreds of risk loci. We have identified almost 100 such loci explaining over 50% of the heritability and our ongoing common variant mapping efforts in ~36,000 cases and ~44,000 controls will substantially increase this number. To complement these efforts we have interrogated 250,000 low-frequency non-synonymous (NS) coding variants across all exons in the genome using Illumina's exome chip in a total of ~40,000 cases and ~60,000 controls of European descent. In our interim analysis, we have analyzed 26,231 cases and 24,031 controls in 14 country-level strata following stringent quality control to eliminate technical artefacts and population stratification. We tested >80,000 NS polymorphic variants for association to MS risk and find that two low-frequency variants (MAF < 0.05) show convincing evidence of association ($p < 6.25 \times 10^{-7}$, Bonferroni-adjusted p < 0.05 for number of variants tested). These variants are in genes PRF1 (preforin 1, OR = 1.2) and GALC (galactosylceramidase, OR = 0.77). A further variant in HDAC7, a histone deacetylase, has suggestive evidence of association ($p = 1.97 \times 10^{-5}$, OR = 0.77). We are currently analyzing our remaining samples to perform a joint analysis across approximately 100,000 cases and controls. Recognizing that even in a cohort of this size, individual variant tests are underpowered to detect small effects on risk, we are performing gene burden tests to identify genes harboring more low-frequency MS risk variants than expected by chance. As not all genes tolerate substitutions equally, we shall also look for genes with strong mutational constraint and association to MS risk. Overall, our results suggest that low-frequency variation identifies a number of MS risk genes in both known risk loci and elsewhere in the genome. These effects account for a small but significant proportion of disease risk heritability and reveal novel risk genes and aspects of MS susceptibility biology.

796S

Genomic Insights into Innate immunity against Viral Respiratory Infections in Pediatric Population. *S. Asgari[1], C. Hammer[1], I. Bartha[1], P. McLaren[1], L. Schlapbach[2], J. Fellay[1].* 1) School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; 2) Paediatric Critical Care Research Group, Mater Research, University of Queensland, Brisbane, Australia.

Background: Respiratory viruses are the most common pathogens leading to non-elective admission to Pediatric Intensive Care Unit. Human genetic variation plays a key role in determining individual responses after exposure to infectious agents. Current knowledge on the genetic basis of susceptibility to common respiratory infections is very limited. Study objectives: This project aims at identifying and functionally characterizing rare human genetic variants conferring unusual susceptibility to viral respiratory infections in the pediatric population. Methods: Previously healthy children below 4 years of age that required respiratory support for a viral respiratory disease were prospectively recruited in Switzerland and Australia between 2011 and 2014. Exome capture was done using Agilent SureSelect capture kits, and prepared libraries were sequenced on Illumina HiSeq2000. Short sequencing reads were mapped to hg19. Variant calling was carried out using GATK3.2 Haplotypecaller. After read mapping to hg19, variant calling and annotation, and filtration for potentially high-impact changes, we compared allelic frequencies between our cases and controls from the Exome Variant Server and the 1000 Genomes Project, both at single variant and at gene levels. In addition, we tested all genes for potential enrichment in rare (MAF <1%) loss-of-function variants (LoFs) using a binomial test. Results: A total of 120 patients were recruited and sequenced. Single variant and gene-based association testing did not show any significant difference between cases and controls. We observed a significant enrichment (p<2.5E-06) for LoFs in 128 genes; two of those were found in homozygous form, including IFIH1, which encodes a virus-specific pattern recognition receptor. A total of 9 patients carried a LoF in this gene. We studied by RNA sequencing a variant mapping to an essential splice site, which was found both in homozygous and heterozygous form in our study population (N=1 and 3, respectively); the analysis showed that the rare allele leads to the skipping of an exon coding for a protein domain involved in viral RNA recognition. Conclusions: This study provides proof-of-concept that rare genetic variations can be responsible for severe presentations of common viral infections in previously healthy children. Using exome sequencing, we successfully identified potentially deleterious mutations. Further studies are needed to explore the functional impact of these variants.

797M

Evaluating the impact of rare functional polymorphisms in pediatric sepsis. *A. Bittencourt Piccini[1,2], L. Schlapbach[4], I. Bartha[1,2,3], P. J. McLaren[1,2,3], J. Fellay[1,2,3], Swiss Pediatric Sepsis Study.* 1) School of Life Science, EPFL, Lausanne, VD, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne Switzerland; 3))Institute of Microbiology, University Hospital of Lausanne, (CHUV), Lausanne, Switzerland; 4) Paediatric Critical Care Research Group, Mater Children's Hospital, Brisbane, Australia.

Sepsis is defined as a systemic inflammatory response syndrome (SIRS) in response to an infectious process. It is one of the leading causes of infant mortality with over 1 million neonatal deaths worldwide every year. Up to 50% of lethal infections occur in apparently healthy children. A number of genetic variants have been identified through candidate gene studies, but they have not been confirmed independently. Through a non-biased exome sequencing approach, we scanned 43 samples from the Swiss Pediatric Sepsis Study for genetic variants that may underlie their susceptibility to sepsis. After sequencing, read alignment and quality control, we identified an average of 42,473 variants per individual, for a total of 277,625 variant loci. The majority of these variants are single nucleotide changes. Individuals that did not cluster with European ancestry population were filtered out. Using Snpeff software for functional annotation, we found that 1395 genes harbored at least one loss of function variant. We also ran a gene-based analysis of variant burden using VAAST software, which identified 9 genes that were significantly enriched for potentially deleterious variants in our cases vs. population controls. These preliminary findings require confirmation in additional samples. Although no causal gene or variant has been identified in this first analysis, candidate genes were identified that could be implicated in unusual susceptibility to pediatric sepsis. Ongoing recruitment, sequencing and bioinformatic analyses will permit further exploration of human genetic variation that might play a role in the pathogenesis of sepsis.

798T

Exome sequencing of 487 Community Acquired Pneumonia patients. *K.S. Elliott[1], A. Ndungu[1], T.C. Mills[1], A.L. Rautanen[1], P. Hutton[2], C. Garrard[2], A. Gordon[3], C.M. Hinds[4], M. Lathrop[5], A.V.S. Hill[1], S.J. Chapman[1].* 1) Wellcome Trust Centre Human Genetics, University of Oxford, Oxford, UK; 2) Intensive Care Unit, John Radcliffe Hospital, Oxford, UK; 3) Anaesthetics, Pain Medicine and Intensive care, Imperial College, London, UK; 4) William Harvey Research Institute, Queen Mary University of London, London EC1M 6BQ, UK; 5) McGill University-Génome Québec Innovation Centre, Montreal, Canada.

Respiratory infection is the largest contributor to global disease burden and pneumonia kills over one million children each year. A major genetic component of the infectious disease was demonstrated by a study of Danish adoptee children where a 5.8-fold increased risk of death from infectious disease was observed if one of their biological parents had died prematurely from infection. Severe bacterial disease may exert enormous selective pressure leading to the finding of rare susceptibility variants of relatively recent origin. In order to identify such variants an exome sequencing study was undertaken. DNA samples from 487 adult UK individuals admitted to an intensive care unit with severe community-acquired pneumonia (CAP) were collected as part of a study of genetic predictors of death from sepsis in critically ill patients (Genomic Advances in Sepsis [GAinS]). Analysis of sepsis susceptibility was performed on a discovery cohort of 270 CAP samples compared to the UK10K ALSPAC control dataset. After stringent QC criteria were applied, 135,392 variants were identified. Of these, 43 reached ExWAS significant threshold for association ($p < 3.6 \times 10^{-7}$) and an additional 63 variants were suggestive ($p < 1 \times 10^{-4}$). The exomes from the remaining 217 CAP patients are being analysed as a replication dataset against the UK10K TWINSUK control dataset. The sepsis outcome phenotype was also analysed, measured as 28 day mortality post-ICU admission within the 487 combined CAP cohorts (deaths n=237, survivors n =237, unknown n=13). In single variant analysis of - sepsis outcome, seven variants were identified reaching the ExWAS significance threshold including variants in two related genes known to be involved in thrombosis. Collapsing methods with rare deleterious variants are being performed to detect gene centric associations. Identification of novel, large-effect genetic variants has the potential to significantly expand current understanding of sepsis biology and may have clinical applications.

799S

Genome-wide exome array analyses reveal novel rare variants for refractive error in Asia populations. *Q. Fan[1], J. Liao[2], CC. Khor[3], YY. Teo[1,4], SM. SAW[1,2], TY. Wong[1,2,5], CY. Cheng[1,2,5].* 1) SSH School of Public Health, National University of Singapore, Singapore, Singapore; 2) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore; 3) Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore; 4) Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore; 5) Department of Ophthalmology, National University of Singapore, Singapore, Singapore.

As one of the leading causes of visual impairment, myopia poses a significant burden to public health in Chinese population. Genome-wide association studies have identified genetic loci contributing to refractive error, however, a substantial portion of the heritability remains unknown. We conducted exom array analysis for low-frequency (MAF 1-5%) and rare variants (MAF < 1%) on spherical equivalent in 2,253 Chinese and 2,091 Malay adults residing in Singapore. We genotyped the whole genome exome variants using the Illumina HumanExome Beadchip. Among 274,00 variants passing quality control, 84,000 (31%) were polymorphic variants. We conducted the regression-based association test for single-variant (MAF > 0.1%) and burden test (SKAT) for 53,991 nonsynomous variants (MAF < 5 %) within 14,130 genes. One gene harboring 4 rare/low frequency nonsynomous variants (cumulative MAF = 3.6%) on chromosome 1 was identified to be associated with SE variation after Bonferroni correction for the number of genes ($P = 2.25 \times 10^{-7}$, respectively). Conditional analysis suggested the signal of the top gene was mainly driven by the single low frequency variant within this gene. This protein coding gene is a part of the epidermal growth factor receptor (EGFR) pathway, whereas EGFR signaling has been implicated in the myopia development in animal models. Six genes containing at least two rare nonsynomous variants also exhibited suggestive association with SE ($P < 5.0 \times 10^{-5}$). Our data suggest the likely role of rare genetic variants in influencing refractive error. The interpretation of the gene-based tests needs carefully considering the effects of single rare variants where the large sample size is required.

800M

Analysis of the rare variant burden in the exomes of candidate HIV-target genes in relation to HIV-acquisition and AIDS-progression. *M.C. Turchin[1], S. Penugonda[2], E-Y. Kim[2], K. Kunstman[2], M. Stephens[1,3], S.M. Wolinsky[2].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) The Feinberg School of Medicine, Northwestern University, Division of Infectious Diseases, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL.

Multiple genome-wide association studies (GWAS) have been conducted attempting to link common human genetic variation (minor allele frequency, MAF, >5%) to various aspects of HIV and AIDS pathology. Despite using large sample sizes (up to 6,000s) and samples representing ancestries beyond Western Europeans (Fellay et al. 2009, Pelak et al. 2010, McLaren et al. 2013), the majority of associations have only been found within the human leukocyte antigen (HLA) region. Here, we present the results of a more focused gene-exome sequencing study utilizing ~550 genes whose protein-products have previously been implicated as significantly associating with the 18 HIV-1 proteins (Jäger et al. 2012). These ~550 genes were sequenced in over 900 individuals, the majority of which are of Western European descent. Individuals in this sample were classified as either sero-positive or sero-negative for associations involving HIV-acquisition, with HIV-positive individuals further being classified as slow/very-slow progressors or rapid/very-rapid progressors (where applicable), thus allowing us to look at associations regarding AIDS-progression as well. Analyses were conducted focusing on both the rare variant burden (MAF <5%) in the exomes of these genes as well as variants that are classified as 'damaging' as designated by multiple functional-prediction algorithms. Overall, we find a handful of genes outside the HLA region that appear to contain an excess of rare, 'damaging' variants in one class of individuals versus another for both HIV-acquisition and AIDS- progression. These results provide further insight into the biology of HIV-human interactions, and with proper functional follow-up and validation, may provide future directions for the development of HIV-drug targets.

Fellay J et al. (2009) Common Genetic Variation and the Control of HIV-1 in Humans. PLoS Genet 5(12): e1000791

Jäger S et al. (2012) Global landscape of HIV-human protein complexes. Nature 481: 365-370.

McLaren PJ et al. (2013) Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. PLoS Pathog 9(7): e1003515

Pelak K et al. (2010) Host determinants of HIV-1 control in African Americans.J Infect Dis 201(8): 1141-9.

801T

The application of the new D5000 ScreenTape assay in larger NGS library quality control. *A. Inche[1], L. Prieto-Lafuente[1], C. McDonald[1], E. Brown[1], B. McHoull[1], M. Liu[2], R. Salowsky[3].* 1) Agilent Technologies Ltd, Edinburgh, United Kingdom; 2) Agilent Technologies, La Jolla, CA, USA; 3) Agilent Technologies GmbH, Waldbronn, Germany.

The current winners of the next generation sequencing (NGS) race are the short read sequencers. As this technology continues to be rapidly adopted, smarter strategies are being developed by manufacturers and in the wider NGS community to address some of the short comings of these short reads. Some of the solutions currently available have increased the overall library size up to and above 2000bp. The Agilent 2200 TapeStation is now a very well accepted QC tool within the NGS library preparation workflow. However, it has been somewhat tricky to QC these larger libraries on the current assay portfolio for this system. Here we present data from some of these larger libraries as analysed on the new D5000 ScreenTape assay for the 2200 TapeStation, which should help streamline sample QC of these libraries. The data shown includes samples from the Illumina Nextera Tagmentation kit which typically generates libraries up to and over 2000bp. We also present an overview of the performance characteristics of the D5000 ScreenTape assay.

802S

Using Exome Sequencing Followed By Genotyping To Identify Susceptibility Gene For Morbid Obesity. *H. Jiao[1], P. Arner[2], P. Gerdhem[3,4], R.J. Strawbridge[5], E. Näslund[6], A. Thorell[7], A. Hamsten[5], J. Kere[1], I. Dahlman[2].* 1) Department of Biosciences and Nutrition, Karolinska Institutet, SE-141 57 Stockholm, Sweden; 2) Department of Medicine (H7), Huddinge, Karolinska Institutet, SE-141 86 Stockholm, Sweden; 3) Department of Orthopaedics, Karolinska University Hospital, SE-141 86 Stockholm; 4) CLINTEC, Karolinska Institutet, SE-141 86 Stockholm, Sweden; 5) Cardiovascular Genetics and Genomics Group, Atherosclerosis Research Unit, Department of Medicine, Solna, Karolinska Institutet, SE-171 76 Stockholm; 6) Department of Clinical Sciences, Danderyd Hospital, Karolinska Institutet, 182 88 Stockholm, Sweden; 7) Department of Clinical Science at Danderyds Hospital and Department of Surgery, Ersta Hospital, Stockholm, Karolinska Institutet, Sweden.

We used exome sequencing to identify variants associated with morbid obesity. DNA from 100 morbidly obese adult subjects and 100 controls were pooled (n=10/pool), subjected to exome capture, and subsequent sequencing. At least 100 million sequencing reads were obtained from each pool. After several filtering steps and comparisons of observed frequencies of variants between obese and non-obese control pools, we systematically selected 144 obesity-enriched non-synonymous, splicing site or 5'upstream single nucleotide variants for validation. We first genotyped 494 adult subjects with morbid obesity and 496 controls. Five obesity-associated variants (nominal p-value < 0.05) were subsequently genotyped in 1,425 morbidly obese and 782 controls. Out of the five variants only one was confirmed. The variant showed strong association with body mass index ($p=6.28\times10-5$) in joint analysis of all 3,197 genotyped subjects, and had an odds ratio of 1.32 for obesity association. In conclusion, using exome sequencing we identified a low-frequency coding variant that was associated with morbid obesity. The corresponding gene display reduced body weight in mice. It may be involved in the development of excess body fat.

803M

Identification of genes involved in functional recovery after stroke through exome sequencing of extreme phenotypes. *R. Rabionet[1], M. Mola[2], G. Escaramís[1], H. Susak[1], C. Soriano[2], C. Carrera[3], S. Ossowski[1], I. Fernandez-Cadenas[3], J. Roquer[2], J. Jimenez-Conde[2], X. Estivill[1].* 1) Bioinformatics and Genomics, Center for Genomic Regulation, Barcelona, Spain; 2) Neurovascular Research Group (NEUVAS), IMIM, Barcelona, Spain; 3) Fundació Docencia i Recerca Mutuaterrassa, Terrassa, Barcelona, Spain.

Cerebrovascular diseases are the second most important cause of death in Spain, and together with other neurodegenerative diseases, they are the leading cause of disability in adults. Variability in functional outcome after a stroke can be influenced by many factors. Irrespective of clinical factors such as age, stroke etiological subtype, vascular stenoses, location of the injury and the size of the affected area, inter-individual variation in capacity of neuronal recovery is considerable. A number of systems and metabolic pathways are important for response to cerebral ischemic damage, and their activity may be modulated by variation in the genes that encode their various components. We aim to identify genetic variants, genes and pathways influencing the functional recovery process. With this aim, we have selected 81 patients with extreme phenotypes (36 bad vs 45 good outcome), suffering an anterior territorial ischemic stroke, with similar stroke severity, and matched for basal functional level, age and gender, from a cohort of over 4000 stroke cases. These patients underwent exome sequencing (Nimblegen v3 and Illumina sequencing). Downstream analysis was performed using a mixed-model association test, with variants collapsed by gene and weighted by their frequency and condel scores, with the aim of identifying genes with an accumulation of variants in either set of samples. This provided a set of aprox. 400 nominally significant genes (although none passed multiple testing correction), which includes some relevant genes for stroke risk and stroke outcome. Further analysis including CNV analysis and integration with GWAS results from this and other cohorts of stroke cases is underway. These findings will be validated in the extended cohort.

804T

Lessons Learned From the Sequencing of Severe Insulin Resistance Exomes. *F. Payne[1], W. Bottomley[1], I. Isaac[2], S. O'Rahilly[2], D.B. Savage[2], R.K. Semple[2], I. Barroso[1], The UK10K Consortium.* 1) Metabolic Disease Group, Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Cambridge, United Kingdom.

The UK10K Consortium (http://www.uk10k.org/) is a collaboration between multiple research centres, mainly in the UK, aiming to uncover genetic variants contributing to disease and health status by sequencing 10,000 people. As part of the Rare Disease Group, the exomes of a 1,000 individuals with rare, extreme conditions have been sequenced to an average read depth of 72x using next generation sequencing technology (NGS), with an aim to detect novel causal variants. Amongst these, we focused our analysis on patients with disorders of insulin action of unknown aetiology. These are a group of highly heterogeneous conditions encompassing both severe insulin resistance (SIR) syndrome (a collection of rare disorders of extreme resistance to the glucose-lowering effects of insulin) and unrestrained metabolic or mitogenic insulin-like activity in the absence of insulin. In the discovery phase, 60 probands and their family members (125 samples in total) were whole-exome sequenced, with follow-up in a further 250 index cases by targeted sequencing of 76 genes selected either as good candidates for disease from previous NGS or GWAS studies, or already known to be disrupted in individuals with disorders of insulin action (e.g. *INSR, LMNA, PPARG, AGPAT2* and *AKT2*). After sample quality control, 117 discovery and 248 follow-up samples were carried through for further analysis. To prioritise putative causative variants after removing those with low sequencing depth and quality, we filtered using the 1000 Genomes (http://www.1000genomes.org/), NHLBI Exome Sequencing Project (ESP: http://evs.gs.washington.edu/EVS/) and the remaining UK10K exomes and genomes to exclude common variants and focused on those predicted to alter protein sequence (potentially functional). Family trios were also assessed to detect variants with a high probability of being de novo using DeNovoGear (https://sourceforge.net/projects/denovogear/). Possible candidates were confirmed by Sanger sequencing and co-segregation with available family members was examined wherever possible. Combined analysis of discovery and replication samples has provided possible diagnosis to 5.6% of the index cases to date. These include mutations within genes known to be disrupted in syndromes of SIR (*INSR, PPARG, AGPAT2, ALMS1*) and in strong candidates either with a role in insulin action (*MTOR, RICTOR, SNAP23, TBC1D4*) or those within pathways relevant to phenotypic features (*POC1A*).

805S

Disease associated variants in healthy centenarian exomes. *L.C. Tindale[1,2], S. Leach[1], A.R. Brooks-Wilson[1,2].* 1) Genome Sciences Centre, BC Cancer Research Centre, Vancouver, BC, Canada; 2) Simon Fraser University, Burnaby, BC, Canada.

Many people strive to live to 100, but very few attain this goal. The genetic component to longevity has been established through twin and family studies, but the search for specific variants and mechanisms to explain this rare and complex phenotype is ongoing.

We resequenced the exomes of four healthy centenarians who had never been diagnosed with cancer, cardiovascular disease, diabetes, Alzheimer disease, or major pulmonary disease, as well as four mid-life individuals. The objective of this pilot study was to gain insight into whether the centenarians' exceptional health and longevity were a result of carrying fewer deleterious variants than typical individuals. We hypothesized that the centenarians would have 1) fewer common disease associated genome-wide association study (GWAS) variants, 2) on average fewer variants predicted to be damaging, and 3) few or no high penetrance variants in known disease genes.

The NHGRI Catalog of Published GWAS was filtered for exonic variants associated with common chronic diseases. The proportion of disease associated alleles of GWAS SNPs was compared between the centenarians and controls, as well as to published populations from HapMap and the NHLBI Exome Sequencing Project; however, no difference was detected between the centenarians and comparison groups. Analysis of predicted deleterious variants also showed no evidence of a difference in the burden of SNPs and insertions or deletions predicted to have a high functional impact, between centenarians and controls. Lastly, we are in the process of screening for variants in the 57 genes classified by the American College of Medical Genetics and Genomics as carrying clinically relevant variants.

Our findings support the interpretation that attaining extreme longevity does not appear to be a result of carrying fewer common disease variants, or having a lower burden of predicted damaging variants. We suggest that future studies should focus on seeking rare longevity-promoting variants that contribute to long term good health.

806M

Genetic variation among Multiple Sclerosis in Saudi Patients. *M. Albalwi[1, 2, 3], I. Alabdulkareem[2, 3], W. Alharbi[2], M. Balow[2], Z. Rabhan[2], M. Aljumah[2,3].* 1) Department of Pathology and Laboratory Medicine, King Abdulaziz Medical City, National Guard Health Affairs, Riyadh, Saudi Arabia; 2) King Abdullah International Medical Research Center, Medical Genomic Research Department, Riyadh, Saudi Arabia; 3) King Saud Bin Abdulaziz University for Health Sciences, College of Medicine, Riyadh, Saudi Arabia.

Multiple Sclerosis (MS) is an immune-mediated, inflammatory and demyelinating disorder of the central nervous system (CNS), which includes damage to axons, oligodendrocytes and neurons. After trauma, MS is the most common acquired neurological disorder of young adults. The disease shows a spectrum of severity, ranging from an asymptomatic pathological process, to mildly symptomatic, to severely disabling disease. Although the pathogenesis of MS is poorly understood, available evidence suggests that both genetic and environmental components play important roles in disease development, both independently and interactively. The contribution of genetics to MS is supported by many reports showing familial aggregation of the disease, high concordance rates among twins, and an increased risk among relatives of patients with MS. The role of genetics in MS and its interaction with environmental triggers are currently being extensively studied. However, the purpose of the study is to identify the genetic variation and the risk factors that may be associated with Multiple Sclerosis in the affected patients from Saudi Arabia. Using HLA high-resolution genotype sequencing and whole exom sequencing approch on FLX+ Titanium chemistry and solid 5500 next generation sequencing revealed presence of scientific variation evidence shared among our MS patients. Further more, our analysis suggesting that some of these variations findings may contribute to drugs effect and pathogenesis of MS and others autoimmune diseases in Saudi population.

807T

Admixture mapping of exome genotyping data implicates region 15q21.2-22.3 with keloid risk in African Americans. *K.S. Tsosie[1,2], D.R. Velez Edwards[1,3,4,5], S.M. Williams[6], T.L. Edwards[1,2,3,4], S.B. Russell[1,7].* 1) Center for Human Genetics, Vanderbilt University, Nashville, TN; 2) Division of Epidemiology, Department of Medicine, Vanderbilt University, Nashville, TN; 3) Vanderbilt Epidemiology Center; 4) Institute for Medicine and Public Health; 5) Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN; 6) Department of Genetics, Geisel School of Medicine, Dartmouth University, Hanover, NH; 7) Division of Dermatology, Department of Medicine, Vanderbilt University, Nashville, TN.

Keloids (MIM 148100) are benign dermal fibrotic tumors with no effective clinical remedy that affect people of recent African ancestry approximately 20 times more than individuals of Caucasian descent. Possible related fibroproliferative diseases with increased prevalence in African populations include hypertension, nephrosclerosis, allergic disease, and uterine fibroma. Familial aggregation and ancestral differences in risk among geographic subpopulations strongly suggests a genetic association between African ancestry, keloids and fibroproliferative disease risk. There are no published genome-wide studies of keloid risk in African ancestry subjects. We conducted admixture mapping (AM) and whole exome association in 478 African Americans (AAs: 122 cases, 356 controls) with exome arrays to identify regions of local ancestry and SNP genotypes under AM peaks associated with keloid risk. **Results:** The most significant association with keloids discovered by AM was observed on chr15q21.2-22.3. This 5Mb region includes *NEDD4*, which was previously implicated in keloid formation by GWAS in Japanese and later validated in Chinese. Though our study nominally replicated this finding by AM and genotype association, the most significant SNP genotype association under the AM peak was observed at *MYO1E* (rs747722, odds ratio [OR]=4.41, 95% confidence interval [CI]=2.29-8.50, p=9.07×10^{-6}). A scan of all common genotype associations also identified associations at *MYO7A* (rs35641839, OR=4.71, 95% CI=2.38-9.32, p= 8.34×10^{-6}) at chr11q13.5. GWAS have linked the chr15q21.2-22.3 region with hypertension in AAs, asthma in Europeans, and atherosclerosis in a Finnish cohort, providing evidence for common genetic elements. Examination of earlier microarray data of fibroblasts from keloids and normal scars that included some subjects from this study also implicated chr15q21.2-22.3 as a causal region for keloids, with increased expression of *MYO1E* in keloids compared to normal scars. Notably, *MYO1E* has been shown to be a crucial component of the invadosome, a structure involved in matrix degradation and invasion and thus may have a functional role in the keloid phenotype. **Conclusion:** This study is the first to use AM and exome array association analysis to explore the genetics of keloids in AAs. Our findings, strengthened by support from expression data, further elucidate a potential region on chr15q21.2-22.3 for a role in risk of keloids in AAs, Japanese, and Chinese populations.

808S

Exome sequencing in pooled DNA samples identifies a potential candidate variant for preeclampsia. *T. Kaartokallio[1], J. Wang[2,3], H. Jiao[2,3,4], J. Kere[1,2,3,4,5], H. Laivuori[1,6,7].* 1) Haartman Institute, Medical Genetics, University of Helsinki, Helsinki, Finland; 2) Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden; 3) Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden; 4) Clinical Research Centre, Karolinska Institutet, Stockholm, Sweden; 5) Folkhälsan Institute of Genetics, Helsinki, Finland; 6) Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Helsinki, Finland; 7) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

Preeclampsia (PE), a common vascular pregnancy disorder, increases the risk for mortality and morbidity for both mother and child. As PE is a placental disease that reduces reproductive success, it is plausible that variants predisposing to this disease are under negative evolutionary selection and do not reach high frequencies in a population. In this study, our aim was to identify low-frequency variants that predispose to PE in a Finnish population by utilizing exome sequencing in pooled DNA samples. The exonic regions of a 100 preeclamptic women were sequenced in pools of ten using Illumina platform. The reads were aligned to the human reference genome with the BWA software package, and variants were called with SAMtools and annotated with ANNOVAR. In the filtering step, we selected missense, nonsense and splice site variants that were present in at least two PE pools and had minor allele frequency (MAF) ≤ 0.05 in the 1000G_EUR data. MAFs of the variants in our data were estimated by pooling all the reads at each position together and calculating the percentage of the reads supporting the minor allele. Estimated MAFs were then compared to MAFs in the 1000G_EUR data, in the SISu (Sequencing Initiative Suomi) data (a data set currently containing exome data from ~3300 Finns), and in scoliosis exome seq data, which was produced with a strategy similar to ours. Fifty-nine variants that fulfilled the above filtering criteria and were enriched in our data compared to the reference data (PE_MAF/reference_MAF ≥ 1.5) were selected for validation. Forty-eight of these variants were successfully Sequenom genotyped in 180 pregnant controls and in 180 PE cases, including the original cases, to confirm the presence of the variants in the original samples and to validate the MAF difference between the cases and controls/general population. Next, twenty-eight of the variants with OR ≥ 1.3 or ≤ 0.67 were selected for genotyping in 1456 PE cases and in 704 pregnant controls to test the association of the variants with PE. To increase the power, genotypes from the SISu data were included in the analysis. With this approach, we identified a nominal association of four variants with PE. The strongest association was seen for a variant in the *TP53BP* gene (uncorrected p 0.006). Replication genotyping of this variant is currently underway in a Scandinavian PE case-control cohort.

809M

High burden of deleterious variants and the genetic basis of speech sound disorders. *H.A. Voss-Hoynes[1], C.M. Stein[1], W.S. Bush[1], B. Truitt[1], L. Freebarin[2], B.A. Lewis[2], S.K. Iyengar[1]*. 1) Department of Epidemiology and Biostatistics and; 2) Communication Sciences, Case Western Reserve University, Cleveland, Ohio.

Speech sound disorders (SSD) are communication disorders occurring in 16% of 3 year olds and cost an estimated $30 - $154 billion annually in lost productivity, special education, and medical care. SSD exhibits high heritability, and for childhood apraxia of speech (CAS), a severe subtype of SSD, there are a few candidate genes such as FOXP2. However, apart from this knowledge, a comprehensive understanding of SSD genetics remains to be developed. We hypothesize that SSD is similar to neurodevelopmental disorders such as autism and intellectual disability and occurs not due to a single variant, but rather as the result of a high burden of multiple variants. The present study to test this hypothesis is a 26-year longitudinal cohort study based in Cleveland, OH, involving 135 cases with SSD, 34 (35%) of whom are affected with CAS, and 213 unaffected parental controls. To examine the genetic component of SSD, we used both exome sequencing and HumanOmni2.5Exome genotyping. Analysis of FOXP2 in exome sequenced (n=11 CAS) and genotyped (n=135 SSD and CAS) samples did not identify any single variant or CNVs that could be considered causative mutations. We expanded our analysis to examine overall variant burden using a single measure of deleteriousness, Combined Annotation Dependent Depletion scores (C-scores), which allows for comparison of the distribution of deleterious variants between groups. Preliminary results on a subset of 7,000 variants indicate that while there is no significant difference between the total number of rare (<1%) variants per individual (mean # case=3.43 vs control=3.26, p=0.5), there are distinctions in the deleteriousness of the variants present. Compared to controls (mean RawC=4.39), the CAS cases alone (mean RawC=6.07), but not all cases, trend toward a higher burden of deleterious variants (p=0.08). Additionally, compared to cases without CAS (mean RawC=4.16), apraxic individuals have a significantly higher burden of deleterious rare variants (p=0.05). Finally, a missense variant in BTBD8, a gene expressed prominently in the fetal, but not the adult brain, exhibited the most extreme frequency difference between apraxic cases vs controls (14% vs 5%). Our results suggest that the genetic etiology of SSD is best examined in an integrative, whole genome manner and imply that human speech evolved through changes in many genes and interactions between them rather than as a single event. Acknowledgements: DC00528, DC012380, T32-HL007567.

810T

Functional follow-up, fine mapping and haplotype meta-analysis improve insight in findings from exome chip analyses and reveal potential novel fasting glucose associations. *S.M. Willems[1,2], S. Wang[3], J. Wessel[4,5], A.Y. Chu[6], H. Yaghootkar[7], L. Lipovich[8,9], M-F. Hivert[10,11,12], P. An[13], Y. Liu[14,15], L.A. Lange[16], J.G. Wilson[17], J.E. Huffman[18], M. Dauriz[19,20,21], B. Hidalgo[22], R.J.F. Loos[14,15,23], T. Frayling[7], D. Siscovick[24,25], J.I. Rotter[26], R.A. Scott[1], J.B. Meigs[19,20], J. Dupuis[3,27], M.O. Goodarzi[28], CHARGE GLYCEMIA-T2D EXOME CHIP WORKING GROUP*. 1) MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK; 2) Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 3) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 4) Fairbanks School of Public Health, Department of Epidemiology, Indianapolis, IN, USA; 5) Indiana University School of Medicine, Department of Medicine, Indianapolis, IN, USA; 6) Division of Preventive Medicine, Brigham and Women's Hospital, Boston MA, USA; 7) Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK; 8) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI, USA; 9) Department of Neurology, Wayne State University School of Medicine, Detroit, MI, USA; 10) Harvard Pilgrim Health Care Institute, Department of Population Medicine, Harvard Medical School, Boston, MA, USA; 11) Division of Endocrinology and Metabolism, Department of Medicine, Université de Sherbrooke, Sherbrooke, Québec, Canada; 12) Department of Epidemiology, University of Washington, Seattle, Washington, USA; 13) Division of Statistical Genomics and Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA; 14) The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY, USA; 15) The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, NY, USA; 16) Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; 17) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA; 18) MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Edinburgh, Scotland, UK; 19) Massachusetts General Hospital, General Medicine Division, Boston, MA, USA; 20) Department of Medicine, Harvard Medical School, Boston, MA, USA; 21) Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, University of Verona Medical School and Hospital Trust of Verona, Verona, Italy; 22) Section on Statistical Genetics, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL; 23) The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY, USA; 24) New York Academy of Medicine, New York, New York, USA; 25) Cardiovascular Health Research Unit, Departments of Medicine and Epidemiology, University of Washington, Seattle, WA, USA; 26) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; 27) National Heart, Lung, and Blood Institute (NHLBI) Framingham Heart Study, Framingham, MA, USA; 28) Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, USA.

To explore the role of coding single nucleotide variants (SNVs) on fasting glucose (FG) levels, we performed exome-wide association analysis in up to 60,564 non-diabetic individuals of European (84%) and African (16%) ancestry from 23 studies with Illumina Exome chip data. We performed single variant analyses for all coding and non-coding SNVs with MAF>0.02%. Gene-based tests included rare (MAF<1%) SNVs predicted to be protein-altering. Single variant analyses revealed two novel FG-associated loci: *GLP1R* (SNV A316T) and *ABO* (four non-coding SNVs, lead rs651007). Follow-up of *GLP1R* in exome sequencing data from up to 14,118 individuals, performed to investigate association of sequence variation not captured by the exome chip, identified an intronic SNV 100 bp from A316T (rs761386, MAF=3%, β=0.09, $p=7.6\times10^{-6}$) that was more strongly associated with FG than A316T (p=0.01). rs761386 was close to a splice site, suggesting a possible effect on *GLP1R* pre-mRNA splicing. Follow-up of the four *ABO* SNVs using GTEx and ENCODE Consortium resources identified that rs507666 resides near the transcription start site of a long non-coding RNA that is antisense to exon 1 of *ABO* and expressed in pancreatic islets. rs507666 is also an eQTL for the glucose transporter *SLC2A6* ($p=1.1\times10^{-4}$). Gene-based analyses identified a set of 15 SNVs associated with FG in the previously established *G6PC2* locus. To further investigate the combined effect of these SNVs, we performed haplotype meta-analysis showing significant association ($p=1.1\times10^{-17}$) for haplotypes with the 15 SNVs. Haplotypes carrying the single rare allele at R283X (MAF=0.26%, $p=2.8\times10^{-10}$), P324S (MAF=0.19%, $p=1.4\times10^{-7}$) or S207Y (MAF=0.59%, $p=1.5\times10^{-6}$) were most strongly associated with FG compared to the most common haplotype. The association at the three SNVs was not explained by the known common intronic *G6PC2* SNV rs560887, which is close to the splice acceptor of intron 3 and may be implicated in *G6PC2* pre-mRNA splicing. rs560887 is also near the transcription start site of expressed sequence tag DB031634, a potential cryptic minor isoform of *G6PC2* mRNA. Together, these findings suggest that both coding and non-coding SNVs in *G6PC2* affect glucose homeostasis. In conclusion, further follow-up of exome chip results with exome sequencing, functional annotation and haplotype meta-analysis offers new insight into the functionality of observed associations and reveals novel variants implicated in regulation of FG levels.

811S

Exome Array Analysis of Quantitative Traits related to Glaucoma. *A.I. Iglesias Gonzalez[1], H. Springelkamp[1,2], S. van der Lee[1], N. Amin[1], C.C. Klaver[1,2], C.M. van Duijn[1].* 1) Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands; 2) Department of Ophthalmology, Erasmus MC, Rotterdam, The Netherlands.

Purpose: Intraocular Pressure (IOP) and Vertical Cup to Disc Ratio (VCDR) are extensively studied endophenotypes of glaucoma. Although there has been major progress identifying common variants with small effect, there has been limited progress identifying rare variants. To evaluate the association of exonic variants to IOP and VCDR in Caucasians, we analysed exome-array and exome-sequencing data in both a population-based and a family-based study. **Methods**: The Illumina HumanExome BeadChip containing ~ 250,000 variants was used to genotype 3,163 subjects from the Rotterdam Study I (RS-I), a population-based cohort, and 1,512 subjects from the Erasmus Rucphen Family (ERF) Study, a family-based cohort. In addition, overlapping variants of ~1,879 individuals from RS-I (~600) and ERF (~1,279) with exome-sequencing data available were included in the analysis. In total, 5,781 individuals with IOP data and 3,758 individuals with VCDR data were analysed. Rare variants were analysed for an association with IOP and VCDR using both single variant analysis and burden test implemented in SeqMeta; an additive genetic model was assumed adjusting for age, sex and the first five principal components (RS-I) or family structure (ERF). Variants with p-values $<2.0E^{-4}$ were selected for replication in RS-II and RS-III imputed with the Genome of the Netherlands (GoNL). **Results**: Single variant analysis for IOP identified a novel common missense variant rs11541353 (p-value = $2.33E^{-8}$) located in the *NPAS2* gene, involved in a circadian mechanism known as contrast sensitivity which is regulated by retinal ganglion cells. This finding was replicated using GoNL imputations. Gene-based analysis showed significant association (p-value $<1.85E^{-6}$) between IOP and three genes: *RNASEH2B*, *SPOPL* and *IVL* when using a minor allele frequency (MAF) upper bound of 5%. Gene-based analysis for VCDR identified one novel significant association (p-value $<1.85E^{-6}$) with *CDK13* which belongs to the family of cyclin-dependent kinases and interact with *CDKN2B*, a key gene involved in glaucoma. **Conclusions**: We have identified multiple genes associated with IOP. *NPAS2* is involved in the circadian clock mechanism. Changes in circadian rhythm have been found in glaucoma patients. *RNASEH2B* is involved in DNA replication, *SPOPL* belongs to the ubiquitin ligase complex, and *IVL* is important for keratinocyte differentiation. In addition, in the VCDR analysis we found a new gene involved in the cell cycle pathway.

812M

Relationship between neutrophil count, white blood count and *TCIRG1* variation. *E.A. Rosenthal[1], V. Makaryan[2], D.C. Dale[2], D.R. Crosslin[1,3], D.A. Nickerson[3], A.P. Reiner[4,5], G.P. Jarvik[1], NHLBI GO ESP.* 1) Dept Med Gen, Univ of Washington School of Medicine, Seattle, WA; 2) Div of GIM, Univ of Washington School of Medicine, Seattle, WA; 3) Dept Genome Sciences, Univ of Washington, Seattle, WA; 4) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 5) Dept of Epidemiology, Univ of Washington School of Public Health, Seattle, WA.

Congenital neutropenia is a hematological condition characterized by low neutrophil counts and recurrent bacterial infections. Recently, we reported linkage and association between congenital neutropenia and a novel *TCIRG1* missense mutation (NG_007878.1:c.2206C>A) in a five generation pedigree (PMID:24753205). We investigated the role of *TCIRG1* SNVs in neutrophil count (NPC) and total white blood count (WBC), of which neutrophils are the major component, using data from the exome sequence project (N=3560, WBC range 2×10^9 to 11×10^9 cells/Liter). After quality control, only 934 individuals (367 European ancestry (EA), 563 African ancestry (AA), 4 other ancestry) had measured NPC (NPC range 10-87% of total WBC, correlation with WBC = 0.40), reducing power to detect an association. We tested association with *TCIRG1* SNVs that were predicted to cause a coding or splicing change, using single SNV, burden and Kernel (SKAT) association tests, prioritizing evolutionarily conserved and predicted damaging SNVs. We detected suggestive evidence for association between NPC and the common missense SNV rs36027301 (1-tail p=0.02, β=-2.4), adjusting for age, sex, cohort, race, and *DARC* (rs2814778 and rs12075). For WBC, we analyzed EA (N=1957) and AA (N=1603) individuals separately. In the analysis of EA individuals, WBC was adjusted for age, sex and cohort. In the analysis of AA individuals, WBC was additionally adjusted for the known effects of *DARC*. In all tests, there were no statistically significant association of WBC and *TCIRG1* variants (p>0.2). These results could be due to difficulty in choosing SNVs with a true functional effect versus those that are benign, or low statistical power if variation in *TCIRG1* is associated only with NPC, but not with other WBC components, given the low correlation of WBC and NPC in these data. Further investigations in larger cohorts are underway.

813T

Identification of COPD causal variants by combining GWAS associated SNPs, lung eQTLs, and pathogenicity prediction tools. *M. Lamontagne[1], C. Couture[1], DS. Postma[2], W. Timens[2], K. Hao[3], D. Nickle[3], DD. Sin[4], PD. Pare[4], M. Laviolette[1], Y. Bosse[1,5].* 1) Centre de recherche de l'Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada; 2) University of Groningen, University Medical Center Groningen, GRIAC research institute, Groningen, The Netherlands; 3) Merck & Co. Inc., Rahway, New Jersey, USA; 4) University of British Columbia James Hogg Research Center, Center for Heart and Lung Health, St. Paul's Hospital, Vancouver, BC, Canada; 5) Department of Molecular Medicine, Laval University, Québec, Canada.

Introduction: Chronic obstructive pulmonary disease (COPD) is the fourth most common cause of death worldwide and is predicted to be the third leading cause of mortality by the year 2030. COPD is a complex disease characterized by a progressive airflow obstruction that is not fully reversible. Genome-wide association studies (GWAS) have identified polymorphisms associated with COPD and lung function. However, most of these SNPs are located in non-coding regions (intron and intergenic) and their mechanisms of action are unknown. The aim of the present study is to identify GWAS-associated SNPs that are also associated with gene expression levels in the lung and to verify whether these SNPs are in LD with possible pathogenic variants. **Methods:** Genome-wide association studies on COPD and lung function were reviewed to extract significant SNPs. GWAS-associated SNPs were analyzed to identify expression Quantitative Trait Loci (eQTLs) in a large-scale lung eQTL mapping study including 1,111 patients that underwent lung surgery at three participating sites. eQTLs analyses were performed independently in the three datasets and meta-analyzed using the Fisher's method. Missing genotypes were imputed with MACH and association tests between adjusted expression traits (n = 38,820 probe sets) and GWAS-associated SNPs were performed in PLINK. The Combined Annotation-Dependent Depletion (CADD) method was used to determine the potential pathogenicity of selected SNPs. **Results:** After imputation, 243 SNPs associated with COPD or lung function from previous GWAS were available in the three cohorts. Thirty-three GWAS-associated SNPs were significantly regulating the expression levels of 30 genes (43 probe sets) after correction for multiple testing (Bonferroni P-value $\leq 5.30 \times 10^{-9}$). The average CADD score among the 243 GWAS variants was 4.51 and 26 of them had a score greater than 10. Most of the significant eQTL-SNPs (29/33) were in strong LD ($r2 \geq 0.8$) with SNPs having possible pathogenic effect (CADD score $\geq$ 10). One SNP, rs4846480, was in strong LD with seven SNPs with a CADD score greater than 10. **Conclusion:** By combining genotypes from GWAS-associated SNPs and genome-wide gene expression in the lung, we identified genes that are likely to play a role in the pathogenesis of COPD. We also identified possible functional SNPs by combining GWAS-associated SNPs, lung eQTLs, and the latest tool to predict the pathogenicity of genetic variants.

814S

Association of *IRS2* gene polymorphism G1057D with obesity in young. *M. Martinez Lopez[1], C. Rodriguez-Perez[1], J.M Magaña-Cerino[1], R. Diaz-Martinez[2].* 1) Division Academica Ciencias de la Salud, Universidad Juarez Autonoma de Tabasco, Villahermosa, Tabasco, Mexico; 2) Hospital Civil de Guadalajara, Guadalajara, Mexico.

Obesity is a public health problem in the country and is associated with insulin resistance (IR). The substrate insulin receptor 2 (IRS2) in the signaling pathway of insulin is considered essential in the development and/or survival of β cells. The polymorphic variant of the G1057D IRS2 gene is associated with the pathogenesis of obesity and an increased risk of developing type 2 diabetes (T2D) in adulthood. Objective was to determine whether the IRS2 G1057D polymorphism gene is associated with obesity in young people. Materials and Methods a descriptive study was conducted on students of the Universidad Juarez Autonomy de Tabasco. the control group (n = 94) a group of cases was divided into obese with a history of diabetes (OCAD) (n = 94) and obese with no history of diabetes (OSAD) (n = 94). Peripheral blood samples for biochemical determination and for the extraction of genomic DNA were taken. Genotyping was performed by real time PCR with HRM and direct DNA sequencing. Statistical analysis One-way ANOVA was used, and the post hoc Dunnett t-test for metabolic characterization, for the association of the polymorphism with obesity, the X2 test was used and the OR was calculated on the significant value of $p<0.05$. Results obese youth had alterations in the lipid profile observed lower concentrations of HDL cholesterol and increased triglycerides $p < 0.05$. The allele frequency of IRS2 gene (G = 0.74 and D = 0.26) was found in Hardy Weinberg equilibrium in the population. The GD allele frequency was 43.1% in obese and 22.3% in the control. The wild-type GG, which has 45.2% of obese and 76.6% of the control. The DD homozygote was identified in 11.7% of cases and 1.1% of the control group. The G1057D variant was associated with obesity with an OR = 2.41 (95% CI 1.29-4.50). Conclusion metabolism between groups of young normal weight and obese is different. The G1057D IRS2 gene polymorphism is very common among young obese is statistically associated with it. This mutation is in equilibrium between the study populations. Therefore we can say that this mutation may be responsible for cases of obesity in young people having a unhealthy lifestyle.

815M

Revealing the detailed MHC implication in seven common diseases from the WTCCC by HLA imputation. *N. Vince[1,2], A. Bashirova[1,2], G. Nelson[1], M. Carrington[1,2].* 1) Cancer and Inflammation Program, Laboratory of Experimental Immunology, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD; 2) Ragon Institute of MGH, MIT and Harvard, Cambridge, MA.

The Wellcome Trust Case Control Consortium (WTCCC) comprise a study of 7 complex human diseases each of which comprise approximately 2000 individuals: bipolar disorder (BD), coronary artery disease (CAD), Crohn's disease (CD), hypertension (HT), rheumatoid arthritis (RA), type 1 diabetes (T1D), type 2 diabetes (T2D). The genome wide association study (GWAS) published in 2007 showed various associations across the genome, and particularly strong associations within the MHC for RA and T1D. Most HLA alleles are not efficiently tagged by any GWAS SNPs, and therefore, HLA allelic effects can be missed in SNP analyses. To test for allelic effects, imputation approaches have been recently developed. We used the SNP2HLA software to impute the HLA class I and class II alleles in the 7 disease cohorts. Imputed HLA class I data was further applied to predict HLA-C level of expression. Each of the 7 disease cohorts were compared to 3000 controls from the WTCCC, all with UK ancestry. Logistic regression with stepwise selection as a statistical model was used to ascertain the independency of each association. As expected RA and T1D show the strongest association with HLA: HLA-DQA1*0501 (P=7.1E-15, OR=0.43) and HLA-DQB1*0301 (P=7.2E-26, OR=0.18), respectively. In addition, HLA-DRB1*0103 shows a strong association with CD (P=9.0E-13, OR=3.51). The other 4 studied diseases have less significant associations with HLA, which need to be replicated elsewhere for confirmation: HT, HLA-DRB1*1101 (P=0.002, OR=1.55); BD, HLA-DPB1*0101 (P=0.002, OR= 0.72); T2D, HLA-B*3701 (P=0.003, OR=0.54); CAD, HLA-A*0205 (P=0.005, OR=0.44). Apart from the HLA alleles, the HLA-C expression level associates with CD: P=4.3E-6, OR=1.39. Thus, available GWAS data can be used for imputing HLA alleles and exploring the HLA allelic effects on human diseases.

816T

A polymorphism in the peptidyl arginine deiminase type IV gene (*PADI4*) is associated with radiographic joint destruction in patients with rheumatoid arthritis who are negative for anti-citrullinated peptide antibody (ACPA). *K. Ikari, S. Yoshida, K. Yano, A. Taniguchi, H. Yamanaka, S. Momohara.* Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan.

Rheumatoid arthritis (RA) is a complex polygenic disease of unknown etiology and is characterized by progressive joint damage. Anti-citrullinated peptide antibodies (ACPA) are the most specific autoantibody for RA. Genetic polymorphisms in the *PADI4* gene, encoding the citrullinating enzyme peptidyl arginine deiminase 4, have been associated with susceptibility to RA in several populations. *PADI4* polymorphisms have also been reported to be associated with joint damage in patients with RA. In the present study, we investigated whether a polymorphism in *PADI4* is associated with radiographic joint destruction in ACPA-negative patients with RA.

This study used DNA samples from 122 Japanese ACPA-negative patients with RA who satisfied the American College of Rheumatology 1987 revised criteria for RA. Most of patients were female (81.1%), 51.6% were rheumatoid factor (RF) positive, and the mean age was 55 years. We used the Sharp/ van der Heijde score (SHS) of the hands at a 5-year disease duration to measure joint damage. Single nucleotide polymorphism (SNP) rs2240340 was selected for the study because it has the best evidence of association with RA in the Japanese population. Genotyping was performed by using a TaqMan assay. The genetic risk of joint damage associated with rs2240340 was assessed by multiple regression analysis adjusted for the possible genetic risk associated with the shared epitope (SE) alleles of the class II, DR beta 1 major histocompatibility complex gene (HLA-DRB1). These alleles are thought to be associated with joint damage in RA patients.

The *PADI4* SNP was significantly associated with radiographic joint destruction in the Japanese population (P = 0.0395). The result remains significant when the analysis was adjusted by RF status (cutoff = 15.0 IU/ml) with HLA-DRB1 SE alleles (P = 0.0287). A SNP in *PADI4* may contribute to joint destruction in ACPA-negative patients with RA. The *PADI4* gene is likely to play a role in the disease progression of RA in addition to its role in the formation of ACPA. The results of this study provide important information about the risks of progressive joint damage in patients with RA.

817S

Risk for nonsyndromic cleft lip and palate from rare coding variants. *K. Asrani[1], W. Yang[2], J. Rine[1], E. Lammer[3], G. Shaw[2], N. Marini[1].* 1) California Institute for Quantitative Biosciences, University of California, Berkeley, CA; 2) Department of Pediatrics, Stanford University Medical Center, Stanford, CA; 3) Childrens Hospital Oakland Research Institute, Oakland, CA.

Cleft lip and/or cleft palate are common craniofacial malformations with complex and heterogeneous etiologies, reflecting both genetic and environmental factors. Both linkage and association studies have shown multiple genes influence risk for clefts. Murine models of clefts as well as studies on normal lip/palate development suggest additional genetic loci. However, causal variants have yet to be identified. Association studies, whether case-control or family-based, have little power to detect unknown rare variants that may be causal. Thus, our goal was to identify rare, potentially causal variants among a list of rationally chosen candidate genes. To this end, we sequenced the exons of 51 candidate genes in 322 ethnically diverse cases and 150 non-malformed controls. 31 target genes are involved in folate metabolism and comprise nearly all cellular enzymes that utilize a folate cofactor. 20 additional genes have been implicated in disease risk by murine models of clefting (e.g. WNT9B), studies of normal craniofacial development (e.g. BMPR1B), or by multiple association studies (e.g. IRF6). Our findings indicate that rare, protein-altering variants (missense, nonsense, frame-shift) confer risk for nonsyndromic clefts. First, the aggregate burden of rare, coding alleles was considerably higher in cases than controls. Overall, cases were 33% more likely to harbor a rare, protein-altering change in 51 genes (case mean = 1.52 mutations/individual: control mean = 1.15 mutations/individual). Second, mutations that are likely to result in functional impairment (nonsense, frame-shift) were significantly enriched among cases. 15 of 322 cases (~5%) harbored one such change within this set of 51 genes, whereas no such changes were found in controls (0 of 150; P-value = 0.01). Third, in gene-level analyses, the burden of rare alleles showed a strong case bias for several genes/gene regions previously implicated in cleft risk. For example, within the BHMT/BHMT2/DMGDH gene cluster on chromosome 5, 41 rare alleles were found in the case population versus 9 in controls (OR=2.3/ 95% CI = 1.1-4.9). Other loci with stronger allele burdens in cases were WNT9B, BMP4 and BMPR1B. Thus, we conclude that rare coding variants may confer risk for isolated clefts. Additional rare and common variant association analyses on this dataset will be discussed.

818M

Screening of CDH1 mutations in a Brazilian sample of nonsyndromic cleft lip / palate individuals. *L.A. Brito, C. Malcher, G. Yamamoto, S.G. Ferreira, M. Naslavsky, M. Aguena, M. Zatz, M.R. Passos-Bueno.* Human Genome Research Center, Institute of Biosciences, University of São Paulo, Brazil.

The epithelial cadherin is a calcium-dependent cell adhesion molecule, essential to the formation of adherens junction. Mutations in its encoding gene, CDH1, have been largely associated with different types of cancer, including gastric, lobular breast, endometrial, ovarian and prostate cancer. Recent studies have suggested a potential involvement of mutations in this gene with cleft lip with or without palate (CL/P) in individuals without known history of cancer in the family. Nonsyndromic CL/P is a complex trait, and, although many associations with common variants have been reported, they confer small risk and do not explain the whole heritability of the disease. To investigate the role of rare (<1%) CDH1 variants in susceptibility to nonsyndromic CL/P in the Brazilian population, we analyzed the coding regions and exon-intron boundaries of CDH1 in 199 affected individuals (sequenced by Sanger method) and 609 controls (from our in-house control exome database - sequenced with HiScan SQ - Illumina Inc; mean coverage of 100x). We found a total of 15 non-described variants in our affected group, 11 of those were not present in our control sample: 2 missense, 1 sinonymous, 6 intronic (near exon boundaries), 1 upstream and 1 in 3'UTR. The missense variants (p.D254N and p.R784H) were found in familial cases of CL/P, and are potentially causal, since they are segregating in the remaining affected family affected members and are predicted to damage the protein function, according to in silico analyses (scores for both variants: Polyphen-2 = 1.0, Probably damaging; SIFT <0.023, Damaging; PROVEAN <-4.36, Deleterious). Kernel association test (SKAT), performed only with rare variants in the CDH1 regions covered by both methods (Sanger and exome), did not detect statistical difference between affected and unaffected groups (p=0.07). Our results support the involvement of rare CDH1 variants in nonsyndromic CL/P etiology; however, collectively, CDH1 variants do not seem to be a common cause of CL/P in the Brazilian population, since no association was detected. In addition, since the carriers of the two missense mutations belong to familial cases of CL/P, we encourage the mutational screening of CDH1 mutations only in families segregating nonsyndromic CL/P. FAPESP, CNPq.

819T

Microsatellite (AT)n in the 3' UTR of the CTLA4 gene is associated with coronary artery aneurysm of Kawasaki disease. *H. Chi[1,2,3,4], MR. Chen[1,3], NC. Chiu[1,3], FY. Huang[1,7], TN. Huang[1,2], WF. Chen[5], CL. Lin[5], YJ. Lee[1,5,6,7].* 1) Pediatrics, Mackay Memorial Hosp, Taipei, Taiwan; 2) Graduate Institute of Clinical Medicine, National Taiwan University College of Medicine; 3) Department of Nursing, Mackay Junior College of Medicine, Nursing, and Management; 4) Department of Medicine, Mackay Medical College; 5) Department of Medical Research, Mackay Memorial Hospital Tamsui District; 6) Institute of Biomedical Sciences, Mackay Medical College; 7) School of Medicine, Taipei Medical University.

Background Kawasaki disease (KD) is a febrile disease of children complicated with vasculitis of large coronary arteries and potential aneurysm formation. The *CTLA4* gene encodes a T-cell receptor, cytotoxic T-lymphocyte-associated protein 4 (CTLA4), which involved in controlling the proliferation and apoptosis of T lymphocytes. A microsatellite (AT)n repeat in the 3' untranslated region has been reported to be associated with the organ-specific autoimmune disorders in several racial groups. CTLA4 expression is increased during the acute stage of KD and significantly higher than at the convalescent stage. This suggests that CTLA4 may be important in pathogenesis of KD or CALs. We investigated whether the *CTLA4* gene was associated with CALs in Han Chinese children with KD. **Material and methods** The patients were 576 unrelated children (339 boys, 237 girls) with KD. Their age at diagnosis was 1.9 ± 1.7 years (range 0.1 - 10.2 years). We typed 3'UTR (AT)n microsatellite using fluorescence-based methods. Polymerase chain reaction (PCR) primers were 5'-GCCAGTGATGC-TAAAGGTTG -3' (forward) and 5'-ACACAAAAACATACGTGGCTC- 3' (reverse). The forward primer was labeled with fluorescent dyes. PCR products were electrophoresed on an ABI 3730 DNA analyzer and their size was analyzed with GENESCAN 4.0. We designate the alleles according to the number of repeat units. **Statistical analysis** Statistical difference in allele distribution between patients with CALs and those without CALs were assessed by the chi-square test. Odds ratios and 95% confidence intervals were also calculated. Only those alleles of >2.0% in frequency in either patients or control were compared. The Bonferroni correction, $Pc = 1 - (1-P)n$, was used for multiple comparisons where *Pc* is the corrected *P* value, *P* the uncorrected value and n is the number of comparisons. In this study, n is 8 for each allele. A *Pc* value of less than 0.05 was considered statistically significant. **Results** The PCR products were 95-141 base pair (bp) in length corresponding to alleles 8-31. Alleles 16, 8, 15, and 17 were major alleles with a frequency of >10%. Allele 8 was 79/400 (19.8%) in patients with CALs and significantly less frequent compared to that in patients without CALs which was 210/752 (27.9%); OR= 0.64 (0.47-0.85), $P = 0.0023$, $Pc = 0.018$. **Conclusions** The microsatellite (AT)n in the 3'UTR of the CTLA4 gene is associated with CALs in KD. Allele 8 conferred protection against the formation of CALs in patients with KD.

820S

Interaction between *PTPN2* and HLA-DRB1 SE alleles in rheumatoid arthritis. *M. Houtman, K. Shchetynsky, L. Padyukov.* Rheumatology Unit, Department of Medicine Solna, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden.

Background: One of the most common chronic autoimmune disorders worldwide is rheumatoid arthritis (RA). It has a strong genetic component and over 100 risk loci have been discovered in genome-wide association studies (GWAS). The major risk factor for RA is HLA-DRB1 shared epitope (SE) alleles. Outside of this region, a recently identified candidate gene for RA is protein tyrosine phosphatase non-receptor type 2 (*PTPN2*), a suggested phosphatase in cytokine receptor signaling. However, the functional consequences of genetic variations in the *PTPN2* region remain undefined. We therefore aimed to identify additional susceptibility SNPs in *PTPN2* through the analysis of statistical interactions between SNPs in *PTPN2* and HLA-DRB1 SE alleles in development of RA. **Methods:** Genotypes from two independent cohorts were used in this study, the Swedish EIRA study (3151 patients with RA and 2247 matched healthy controls) and NARAC (873 patients with RA and 1196 controls). Interactions were investigated by calculating the attributable proportion due to interaction (AP) between 11 SNPS in the *PTPN2* region and HLA-DRB1 SE alleles. **Results:** We found *PTPN2* SNPs rs657555 and rs11080606 in statistical interaction with HLA-DRB1 SE alleles in autoantibody positive RA in the two independent cohorts EIRA and NARAC (AP = 0.208 [0.071-0.345] and AP = 0.209 [0.019-0.399], respectively). When individuals from the EIRA study were analyzed only *PTPN2* SNP rs657555 demonstrated association with RA ($p = 6.13\times10^{-5}$; OR = 1.23 [1.11-1.37]). This is an intronic SNP in region between exons 2 and 3 that does not influence directly known splicing elements. **Conclusion:** Through the analysis of statistical gene-gene interactions between SNPs in *PTPN2* and HLA-DRB1 SE alleles, we identified a new candidate SNP in the *PTPN2* region that distinguish a subgroup of RA cases from healthy controls. The potential link between *PTPN2* and SE in RA should be further investigated in functional studies in relation to different autoimmune diseases.

821M

Genetic variants of SMADs in the TGF-β/SMAD signal pathway are related specifically to susceptibility to ulcerative colitis in Japanese patients. *T. Inamine[1], S. Suzuki[1], A. Yamashita[1], S. Fukuda[1], S. Kondo[1], H. Isomoto[2], K. Tsukamoto[1].* 1) Dept Pharmacotherapeutics, Nagasaki Univ Grad Sch, Nagasaki, Japan; 2) Dept Gastroenterol and Hepatol, Nagasaki Univ Grad Sch, Nagasaki, Japan.

PURPOSE Inflammatory bowel diseases (IBD), comprised of Crohn's disease (CD) and ulcerative colitis (UC), is attributed to inappropriate inflammatory response in the intestinal epithelia of patients. TGF-β/SMAD signals play a key role in differentiation of naive CD4+ T cells to Th17 cells or regulatory T (Treg) cells. In order to identify genetic determinants of IBD, we investigated an association between susceptibility to IBD and SMADs polymorphisms in the Japanese population. METHODS The study subjects consisted of 108 patients with UC, 81 patients with CD, and 199 unrelated healthy control subjects. A total of 21 tag single nucleotide polymorphisms (SNPs) in four genes (*SMAD2, SMAD3, SMAD4,* and *SMAD7*), which are involved in the TGF-β/SMAD signal pathway, were genotyped by PCR-restriction fragment length polymorphism, -direct DNA sequencing, or -high resolution melting curve analysis. The frequencies of alleles and genotypes were compared between control subjects and UC patients or CD patients by chi-square test or Fisher's exact test in three inheritance models: the allele, the minor allele dominant, the minor allele recessive models. RESULTS Five SNPs (rs13381619, rs9955626, rs1792658, rs1792684, and rs1792671) of *SMAD2*, rs4147358 of *SMAD3*, two SNPs (rs7229678 and rs9304407) of *SMAD4*, and rs12956924 of *SMAD7* showed the significant association with susceptibility only to UC. In each gene, rs13381619 of *SMAD2*, rs4147358 of *SMAD3*, rs9304407 of *SMAD4*, and rs12956924 of *SMAD7* showed the strongest association ($P = 0.0007$ in the minor allele dominant model, 0.0205 in the minor allele dominant model, 0.0050 in the minor allele recessive model, and 0.0011 in the minor allele recessive model, respectively). In addition to UC, rs1792658 of *SMAD2* also showed the significant association with susceptibility to CD ($P = 0.0409$ in the minor allele recessive model). CONCLUSION Genetic variants of SMAD signaling molecules in the TGF-β/SMAD pathway may disturb the signal transduction and alter the balance of differentiation to Th17 and Treg, leading to dysregulation of immune response and eventually resulting in the development of IBD, especially UC. However, a genome-wide association study in European ancestry has shown the association of SMAD3 with susceptibility to CD. Although the TGF-β/SMAD signal pathway is crucial in the etiology of IBD, genetic variants of the key SMAD molecules in this pathway may be different between Caucasian and Japanese patients.

822T

Associations of BST2 Polymorphisms with HIV-1 Acquisition in African American and European American People who Inject Drugs. *E.O. Johnson[1], D.B. Hancock[1], N.C. Gaddis[2], N.L. Saccone[4], L.J. Bierut[3], A.H. Karl[5].* 1) Behavioral Hlth Epidemiology, RTI International, Research Triangle Park, NC; 2) Research Computing Division, RTI International, Research Triangle Park, NC; 3) Dept of Psychiatry, Washington University School of Medicine in St. Louis, MO; 4) Dept of Genetics, Washington University School of Medicine in St. Louis, MO; 5) Urban Health Program, RTI International, San Francisco, CA.

An estimated 50% of the variability in acquiring human immunodeficiency virus type I (HIV-1) upon exposure is attributable to host genetic factors. The only genetic polymorphism conclusively associated with risk of acquiring HIV-1, a low frequency 32-base pair deletion in the chemokine (C-C motif) receptor 5 (CCR5) gene, accounts for little of the population variability. Identification of other genetic polymorphisms underlying HIV-1 susceptibility is needed to better understand disease pathogenesis. The BST2 gene encodes a host restriction factor that suppresses the release of HIV-1 particles by tethering them to the cell surface based on in vitro studies. However, little is known about the influence of BST2 polymorphisms on HIV-1 acquisition and disease progression in vivo. In this study, we tested a dense set of polymorphisms spanning BST2 for association with HIV acquisition and viral load using 955 HIV antibody negative cases and 2,181 HIV antibody positive controls from the Urban Health Study of European American and African American people who inject drugs. We tested 470 single nucleotide polymorphisms (SNPs) and insertions/deletions (indels), genotyped or imputed from 1000 Genomes across BST2 and its flanking regions. The novel SNP rs113189798 was associated with HIV acquisition, exceeding the corrected P value threshold for multiple testing. The rs113189798-G allele, which occurred at frequencies of 16% in AAs and 4% in EAs, was associated with reduced risk of HIV-1 acquisition across both ancestry groups (meta-analysis $P=1.43\times10^{-4}$): odds ratio (95% confidence interval) of 1.22 (1.01-1.49) in AAs and 2.17 (1.43-3.33) in EAs. This SNP is located 17.8 kb downstream of BST2. The SNP rs12609479, which was previously reported for affecting BST2 expression but not previously linked to HIV acquisition, was nominally associated with HIV-1 acquisition in the UHS (meta-analysis P=0.036 across the ancestry groups). The rs12609479-A allele is predicted to increase BST2 expression and subsequently decrease risk of acquiring HIV-1. Rs113189798 and rs12609479 are only weakly correlated (r^2=0.2-0.4), and represent distinct association signals. No significant associations were found for HIV viral load. Our findings provide support to BST2 as a genetic susceptibility factor for HIV-1 acquisition: identifying a novel SNP association for rs13189798 and linking the previously reported regulatory SNP rs12609479 to risk of acquiring HIV-1.

823S

Assessing genetic association between RASGRP3 and SLE susceptibility. *X. Kim-Howard[1], C. Sun[1], A. Adler[1], H. Zhang[2], L.H. Lian[3], K.H. Chua[4], S.-C. Bae[5], S. Nath[1].* 1) Oklahoma Med Res Foundation, Oklahoma City, OK; 2) Peking University First Hospital, Beijing, China; 3) Department of Molecular Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia; 4) Department of Biomedical Science, University of Malaya, 50603 Kuala Lumpur, Malaysia; 5) Hanyang University Hospital for Rheumatic Diseases, Seoul 133-791, Korea.

Systemic lupus erythematosus (SLE) is a chronic, multisystem autoimmune disease with diverse clinical manifestations. Despite evidence for a distinct genetic architecture of SLE in Asians, little research has investigate the full spectrum of genetic variants that influence Asian SLE. Recent GWAS in Han Chinese identified association between SLE and RASGRP3 (rs13385731). RASGRP3 is responsible for Ras-ERK signaling mediated by B-cell receptor ligation in B cells and is involved in immunoglobulin production. It is not known whether rs13385731 is the primary variant responsible for SLE association with RASGRP3, or if it is in linkage disequilibrium with another causal variant. Objectives of this study were to: (a) perform comprehensive analysis using dense fine-mapping in 3 Asian populations, (b) identify robust, independent SLE-predisposing variants, and (c) test whether these variants are associated in ethnically diverse populations. Genotype data from 125 SNPs within RASGRP3 (chr2: 33,656,416-33,794,798 bp) were extracted from the ImmunoChip array, on 2487 cases and 3958 controls from Korean (KR), Han Chinese (CH), and Malaysian Chinese (MC) cohorts. To increase statistical power we performed ethnicity-specific imputation-based association analysis. For replication in other ancestries the most significant SNPs were confirmed by TaqMan assay in European-Americans (EA), African-Americans (AA), and Hispanics (HS). We also assessed association between significant SNPs and SLE clinical sub-phenotypes. After imputation we analyzed 162 SNPs from RASGRP3. The most significant association was at rs13425999 (P(META)=7.5E-8; P(KR)=9.8E-5; P(CH)=1.0E-3; P(MC)=0.03). SLE association with rs13385731 was confirmed (P(KR)=3.2E-4; P(CH)=1.7E-3; P(MC)=0.03; P(META)=4.5E-7). Published rs13385731 was significantly associated in EA and AA (P(EA)=2.1E-3; P(AA)=2.2E-3) but not HS (P(HS)=0.06); rs13425999 was significant in (P(EA)=1.0E-4) but not AA or HS. In KR rs13425999 was collinear with rs13385731 (313 bp apart), either SNP explained association between SLE and RASGRP3. Thus, we replicated RASGRP3 association with SLE in KR, HC, MC and EA,. Preliminary bioinformatic analysis indicate that rs1342599 is located in an open chromatin, attached with several histone marks (H3K27Ac, H3K4Me1), suggesting that it is a likely functional SNP.

824M

Polymorphism in MEN-1 gene is associated with increased risk and earlier age of pituitary adenoma development. *J. Klovins[1], R. Peculis[1], I. Balcere[2], V. Pirags[1,2].* 1) Genome Centre, Latvian Biomedical Research and Study Centre, Riga, Latvia; 2) University of Latvia.

Clinically significant pituitary adenomas affect one individual out of approximately 1000 to 1300 people in general population. So far mutations in number of genes responsible for familial cases of PA contributing up to 5% of all tumors have been described. Genetics of cases not clearly attributed to familial adenoma is unclear with multiple genes potentially involved. In this study, we analyzed 96 tag-SNPs from seven genes known for their involvement in familial PA cases and 3 receptor genes (AIP, SSTR2, SSTR5, DRD2, GNAS, MEN1 and PRKar1a) including their neighboring regions in 143pituitary adenoma patients and 354 age and sex matched controls (1:2.5) in order to investigate potential involvement of these genes in determination of the risk for non-familial pituitary adenomas. We identified one rare SNP in MEN1 locus strongly associated (P=2.1e-4; OR =17.8; CI95[2.18-145.5]) with increased risk of PA development. The same SNP was also associated with younger age at diagnosis of PA (beta=-17.4). We also confirmed previously found association of SNP in SSTR5 gene and two nominal associations of SNPs in DRD2 gene with either risk of PA development or clinical characteristics of the disease. We thus have showed that SNP in MEN1, the gene known to be responsible for familial pituitary tumors also plays role in sporadic cases.

825T

Targeted resequencing of CFH-CFHR genes identifies new putative functional common and rare variants conferring susceptibility to Meningococcal disease. *V. Kumar[1], Z.Y. Phua[1], T.W. Kuijpers[3], F. Martinon-Torres[4,5], A. Salas[6,7], M.L. Hibberd[8], E.D. Carrol[9], W. Zenz[10], M. Levin[2], S. Davila[1].* 1) Human Genetics, Genome Institute of Singapore, Singapore; 2) Division of Infectious Diseases, Department of Medicine, Imperial College London; 3) Division of Pediatric Hematology, Immunology and Infectious diseases, Emma Children's Hospital Academic Medical Center, Amsterdam, The Netherlands; 4) Pediatric Emergency and Critical Care Division, Department of Pediatrics, Hospital Clínico Universitario de Santiago, Santiago de Compostela, Spain; 5) Grupo Gallego de Genética, Vacunas e Investigación Pediátrica, Instituto de Investigación Sanitaria de Santiago, Galicia, Spain; 6) Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses Facultade de Medicina, Universidade de Santiago de Compostela, Santiago de Compostela, Galicia, Spain; 7) Instituto de Medicina Legal, Facultade de Medicina, Universidade de Santiago de Compostela, Santiago de Compostela, Galicia, Spain; 8) Infectious Diseases, Genome Institute of Singapore, Singapore; 9) Institute of Child Health, University of Liverpool, Alder Hey Children's National Health Service Foundation Trust, Liverpool, UK; 10) Department of General Pediatrics, Medical University of Graz, Graz, Austria.

Meningococcal disease (MD) is an infection caused by *Neisseria meningitidis*. In recent years, genome wide association studies (GWAS) have suggested the involvement of a number of genes in determining host susceptibility and progression to disease. For instance, variants within the CFH-CFHR region have been robustly shown to be associated with susceptibility to disease. Yet the putative functional variants within this genomic region remain to be elucidated. Moreover, this reported association accounts for a small fraction of heritability and it is plausible that rare variants with moderate penetrance could contribute to disease susceptibility as well. CFH and CFHR genes share sequences of high homology within chromosome 1 that has made extremely challenging its analysis. Using a Nimblegen capture design we have been able to cover more than 92% of the targeted region (~300 kilobases). We sequence 238 MD cases and 237 controls from Western Europe and perform a case-control analysis. Calling of 3,032 high-quality single nucleotide polymorphism (SNP) with less than 10% sample missingness rate was obtained. Within the CFH region we have identified a SNP (in linkage disequilibrium with the top GWAS SNP, rs1065489) exceeding original GWAS P value. In addition we have identified a rare variant with P value $< 10^{-5}$ showing a genetic effect on the opposite direction to that of GWAS SNP. We are currently genotyping these SNPs in 1,300 MD cases and 3,000 controls to validate our results.

826S

Candidate-gene association study of sciatica. *S. Lemmelä[1], S. Solovieva[1], R. Shiri[1], M. Heliövaara[2], J. Viikari[3], O.T. Raitakari[4,5], T. Lehtimäki[6], E. Viikari-Juntura[1], K. Husgafvel-Pursiainen[1].* 1) Health and Work Ability, Finnish Institute of Occupational Health, Helsinki, Finland; 2) Population Health Unit, National Institute for Health and Welfare, Helsinki, Finland; 3) Department of Medicine, University of Turku, and Division of Medicine, Turku University Hospital, Turku, Finland; 4) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 5) Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; 6) Department of Clinical Chemistry, Fimlab Laboratories, University of Tampere School of Medicine, Tampere, Finland.

Sciatica is a common and often disabling low back disorder (LBD) in working-age populations. It presents as pain radiating from the back to the leg, usually caused by compression or irritation of one of the lumbosacral nerve roots. It is a complex disorder with relatively high heritability (35-75%) but poorly understood molecular mechanisms. Candidate gene studies have mostly focused on genes associated with LBD, such as those related to cartilage structure and stability, pain signaling, obesity or inflammation[1]. We investigated 11 SNPs in genes previously associated with LBDs (sciatica, lumbar disc degeneration [LDD] or low back pain), osteoarthritis or nicotine addiction (smoking is a known risk factor for LBD). The study comprised two large Finnish population cohorts; Young Finns Study (YFS; 171 sciatica cases, 1777 controls) and Health 2000 (H2000; 294 sciatica cases, 5296 controls). Four candidate gene SNPs were genotyped in both populations; rs1107946 of COL1A1, rs2294995 of COL9A3, rs7775 of FRZB and rs1317286 of CHRNA3. Seven additional SNPs were genotyped in YFS; rs61734651 of COL9A3, rs1799907 of COL11A2, rs1800587 of IL1A, rs1800796 of IL6, rs731236 of VDR, rs288326 of FRZB, rs16969968 of CHRNA5 as well as rs2187689 in HLA gene region in H2000. Genotyping was done by qPCR using Taqman assays or OpenArray system. Allele and genotype frequency differences between cases and controls were estimated by Pearson's X2-test with Yates' continuity correction in R program. Odds ratios with 95% confidence intervals were estimated using R. We found an association between rs2187689 on HLA region (6p21.32) and sciatica in H2000 (p=0.04; OR=1.31, 95%CI 1.02-1.69). This SNP was selected due to its previous association with LDD in a GWAS of Northern Europeans[2]. None of the other SNPs were significantly associated with sciatica. Further analysis taking into account work-related and lifestyle factors will be carried out. Current findings, based on relatively large population cohorts, are in line with earlier studies in suggesting that common candidate genes may have a more limited role alone but a wider range of genes each with a subtle effect as well as gene-environment interaction are likely involved. GWA studies and exome/whole-genome sequencing approaches may aid in identifying novel candidate loci for low back disorders, including sciatica. [1]Eskola, Lemmelä et al., PLoSOne 7(11):e49995, 2012. [2]Williams et al., Ann Rheum Dis 72(7):1141-8, 2013.

827M

Association study between NOD2 and CCDC122-LACC1 genes and leprosy in Brazilians. *C.S. Marques[1], H. Salomão[2], V.M. Fava[2], L.E.A. Arnez[1], E.P. Amaral[3], C.C. Cardoso[1,4], I.M.F. Dias-Batista[5], W.L. Silva[5], P. Medeiros[5], M.C.L. Virmond[5], F.C.F. Lana[3], A.G. Pacheco[6], M.O. Moraes[1], M.T. Mira[2], A.C. Pereira[5].* 1) Leprosy Laboratory, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Rio de Janeiro, Brazil; 2) Core for Advanced Molecular Investigation, Graduate Program in Health Sciences, School of Medicine, PUCPR, Curitiba, Brazil; 3) Departamento de Enfermagem Materno-Infantil e Saúde Pública Federal de Minas Gerais, UFMG, Brazil; 4) Laboratório de Virologia Molecular, Departamento de Genética, UFRJ, Rio de Janeiro, Brazil; 5) Instituto Lauro de Souza Lima, Bauru, Brazil; 6) Programa de Computação Científica, FIOCRUZ, Rio de Janeiro, Brazil.

Leprosy is a complex disease with phenotypes strongly influenced by genetic variations. Previously, a Chinese genome-wide association study (GWAS) identified novel genes and pathways associated with leprosy susceptibility, which was only partially replicated by independent studies in different ethnicities. The aim of our study was to perform a validation and replication study of the Chinese GWAS in Brazilians, using a stepwise strategy that involved two family-based and three independent case-control samples, totalizing 3,614 individuals enrolled. We selected 36 tag SNPs at five candidate genes from the Chinese study (CCDC122-LACC1, NOD2, TNFSF15 and RIPK2) which were tested in the discovery sample from Prata Village, a former leprosy colony. The DNA samples were genotyped using allelic discrimination Real Time PCR (TaqMan assay, Applied Biosystems). The family-based association analysis was performed using Transmission Disequilibrium Test (TDT) in the FBAT software, version 2.0.2. In case-control studies, comparative analyses for allelic, genotype and carrier frequencies were performed using an unconditional logistic regression model in R environment. Also, overall analysis combining the case-control samples was performed controlling for geographic region, gender and ethnicity. Linkage disequilibrium was estimated by Haploview software, version 4.2. In the discovery sample we observed an association between leprosy and tag SNPs at NOD2 (rs8057431-A) and CCDC122-LACC1 (rs4942254-C) alleles, both under-transmitted to affected offspring, indicating protection to leprosy. After, these associations were consistently detected in all replication populations from Brazil. The combined analysis showed the following Odds Ratios towards leprosy: rs8057431-AA (OR= 0.49, P= 1.39e-06) and rs4942254-CC (OR= 0.72, P = 0.003). These results indicate an association between NOD2 and CCDC122-LACC1 genes with leprosy protection in Brazilians, and suggest them as important markers to this disease across diverse populations.

828T

The role of SIRT2 in human longevity: converging evidence from gene expression, epigenetics and genetic variation. *D.R. Mazzotti[1], C. Guindalini[1], W.A.S. Moraes[1], M.L. Andersen[1], R. Pellegrino[1,6,7], L.R.A. Bittencourt[1], B. Borsatto-Galera[2], M.A.C. Smith[3], M.S. Cendoroglo[4], L.R. Ramos[5], S. Tufik[1].* 1) Deparamento dePsicobiologia, Universidade Federal de São Paulo, São Paulo, Brazil; 2) Departamento de Ciências Básicas em Saúde, Faculdade de Ciências Médicas, Universidade Federal do Mato Grosso Cuiabá, Brazil; 3) Departamento de Morfologia e Genética, Universidade Federal de São Paulo, São Paulo, Brazil; 4) Disciplina de Geriatria e Gerontologia, Universidade Federal de São Paulo, São Paulo, Brazil; 5) Departamento de Medicina Preventiva, Universidade Federal de São Paulo, São Paulo, Brazil; 6) Center for Applied Genomics - The Children's Hospital of Philadelphia, Philadelphia, United States; 7) Center for Sleep and Circadian Neurobiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, United States.

The increase in life expectancy indicates that aging at the population level is unavoidable. The identification of factors that may help the development of actions to promote healthy aging is fundamental. Using an integrative approach, we aimed to identify whole blood oxidative stress gene expression changes and associated epigenetic mechanisms as well as genetic variants associated with human longevity. For the gene expression and epigenetic study, oldest old individuals (N=10, 85 to 105 years old), older adults (N= 13, 60-70 years old) and young adults (N=15, 20-30 years old) had oxidative stress and antioxidant defense related gene expression and microRNA expression evaluated by qRT-PCR, and DNA methylation by bisulfite sequencing. For the association study, 2025 individuals (20-105 years) were genotyped for 9 polymorphisms within differentially expressed genes. We found significant up-regulation of SIRT2 expression in the oldest old individuals (fold-change=1.438), accompanied by lower DNA methylation rate in specific CpG sites and down-regulation of miR-1275, a microRNA predicted to regulate SIRT2 expression. Strong and significant associations between two SIRT2 polymorphisms (rs10410544 and rs4802998) and longevity (individuals >90 years-old) were found (OR=2.577; 95%CI=1.472-4.510 and OR=2.365; 95%CI=1.281-4.366, respectively and adjusted for gender and genetic ancestry proportions). We described SIRT2 as a novel marker whose fully characterization is essential for understanding the mechanisms behind molecular regulation of healthy aging and longevity in humans.

829S

Association of IL10 variants with visceral leishmaniasis in Indian population. *A. Mishra[1,2], S. Nizammuddin[1], G. Arekatla[1], S. Prakash[1], N. Tupperwar[1], K. Thangaraj[1].* 1) Centre for Cellular and Molecular Biology, Hyderabad, Andhra Pradesh, India; 2) (DBT-RA scheme) Department of Biotechnology, Govt. of India.

Visceral leishmaniasis (VL) is a multi-factorial disease in which host genetics play a significant role in determining the disease outcome. T cells generate specific and memory T cells, in response to intracellular parasitic infections. The immunological role of anti-inflammatory IL10 cytokine, which is secreted by Th2 and group of cells, is well documented in several studies in parasite infections but its genetic role has not been attempted in Indian VL. VL patients display high level of IL10 in blood serum and elevated levels of mRNA. Therefore the aim of this study is to evaluate the role of IL10 gene in Indian VL and look for the distribution of disease associated allele / haplotype in diverse Indian populations. We sequenced all the exons and exon-intron boundaries of IL10 gene in 184 VL patients along with 172 ethinically matched control from VL endemic region of India. Our analysis revealed four variations namely rs1518111 (2195 A>G, intron), rs1554286 (2607 C>T, intron), rs3024496 (4976 T>C, 3' UTR) and rs3024498 (5311 A>G, 3' UTR), of these, rs3024498 is significantly associated with VL for genotype frequencies (AG, $\chi 2$ =15.18; p =0.00010, OR =0.376 and for GG, $\chi 2$ =12.97; p =0.00032, OR =0.369) and for allelic frequency ($\chi 2$ =18.87; p =0.00001, OR = 0.515). Further analysis of the above four variations in 1138 individuals from 34 ethnic populations of India, representing different social and linguistic group, showed variable frequency. Interestingly, we found less representation of VL (rs3024498) and high representation of leprosy (rs1554286) associated alleles in tribes compared to castes. Haploview analysis also showed strong to moderate LD variation in 29 Indian populations. Overall, our population data of rs3024498 and rs1554286 suggest, that tribal and lower caste population of India have protected genotype and allele from VL while, more susceptible allele for leprosy. This study has potential medical implication in pharmacogenomics and vaccination programs.

830M

SVEP1 c.2080A>C (p. Gln581His) gene is associated with altered mortality of septic shock. *T. Nakada[1,2], J. Russell[2], J. Boyd[2], S. Thair[2], E. Nakada[2], K. Walley[2].* 1) Department of Emergency and Critical Care Medicine, Chiba University Graduate School of Medicine, Chiba, Japan; 2) University of British Columbia, Critical Care Research Laboratories, Heart + Lung Institute, St. Paul's Hospital, Vancouver, BC, Canada.

Background: Septic shock is a leading cause of death in intensive care units. Genetic factors are associated with altered mortality of septic shock. However the key genetic variations have not been fully elucidated. To identify the key variations, we first identified non-synonymous single nucleotide polymorphisms (SNPs) in conserved genomic regions that are predicted to have significant effects on protein function. We then test the hypothesis that these variants across genome alter clinical outcome of septic shock. **Methods:** Septic shock patients (n=520, European ancestry) were genotyped for 843 non-synonymous SNPs, which we identified in conserved regions of the genome and are predicted to have damaging effects from the protein sequence using PolyPhen-2 and phastCons. The primary outcome variable was 28-day mortality. Productions of adhesion molecules including IL-8 GRO-alpha, MCP-1 and MCP-3 were measured in human umbilical vein endothelial cells (HUVECs) after SVEP1 gene silencing by RNA interference. We measured IL-8 levels in the plasma obtained on admission from the septic shock patients. **Results:** Of 843 non-synonymous SNPs, SVEP1 c.2080A>C (p. Gln581His, rs10817033) was significantly associated with altered 28-day mortality (Armitage trend test, uncorrected $P=3.8 \times 10^{-5}$, Bonferroni corrected P = 0.032). Patients with septic shock having the SVEP1 C allele of SVEP1 c.2080A>C had a significant increase in the hazard of death over the 28-day (hazard ratio 1.72, 95% confidence interval, 1.31 - 2.26, $P=9.7 \times 10^{-5}$) and increased organ dysfunction, and needed more organ support ($P < 0.05$). SVEP1 gene silencing increased IL-8, GRO-alpha, MCP-1 and MCP-3 in HUVECs under LPS stimulation ($P<0.01$). Non-survivors of the septic shock cohort had significantly increased IL-8 levels ($P=0.0005$). There was a trend toward increased plasma IL-8 with increasing C alleles of SVEP1 c.2080A>C, most notably among homozygous carriers, though this was not statistically significant (CC vs. AC vs. AA, P=0.22). SVEP1 c.2080A>C polymorphism alters amino acids in a conserved region in protein sequence and is predicted to be damaging. **Conclusions:** C allele of SVEP1 c.2080A>C (p. Gln581His), a non-synonymous SNP in conserved regions and predicted to have damaging effects on protein structure, was associated with increased 28-day mortality of septic shock. SVEP1 may involve the leukocyte adhesion pathway.

831T

Generalization and fine-mapping of CDKN2B-AS1 for primary open-angle glaucoma in African Americans from the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study. *N. Restrepo[1], R. Goodloe[1], E. Farber-Eger[1], D. Crawford[1,2].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Molecular Physiology & Biophysics, Vanderbilt University, Nashville TN.

Primary open-angle glaucoma (POAG) is the second leading cause of permanent vision loss and blindness in the U.S. Genome-wide and candidate gene association studies have identified several loci associated with POAG risk in populations of European-descent. African Americans are ~15 times as likely to develop permanent vision impairment from glaucoma vs. European Americans, yet few studies have been performed in this population. To begin to fill these gaps in knowledge, we have performed a generalization and fine-mapping study for one POAG-associated region, CDKN2B-AS1, in African Americans from the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study as part of the larger Population Architecture using Genomics and Epidemiology study. A total of 242 CDKN2B-AS1 SNPs were targeted for genotyping using the Metabochip. African American POAG cases (n=135) and controls (n=1,739) were identified in BioVU, the Vanderbilt University Medical Center DNA repository. Cases and controls were identified via a combination of International Classification of Diseases diagnostic codes, Current Procedural Terminology billing codes, and manual review of clinical records. We performed single SNP tests of association for common variants (MAF>0.05) using logistic regression assuming an additive genetic model adjusted for age and sex. We did not replicate associations previously described in European Americans for rs523096, rs564398, rs2157719, rs1412829, rs1063192 at a corrected threshold of 2.06×10-4. These SNPs have minor allele frequencies (MAF) ranging from 0.37-0.45 in HapMap CEU populations compared to 7-9% among EAGLE BioVU African Americans. In addition to striking MAF differences, we observed differences in linkage disequilibrium (LD) for this gene between HapMap CEU and ASW data, and mostly low levels of LD in EAGLE BioVU African Americans when plotted using LocusZoom. Although none of the associations tested survived multiple correction testing, several were nominally associated with POAG including six variants in a 37kbp region (Chr 9: 22,105,026-22,068,305) that were associated at p<0.01 with ORs ranging 1.40-1.44. An exception was a SNP at Chr. 9 (bp 22105026; MAF=0.42) that was associated with an OR of 0.70 (95% CI: 0.54-0.92). Although underpowered, this study highlights the need for further studies among African Americans to determine if lack of generalization is due to differences in LD or differences in the genetic risk factors for POAG.

832S

Associations between variants in motilin genes and infantile hypertrophic pyloric stenosis. *PA. Romitti[1,2], Y. Zhu[2], Y. Cao[2], DM. Kay[3], RJ. Sicko[3], S. Richardson[4], R. Fan[5], A. Liu[5], ML. Browne[4,6], CM. Druschel[4,6], T. Carter[5], M. Caggana[3], LC. Brody[7], JL. Mills[5].*1) Iowa Registry for Congenital and Inherited Disorders; 2) Department of Epidemiology, College of Public Health, The University of Iowa, Iowa City, IA; 3) Division of Genetics, Wadsworth Center, New York State Department of Health, Albany, NY; 4) Congenital Malformations Registry, New York State Department of Health, Albany, NY; 5) Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, DHHS, Bethesda, MD; 6) Department of Epidemiology and Biostatistics, School of Public Health, University at Albany, Rensselaer, NY; 7) Genome Technology Branch, National Human Genome Research Institute, NIH, DHHS, Bethesda, MD.

The onset (days to weeks following birth) and descriptive epidemiology (male and non-Hispanic White excess) of infantile hypertrophic pyloric stenosis (IHPS) suggest its etiology may comprise environmental and genetic factors. Maternal and infant use of erythromycin, a motilin agonist, has been associated with IHPS. We examined associations between variants in the motilin (*MLN*) and motilin receptor (*MLNR*) genes and IHPS. Given the predominance of IHPS among non-Hispanic Whites, we used the population-based New York State (NYS) Congenital Malformations Registry to identify non-Hispanic White cases with IHPS born 1998-2005. A sample of sex- and racial/ethnic-matched controls born during the same time period was randomly selected from unaffected NYS live births. Case and control data were linked with NYS Newborn Screening Program data to obtain residual newborn blood spots. DNA was purified from de-identified blood spots for cases (n=656) and controls (n=656) and genotyped for single nucleotide polymorphisms (SNPs), identified by haplotype tagging, in the *MLN* (12 SNPs) and *MLNR* (1 SNP) genes. Using unconditional logistic regression, we estimated odds ratios (ORs) and 95% confidence intervals (CIs), corrected for multiple comparisons (P<0.0038), between each SNP genotype (heterozygous or homozygous minor vs. homozygous major) and IHPS. Statistically significant increased associations were observed between both *MLN* rs1547668 (homozygous minor: OR=2.78; 95% CI=1.51-5.12, uncorrected P=0.0011) and *MLN* rs4713685 (homozygous minor: OR=2.10; 95% CI=1.38-3.21, uncorrected P=0.0006) and IHPS. Significantly reduced associations were observed between *MLN* rs1547669 (heterozygous: OR=0.69; 95% CI=0.54-0.89, uncorrected P=0.0037; homozygous minor: OR=0.56; 95% CI=0.41-0.76; uncorrected P=0.0003) and IHPS. No significantly increased or decreased associations were observed for the *MLNR* SNP and IHPS. Using this large population-based sample of European ancestry, we observed significantly increased associations in two *MLN* SNPs and significantly decreased associations in another *MLN* SNP and IHPS. The hormone encoded by the *MLN* gene is secreted by cells of the small intestine to regulate gastrointestinal contractions and motility. The associations identified for *MLN* SNPs provide insights into potential genetic susceptibilities for IHPS and support for the observed associations between maternal and infant erythromycin use and this defect.

833M

Intracranial Aneurysm Genetics: A south India perspective. *S. Sathyan[1], L. Koshy[1], H.V. Easwer[2], S. Premkumar[3], J. Alapatt[3], S. Nair[2], R.N. Bhattacharya[2], M. Banerjee[1].* 1) Human Molecular Genetics, Rajiv Gandhi centre for biotechnology, Thiruvananthapuram, KERALA, India; 2) Department of Neurosurgery, SCTIMST, Thiruvananthapuram; 3) Department of Neurosurgery, Calicut Medical College, Calicut, India.

Intracranial aneurysm (IA) is a fairly common condition that is often asymptomatic until the time of rupture. Rupture of cerebral aneurysm is the foremost cause for spontaneous Subarachnoid hemorrhage (SAH). In general population 2-3% of the individuals are likely to harbour intracranial aneurysm. Autopsy studies from India have shown the prevalence of IA ranges from 0.2% to 10.3% with a mean prevalence of 5.3 %. Aneurysmal SAH(aSAH) is associated with mortality rate as high as 40% to 50%. Genetic and environmental factors are reported to influence development of IA.The genes involved in vascular remodelling, endothelial dysfunction and immune response converge in addressing various hypothesis that has been put forward for pathogenesis of IA. Till date five genome wide association studies have been carried pointing out involvement of 4q31.23, 8q21.3, 9p21.3 and 13q.13.1.9p21 genomic loci for IA. The objective of this study is to screen these hypothesis driven candidate genes and validate the positional regions in the GWAS studies in the pathogenesis of intracranial aneurysm in south Indian population. The study populations consisted of 225 radiologically confirmed aneurismal cases and 235 ethnically and age matched controls from Dravidian Malayalam speaking population of Kerala. Genotyping was carried out based on allelic discrimination and Sequencing chemistry. In silico validation of functional effect was carried out. While evaluating the environmental factors our study show that history of hypertension (P-value<0.001,OR= 2.98) and cigarette smoking (P-value<0.001,OR=3.59) where associated with intracranial aneurysm while diabetes showed an protective effect on intracranial aneurysm(P-value=0.009,OR=0.34 (0.15-0.76). While evaluating the pathway focused genetic associations for possible endophenotyping we observe that *MMP2, COL1A2* and *VCAN* genes of extracellular matrix remodeling and proinflammatory cytokines of TNFalpha and IFN gamma of immune response are associated with IA in South Indian population. Their functional implication and interaction in developing a precise ECM/proinflammatory endophenotype would be interesting.Among positional GWAS hits located in 4q31.23, 8q21.3, 9p21.3 and 13q.13.1.9p21 harbouring *EDNRA, SOX17, CDKN2BAS* and *STARD13* respectively we could not replicate any of these regions with exception to 9p21.3 to be associated for IA in South Indian population. This is the first study on genetics of IA in any Indian population.

834T

Association between IRF6 polymorphisms and 8q24region in non-syndromic cleft lip with or without cleft palate in Brazilian population. *L.T. Souza[1,2], T.K. Kowalski[2], J. Ferrari[2], I.L. Monlleo[3], E.M. Ribeiro[3], J. Souza[3], G.F. Leal[3], A.C. Fett-Conte[3], V.G. Silva-Lopes[3], A.K.C. Ribeiro-dos-Santos[4], S.E.B. Santos[4], T.M. Félix[1,2,3,5].* 1) Programa de pós-graduação em Saúde da Criança e do Adolescente-Uinersidade federal do Rio Grande do Sul, UFRGS, Brazil; 2) Laboratório de Medicina Genômica, Hospital de Clínicas de Porto Alegre, HCPA, Brazil; 3) Projeto Crânio-Face Brazil, Brazil; 4) Laboratório de Genética Humana e Médica, Universidade Federal do Pará, UFP, Brazil; 5) Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, HCPA, Brazil.

Cleft lip and palate (CL/P) has a multifactorial inheritance. Identification of genetic risk factors has been the subject of intensive research. Three polymorphisms, rs2235371 and rs642961 in IRF6 and rs987525 in 8q24, have been associated with CL/P risk in several studies. Variants in IRF6 gene are responsible for 12% of clefts, the first variant associated was rs2235371 (V274I) and it is in linkage disequilibrium with rs642961. The rs642961 was suggested to cause disruption of the binding site of transcription factor AP-2α. The function of 8q24 is still unknown. A variant rs987525 (C>A) have involved in the pathogenesis of CL/P in Caucasians and admixed populations. The aim of this study was to evaluate the association between three polymorphisms and CL/P in the Brazilian admixture population. In this study were included subjects with non-syndromic CL/P and their parents that from three different geographical regions of Brazil. This research was approved by the Research Ethics Committee of Hospital de Clínicas de Porto Alegre and all the individuals signed an Informed Consent. For ancestry analysis, a panel of 48 INDELs was selected to measure the proportions of three different ancestries in the probands with CL/P. All selected markers were analyzed by multiplex. The SNPs were analyzed using TaqMan assay (Applied Biosystem). Statistical analysis was performed with FBAT and Haplin softwares. Acestry was analyzed in 228 probands. We select triads whose probands had European ancestry >=0.6 to further genetic analysis of the polymorphisms. We analyzed 151 nuclear families totalizing 390 individuals. MAF for rs2235371 A allele was 0.08, 0.19 for rs642961A allele and 0.41 for rs987525 A allele. TDT analysis showed an overtransmission of G allele rs2235371 (p=0.004) and of A allele for rs987525 (p=0. 048). Haplotype analysis in IRF6 showed association for the children haplotype (rs2235371G and rs642961A) in single and double dose and mother haplotype (rs2235371G and rs642961A) for single dose. This data confirms association of rs2235371 G allele in IRF6 and rs987525 A allele in 8q24 region with non-syndromic CL/P, similar to the previous studies in Caucasian and Mixed populations. The association between these polymorphisms and CL/P is due to European ancestry. We found association with the haplotype (rs2235371 G and rs642961 A) for non-syndromic cleft lip and/or palate.This is the first study that was able to show this association in Brazil.

835S

***MC4R, TMEM18, SH2B1, SEC16B* and *ADIPOQ* gene variants: associations with anthropometric and dietary variables in young children.** *M.R. ZANDONA[1], M.R. VITOLO[1, 2], S. ALMEIDA[1], V.S. MATTEVI[1].* 1) Federal University of Health Sciences of Porto Alegre, PORTO ALEGRE, RIO GRANDE DO SUL, Brazil; 2) Departament of Nutrition, Federal University of Health Sciences of Porto Alegre, Porto Alegre, RS, Brazil.

Background: The prevalence of childhood obesity is increasing worldwide and has become a major health problem. Identification of susceptibility genes in early life could provide the foundations for interventions in lifestyle to avoid obese children to become obese adults. Genome-wide association studies (GWAS) have led to the identification of several loci in the human genome containing genetic variants conferring an increased risk of developing overweight and obesity. Furthermore, association studies in children and adolescents remain useful in the investigation of candidate genes for obesity, in which rare variants might be identified and responsible for unexplained phenotypic variation. In this study we evaluated the influence of genetic variants related to obesity identified by GWAS (*MC4R, TMEM18, SH2B1* and *SEC16B*) and association studies (*ADIPOQ, PPARG* and *LEPR*) on anthropometric phenotypes and food intake in a cohort of 424 children followed-up since birth until 3 years old. Methods: The polymorphisms were genotyped using real-time polymerase chain reaction and the dependent variables were compared among genotypes at the ages of 1 and 3 years old by t-tests and analysis of variance. Multilevel mixed models were also used to analyze genotype effects on the patterns of individual change in anthropometric and dietary variables over time. Results: In this sample 52.6% were boys. At 3 years, overweight prevalence was 43.1%. *TMEM18* rs6548238 was associated with BMI Z-score at 3 years and over the years (P=0.029, P=0.011; respectively). We also found associations with intake of lipid dense foods at 1 year (P=0.045) and total energy intake over the years (P=0.017). *MC4R* rs17782313 was associated with intake of sugar dense foods at 1 year and intake of lipid dense foods over the years (P= 0.023, P=0.020; respectively). *SH2B1* rs7498665 was associated with intake of sugar dense foods at 3 years (P=0.029). *SEC16B* rs10913469 was associated with sum of skinfolds at 1 year in girls (p=0.028). Association between *ADIPOQ* rs17300539 and BMI-Z score was observed at 1 year (P=0.022) and *ADIPOQ* rs266729 was associated with sum of skinfolds over the years (P=0.027). Conclusions: This study provides indications that genetic variants in *TMEM18, MC4R, SH2B1, SEC16B* and *ADIPOQ* genes might be associated with anthropometric phenotypes and food intake in young children.

836M

The QT-interval prolonging variant p.D85N of *KCNE1* associates with reduced levels of insulin after an oral glucose load. *A. Jonsson[1], M.N. Harder[1], A. Stancakova[2], A.P. Gjesing[1], T.S. Ahluwalia[1,3,4], N. Grarup[1], S.S. Torekov[1,5], C. Graff[1,6], A. Linneberg[7], T. Jørgensen[7,8], T.I.A. Sørensen[1,9], J.J. Holst[1,5], M. Laakso[2], J.K. Kanters[5,10], O. Pedersen[1], T. Hansen[1,11].* 1) The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 2) Department of Medicine, University of Eastern Finland, Kuopio, Finland; 3) Copenhagen Prospective Studies on Asthma in Childhood, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 4) The Danish Pediatric Asthma Center, Gentofte Hospital, the Capital Region, Copenhagen, Denmark; 5) Department of Biomedical Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark; 6) Department of Health Science and Technology, Aalborg University, Aalborg, Denmark; 7) Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark; 8) Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 9) Institute of Preventive Medicine, Bispebjerg and Fredriksberg Hospitals, the Capital Region, Copenhagen, Denmark; 10) Gentofte, Aalborg and Herlev University Hospitals, Hellerup, Denmark; 11) Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark.

Aim/hypothesis: Mutations that cause functional changes in an ion channel may result in different diseases, such as cardiac arrhythmias and beta cell dysfunction, depending on which tissue the channel is expressed in. *KCNE1* is expressed in both cardiomyocytes and in pancreatic islets and a low frequent functional variant in this gene, *KCNE1* p.D85N (CEU MAF 0.02), has been shown to affect the QT-interval. We aimed to test for associations between the QT-interval-associated allele of *KCNE1* p.D85N and insulin release from pancreatic beta cells. **Methods:** Associations of the *KCNE1* p.D85N variant with QT-interval and insulin levels were studied in 5,738 Danish individuals from the Inter99 study without known diabetes. An association with 2-h insulin was taken forward for replication in a total of 1,662 Danish and Finnish non-diabetic individuals and the results were combined in a meta-analysis. The beta cell response to an intravenous load of glucose and tolbutamide was examined in 303 individuals from the Danish Family study. **Results:** The QT-interval-prolonging allele of *KCNE1* p.D85N associated with increased QT-interval (β=0.023, p=1.4×10^{-13}) and decreased 2-h insulin (β=-0.149, p=0.013) in Inter99. The association with decreased 2-h insulin stayed significant in the combined meta-analysis (β=-0.174, p= 7.6×10^{-4}). This variant was also associated with decreased intravenous tolbutamide-induced insulin secretion (incremental AUC (mean±SD) 1,574±1,744 vs. 2,398±1,625, p=0.036) in the Danish Family study. **Conclusion/interpretation:** We found that heterozygote carriers of the low frequent QT-interval-prolonging allele of *KCNE1* p.D85N have lower insulin levels after an oral glucose tolerance test and intravenous tolbutamide injection suggesting that these carriers have both impaired glucose and tolbutamide-stimulated insulin secretion.

837T

The role of selected ion channel genes in dental caries. *D. Lewis[1], J. Shaffer[1], E. Feingold[1,2], M. Cooper[3,4], M. Vanyukov[1,5,6], B. Maher[7], S. Reis[8,9], W. McNeil[10], R. Crout[11], R. Weyant[12], S. Levy[13,14], A. Vieira[3,4], M. Marazita[1,3,4,6,8].* 1) Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 3) Center for Craniofacial & Dental Genetics, Department of Oral Biology, School of Dental Medicine, Pittsburgh, PA; 4) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, PA; 5) Department of Pharmaceutical Sciences, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA; 6) Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA; 7) Department of Mental Health, John Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 8) Clinical and Translational Science Institute, School of Medicine, University of Pittsburgh, Pittsburgh PA; 9) Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA; 10) Dental Practice and Rural Health, West Virginia University, Morgantown, WV; 11) Department of Periodontics, School of Dentistry, West Virginia University, Morgantown, WV; 12) Department of Dental Public Health and Information Management, School of Dental Medicine, University of Pittsburgh, Pittsburgh PA; 13) Department of Preventive and Community Dentistry, University of Iowa College of Dentistry, Iowa City, IA; 14) Department of Epidemiology, University of Iowa College of Public Health, Iowa City, IA.

Ion channels play an important role in regulating and maintaining the calcium and pH homeostasis that is critical for tooth development. Mutations in genes encoding for ion channels can be related to a heterogeneous group of diseases called “channelopathies.” One common channelopathy is Timothy Syndrome, a Mendelian disease that affects many parts of the body including the heart, fingers and toes, the nervous system and causes small, misplaced teeth and frequent cavities in children. As part of a larger candidate gene study we investigated 480 single-nucleotide polymorphisms (SNPs) in several ion channel genes, including CACNA1C, CACNA2D1, CACNB2, CACNG2, KCNH1, KCNK5, and KCNK17, several of which are known to play a role in Timothy syndrome or other channelopathies. We tested association of these genes with dental caries in 13 race and age-stratified cohorts from six independent studies of Caucasians and African Americans. We performed analyses independently for each cohort and synthesized results by meta-analysis. After gene-wise adjustment for multiple testing, two SNPs for CACNA2D1 were significantly associated with dental cares via meta-analysis across the five childhood cohorts and in one individual childhood cohort (p-values<0.0007). In adults, genetic association was observed in three individual cohorts for potassium channel genes KCNH1 and KCNK5 (p-values<0.001), but no single SNP was significant via meta-analysis across all eight adult cohorts. These findings strengthen the hypothesis that ion channel genes, particularly those involved in channelopathies, may affect the risk of dental caries. Grants:R01-DE014899, U01-DE018903.

838S

Analysis of Haptoglobin Duplication with Type 2 Diabetes and Diabetic End Stage Kidney Disease. *JN. Adams[1,2,3], LM. Raffield[1,2,3], AJ. Cox[2,3,4], ET. Barton[4], CD. Langefeld[5], BI. Freedman[2,3,6], MCY. Ng[2,3], ND. Palmer[2,3,4], DW. Bowden[2,3,4].* 1) Molecular Genetics and Genomics, Wake Forest School of Medicine, Winston Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 5) Division of Public Health Sciences, Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 6) Department of Internal Medicine - Nephrology, Wake Forest School of Medicine, Winston-Salem, NC, 27157.

Haptoglobin (HP) is an acute phase protein that binds to freely circulating hemoglobin. HP assists in the removal of hemoglobin from the circulatory system, thus preventing its accumulation in the kidney. This accumulation could lead to kidney damage and oxidative damage caused by the heme iron. The HP protein exists as two distinct forms, HP1 and HP2. The longer HP2 form arose from a duplication of exons 3 and 4 and has been associated with cardiovascular disease (CVD) events and mortality in individuals with type 2 diabetes (T2D). HP has also been reported to be associated with T2D in multiple ethnicities, and with end stage kidney disease (ESRD) in individuals with type 1 diabetes. This novel duplication is not tagged or genotyped on conventional genotyping arrays and thus requires direct genotyping of the locus or Western blot. In this study we investigated the association of the *HP* duplication with T2D and diabetic end stage kidney disease (DMESKD) in African-Americans. *HP* genotyping was performed by PCR amplification using two sets of allele-specific primers, followed by resolving the products by agarose gel electrophoresis. Genotyping was performed in 4560 African American individuals: 2022 with T2D and DMESKD, 891 with T2D only, and 1647 controls with neither disease. The *HP* duplication was then analyzed for association with T2D and DMESKD. The *HP* duplication was not associated with T2D in this population (p=0.25). However, analyses revealed that the *HP*2 allele was associated with increased DMESKD risk with an odds ratio (OR) of 1.18 (95% confidence interval (CI) 1.06-1.28, p= 0.0012) when comparing individuals with DMESKD to control individuals. When comparing individuals with DMESKD to individuals with T2D and controls, the *HP*2 allele remained associated with DMESKD with an OR of 1.15 (95% CI 1.05-1.25, p=0.00183). This study suggests an association between the *HP* duplication and susceptibility to DMESKD in the African-American population. HP is a strong biological candidate for kidney disease. HP has been shown to decrease oxidative damage of the kidney. However, the HP2-2 protein has been shown to have decreased antioxidant properties compared to HP1-1. In addition the HP2-2-hemoglobin (Hb) complex is cleared much slower than the HP1-1-Hb complex. This leads to an accumulation of iron in the proximal tubule cells of the kidney. This accumulation leads to increased oxidative stress and damage to the tubule, and increased kidney disease.

839M

Assessment of Common and Rare Variants at Established Type 2 Diabetes and Glucose Homeostasis Loci for Type 2 Diabetes Risk in African Americans. *J.M. Keaton[1,2,3], P. Mudgal[3], B.I. Freedman[4], D.W. Bowden[2,3,5,6], M.C.Y. Ng[2,3].* 1) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Internal Medicine - Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC; 5) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 6) Internal Medicine-Endocrinology, Wake Forest School of Medicine, Winston-Salem, NC.

Type 2 diabetes (T2D) is a metabolic disease characterized by both insulin resistance and impaired insulin secretion. African Americans (AAs) have higher disease prevalence of T2D (12.6%) compared to European Americans (7.1%). Although GWAS of primarily European populations have identified ~70 loci associated with T2D, these loci are usually represented by a common index variant spanning a large region of linkage disequilibrium (LD). GWAS in a single population may not have sufficient resolution to identify causal SNPs, address the impact of rare variants, or assess genetic architecture underlying ethnic disparities in disease prevalence. With these considerations, we examined SNP- and locus-wide association of 44 T2D loci reported in Europeans ($P \leq 5\times10\text{-}8$, odds ratio [OR] 1.07-1.39) for association in an AA cohort consisting of 1617 T2D cases and 676 healthy controls. Considering the pathophysiology of T2D, we also examined 29 reported European glucose homeostasis (GH) loci ($P \leq 5\times10\text{-}8$). Individuals were genotyped using an Affymetrix BioBank array customized for fine-mapping of T2D and GH candidate genes/loci. Our results show that 16 T2D index SNPs, or their proxies ($r^2 \geq 0.5$ in CEU), at the *THADA*, *GRB14*, *IRS1*, *WFS1*, *ANKRD55*, *JAZF1*, *CDKN2A/B*, *TCF7L2*, *KCNQ1*, *MTNR1B*, *KLHDC5*, *HMGA2*, *TSPAN8/LGR5*, *FTO*, *MC4R*, and *CILP2* loci were significantly associated with T2D ($P<0.05$, OR 1.17-1.55) in AAs. Additionally, 6 GH index SNPs or proxies at the *DGKB/TMEM195*, *GLIS3*, *ADRA2A*, *MADD*, *FADS1*, and *VPS13C/C2CD4A/B* loci were significantly associated with T2D ($P<0.05$, OR 1.19-1.35). The strongest association was *TCF7L2* rs7903146 ($P=2.22\times10\text{-}5$, OR=1.35), consistent with current literature. Of the 22 associated loci, the most strongly associated SNP in AAs was identical to or in LD ($r^2 \geq 0.5$) with the European index SNP at 10 loci: *GRB14*, *IRS1*, *ANKRD55*, *JAZF1*, *KCNQ1*, *FTO*, *TCF7L2*, *MADD*, *FADS1*, and *VPS13C/C2CD4A/B*. Gene-based analysis including both common and rare coding variants in SKAT revealed significant association at the IRS1 ($P=0.025$) and TSPAN8 ($P=0.020$) loci, but these associations were lost when only SNPs with predicted functional consequences were included in the model. These findings suggest that loci discovered in European GWAS influence T2D risk in AAs with similar effect sizes, that the lesser degree of LD in AAs may facilitate identification of the causal variants, and that rare coding variants at established T2D loci do not significantly contribute to T2D risk in AAs.

840T

Detailed phenotypic analysis of lipid SNPs reveals divergent effects of *SORT1* on circulating LDL cholesterol, plasma glucose and their respective cardio-metabolic complications. *L.A. Lotta, R.A. Scott, N.J. Wareham.* MRC Epidemiology Unit, School of Clinical Medicine, University of Cambridge, Box 285, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, United Kingdom.

Epidemiologic and genetic evidence suggest that the genetic background of blood lipids may be largely shared with that of glycaemic and anthropometric traits. Recent GWAS have identified >100 SNPs associated with lipid levels, but knowledge of their relevance to other metabolic traits is incomplete. We sought to assess the association of lipid SNPs with metabolic traits in 4,202 participants of the Fenland cohort study. We genotyped or imputed a total of 137 SNPs identified in a recent large-scale GWAS. Using linear regression models, we studied the association of individual SNPs and genetic risk scores of lipid traits (i.e. triglycerides, total-, HDL- and LDL-cholesterol) with detailed glycaemic and anthropometric phenotypes. These included fasting and post-challenge plasma glucose, insulin, HbA1c as well as anthropometric measurements from ultrasound and dual-energy X-ray absorptiometry (DEXA). Quantile-quantile plots of the association of SNPs with metabolic traits were systematically inflated, highlighting considerable pleiotropy. A 63-SNP HDL-increasing genetic score was associated with multiple measures of reduced glycaemia, insulin resistance and adiposity. Individual SNPs with evidence of pleiotropic effects included those at known metabolic-trait or type 2 diabetes (T2D) loci (e.g. *FTO*, *MC4R*, *FADS1-2-3*, *FAM13A* and *GCKR*). The SNPs with the strongest effect across multiple phenotypes were rs1121980 at *FTO* and rs629301 at *SORT1* - a known LDL and coronary artery disease (CAD) gene encoding an insulin-sensitive intracellular sorting protein. The LDL-reducing allele of rs629301 was associated with increased fasting insulin (beta [SE] in SD per allele, 0.07 [0.03]), HOMA-insulin resistance index (0.07 [0.03]) and body fat mass (0.06 [0.03]). DEXA-scan analyses also revealed increased fat mass in each body compartment - in particular the trunk (0.07 [0.03]) - while bone or lean masses were unaffected. In the publicly-available CARDIoGRAM and DIAGRAMv3 datasets, the SNP was associated with CAD (OR [95% CI], 0.90 [0.87-0.93]) and T2D (1.04 [1.00-1.08]) in opposite directions, consistent with the association observed with the corresponding intermediate traits - LDL cholesterol (beta [SE], -0.13 [0.03]) and fasting glucose (0.05 [0.03]). Our study highlights opposite effects of *SORT1* on cholesterol levels and glycaemic traits. This finding is relevant to discussions on the development of lipid-lowering drugs targeting sortilin.

841S

Candidate Genes for Non-syndromic Orofacial Clefts Identified by GWAS Were Assessed in Two African Populations. *A. Butali[1], P.A Mossey[2], W.L Adeyemo[3], E.A Mekonen[4], L. Gaines[1], T. Busch[1], R.O Braimah[5], S.B Aregbesola[5], J. Rigdon[1], C. Emeka[3], O. James[3], O. Ogunlewe[3], A.L Ladeinde[3], F. Abate[4], T. Hailu[4], M. Ibrahim[4], P. Gravem[4], M. Deribew[4], M. Gesses[4], A. Adeyemo[6], M.L Marazita[7], J.C Murray[1].* 1) University of Iowa, Iowa City, IA. U.S.A; 2) University of Dundee, UK; 3) University of Lagos, Nigeria; 4) Addis Ababa University, Ethiopia; 5) Obafemi Awolowo University, Nigeria; 6) National Institutes for Health, Bethesda, MD, U.S.A; 7) University of Pittsburgh, Pittsburgh Pennsylvania, U.S.A.

Genome-wide association studies (GWAS) for non-syndromic cleft lip with or without cleft palate (NSCL/P) have identified 14 new loci. These loci have been replicated in several studies confirming the role of common variants in increasing risk to NSCL/P. Rare variants in these candidate genes have also been reported in resequencing studies. Here, we present analysis of common variants reported in the GWAS studies and rare variants identified following genotyping and Sanger sequencing in two sub-Saharan African populations from Nigeria and Ethiopia. Genotyping of 24 common variants in these 14 loci was done using Fluidigm technology which allows multiplexing of samples. Genotyping of 228 pedigrees was completed with a 95% genotyping call rate. Transmission disequilibrium tests and parent of origin analyses were conducted using PLINK. We sequenced the coding regions of the following genes: MAFB, PAX7, VAX1, ARHGAP29, and IRF6 in 220 probands from Africa [140 Nigerians and 80 Ethiopians] in order to identify rare functional variants. The TDT results for the 24 common variants were not significant ($p>0.05$). However, we observed significant paternal effects for SPRTY2 rs9574565 ($X2=4.5$; $p=0.03$). We also found new rare variants— p.His165Asn in the MAFB gene, p.Asp428Asn in the PAX7, a splice-site variant that creates a new donor splice-site in PAX7 following sequencing. All these variants segregate in parents. It is important to conduct an unbiased genome-wide association and sequencing studies using samples from this understudied African population in order to identify new candidate genes and loci for NSCL/P. Grants: R00-DE022378 (AB) R37-DE008559 (JCM), R01-DE016148 (MLM).

842M

Interaction of immune-related genetic polymorphisms and breastfeeding duration with *Helicobacter pylori* prevalence: the Pasitos Cohort Study. *M.L. Grove[1], K.A. Volcik[2], E.A. Brown[3].* 1) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 2) Department of Biochemistry and Molecular Biology, University of Texas Medical School at Houston, Houston, TX; 3) Center for Infectious Disease, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX.

Helicobacter pylori (*H. pylori*) is a bacteria that colonizes in the mucosal layers of the gastric and sometimes duoadenal or esophageal epithelium in up to 50% of the human population. In 3% of individuals, *H. pylori* contributes to the progression of gastric cancer which is the second leading cause of cancer mortality worldwide. Numerous risk factors have been associated with *H. pylori* infection risk, including genetic variation and environmental contributors such as breastfeeding. Therefore, we investigated the associations of immunological candidate gene loci with *H. pylori* prevalence within the context of breastfeeding (BF) status. We examined 87 SNPs in 27 immune-related genes in 239 Hispanic infants from the Pasitos Cohort Study which were ascertained from two populations residing on the United States (US) and Mexico border at El Paso, Texas and Juarez, respectively. *H. pylori* infection prevalence was ascertained using serum measures, breastfeeding status was categorized as <6 months or ≥6 months duration, and genetic variation was captured with the Sequenom MassARRAY® system using established genotyping methods. Additive logistic regression models were used to test each single nucleotide polymorphism (SNP) by country of origin which included gender, BF status, and a SNPxBF multiplicative interaction term. Models with statistically significant interactions were further stratified by BF categories and the genotypes assessed for additive effects including gender as a covariate. We observed significant SNPxBF interactions for three variants in the US population (*CCR6* *P*=0.003, *IL15* *P*=0.004, and *SFTPD* *P*=0.042), and *IL6R* (*P*=0.037) in the Mexico population. After stratification of these loci by BF status, variants in *CCR6* and *SFTPD* were shown to be associated with 25% and 26% lower prevalence in infants BF≥6 months (rs2022001, odds ratio [OR]=0.25, 95% confidence interval [CI]=0.10-0.67, *P*=0.006; and rs721917, OR=0.26, 95% CI=0.08-0.80, *P*=0.019), respectively. Homozygote infants BF<6 months with the risk-raising allele in *IL15* in the US population and *IL6R* in the Mexico population were 2.6 and 2.1 times more likely to be infected with *H. pylori* (rs10519613, OR=2.66, 95% CI=1.49-4.77, *P*=0.001; and rs6427641, OR=2.12, 95% CI=1.09-4.14, *P*=0.028), respectively. To our knowledge, we are the first to report variations in immune-related genes may interact with BF status to influence *H. pylori* prevalence in Hispanic infants.

843T

Why do some athletes with sickle cell trait suffer from heat illness? *A.C. Stone[1], R. Grieger[1], J. Lund[1], M. Ciambella[1], C. Flansburg[2], L. Madrigal[2].* 1) Human Evolution & Social Chg, Arizona State Univ, Tempe, AZ; 2) Department of Anthropology, University of South Florida, Tampa, FL.

The NCAA recently declared sickle cell trait (SCT) to be a risk factor for sudden illness and death among student athletes. Although its effect on SCT is not fully understood, fetal hemoglobin (HbF) concentration in adults is negatively correlated with disease severity in sickle cell anemia, and the concentration of HbF in adults is found to be highly variable across populations. In order to test whether lower HbF levels in some SCT athletes increases their risk for illness during exercise, we conducted preliminary analysis of five single nucleotide polymorphisms (SNPs) located in the human beta globin gene cluster (rs7482144 and rs10128556), the HBS1L-MYB intergenic interval (rs9402686), and the BCL11a gene (rs10189857 and rs4671393). These SNPs contribute to the heritable variation in HbF levels and are associated with increased HbF concentrations in adults. A sample (n= 22) of NCAA football student athletes was genotyped at these SNPs, and their allele frequencies were compared to those of other populations. For rs7482144 and rs10128556 in the beta globin gene cluster, the minor allele frequencies of both were 0.091 in the sample population. The minor allele frequencies for rs9402686, rs10189857, and rs4671393 were 0.07, 0.34, and 0.41, respectively. These results compared closely with other populations of recent African heritage. The results of this study will be combined with data from medical records to assess whether these polymorphisms can be used to predict susceptibility to exercise related illness in NCAA student athletes with SCT.

844S

Genetic determinants of benign prostatic hyperplasia: associations with prostate volume. *A. Giri[1], J.H. Fowke[1,2], S.S. Motley[1], S. Byerly[1], T.L. Edwards[1,3].* 1) Division of Epidemiology, Institute for Medicine and Public Health, Vanderbilt University, Nashville, TN; 2) Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN; 3) Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

A highly prevalent condition among men over the age of 50, benign prostatic hyperplasia (BPH) is the non-malignant proliferation of the epithelial and stromal cells in the prostate gland. It is usually diagnosed in the presence of enlarged prostate. Components of metabolic disorders such as obesity, elevated lipid profiles, and fasting glucose/insulin levels have been positively associated with BPH. However, whether there is a genetic basis to these associations has not been investigated in detail. Using the Metabochip platform we evaluated over 100,000 common single nucleotide polymorphisms (SNPs) from 257 loci implicated with metabolic disorders throughout the genome in relation to prostate volume (PV). Participants for this study were derived from the Nashville Men's Health Study, a rapid-recruitment protocol targeting men seeking a diagnostic prostate biopsy. Consenting participants provided information regarding individual characteristics, measurement of anthropomorphic traits and blood samples prior to prostate biopsy. PV (ml) was measured using ultrasound. Data from 876 prostate cancer free men and 442 men with low-grade prostate cancer (Gleason Score < 7) served as discovery and validation samples, respectively. We used linear regression in PLINK to evaluate the associations between SNPs (additive coding) and natural-log transformed (ln) PV while adjusting for age, body mass index, height, and 10 genetic ancestry principal components derived using Eigensoft. Inverse variance weighted fixed-effect meta-analysis of beta-coefficients from the discovery and validation sample suggested SNP rs11736129 (near *LOC100131429*) to be significantly associated with ln-PV (beta: 0.16, p-value 1.16×10^{-8}). Other SNPs that were nominally associated ($p < 1\times10^{-4}$) with ln-PV included rs9583484 (intronic SNP in *COL4A2*), rs10146527 (intronic SNP in *NRXN3*), rs9909466 (SNP near *RPL32P31*), and rs2241606 (synonymous SNP in the *SLC12A7* exon). *LOC100131429* is an unprocessed pseudogene that bears sequence resemblance with the *ARMC1* gene, which has been reported to be nominally associated with childhood obesity in Hispanics. Similarly, the *NRXN3* locus has been previously associated with waist-to-hip ratio in the GIANT consortium. *COL4A2* is a component of type IV collagens which provide major structural support to basement membranes of several organs including the prostate. In conclusion, we found several metabolic loci that may be important to the etiology of BPH.

845M

IDENTIFICATION OF SIGNIFICANT ASSOCIATION AND GENE-GENE INTERACTIONS OF POLYMORPHISMS IN THREE INFLAMMATORY GENES CRP, TNF-α, AND LTA FOR LOWER EXTREMITY PERFORMANCE IN COMMUNITY-DWELLING ELDERS IN TAIWAN-TAICHUNG COMMUNITY HEALTH STUDY FOR ELDERS (TCHS-E). *T.C. Li[1,2], C.C. Lin[3,4], C.I. Li[4,5], N.H. Meng[4,6], W.Y. Lin[3,4], C.S. Liu[3,4], C.H. Lin[3,4,7], C.W. Yang[5,7], C.K. Chang[4,6], F.Y. Wu[8], L.N. Liao[8].* 1) Graduate Institute of Biostatistics, China Medical University, Taichung, Taiwan; 2) Department of Healthcare Administration, College of Health Science, Asia University, Taichung, Taiwan; 3) Department of Family Medicine, China Medical University Hospital, Taichung, Taiwan; 4) School of Medicine, College of Medicine, China Medical University, Taichung, Taiwan; 5) Department of Medical Research, China Medical University Hospital, Taichung, Taiwan; 6) Department of Physical Medicine and Rehabilitation, China Medical University Hospital, Taichung, Taiwan; 7) Ph.D. Program for Aging, College of Medicine, China Medical University, Taichung, Taiwan; 8) Department of Public Health, College of Public Health, China Medical University, Taichung, Taiwan.

Purpose: We conducted a cross-sectional study to assess the main effects and possible gene-gene interaction of three inflammatory cytokine gene polymorphisms C-reactive protein (CRP), tumor necrosis factor α (TNF-α), and lymphotoxin α (LTA) on lower extremity performance by investigating eleven single-nucleotide polymorphisms (SNPs) of these three genes in community-dwelling elders in Taiwan. Methods: Five SNPs (rs2794520, rs1205, rs1130864, rs1800947, rs3093059) of CRP gene, three SNPs (rs2239704, rs909253, and rs1041981) of LTA gene, and three SNPs (rs1799964, rs1800629, and rs3093662) of TNF-α gene were genotyped for 472 unrelated elderly subjects (221 women and 251 men; mean age, 73.8±6.1 years). Lower extremity performance measurements were administered to all participants including timed up and go test (TUG, sec), walking speed (m/sec), weight-adjusted leg press (waLP, %), and time to rise from a chair three times (Chair3, sec). Results: Significant genotypic association was found for SNPs rs2794520 and rs1205 (in CRP; P <0.05) on walking speed and SNP rs1800947 (in CRP; P <0.05) on waLP in women; and for SNP rs2239704 (in LTA; P <0.05) on waLP in both men and women. We also identified significant two-locus gene-gene effects involving rs2794520 in CRP gene and rs1799964 in TNF-α gene, and rs1205 in CRP gene and rs1799964 in TNF-α gene for walking speed and Chair3; rs2239704 in LTA gene and rs1799964 in TNF-α gene for TUG; rs1130864 in CRP gene and rs2239704 in LTA gene for walking speed; rs909253 in LTA gene and rs1799964 in TNF-α gene, and rs1041981 in LTA gene and rs1799964 in TNF-α gene for waLP; and rs3093059 in CRP gene and rs1799964 in TNF-α gene for Chair3 in women; rs1800947 in CRP gene and rs1800629 in TNF-α gene for TUG in men. Conclusions: These results support the hypothesis that inflammatory genes are involved in lower extremity performance, most likely via complex gene-gene interactions. In addition, there exists gender difference in the association between inflammatory genes and lower extremity performance.

846T

IDENTIFICATION OF SIGNIFICANT ASSOCIATION AND GENE-GENE INTERACTIONS OF POLYMORPHISMS IN THREE INFLAMMATORY GENES TNF-α, LTA AND IL-6 FOR MARKERS OF APPENDICULAR SKELETAL MUSCLE MASS IN COMMUNITY-DWELLING ELDERS IN TAIWAN-TAICHUNG COMMUNITY HEALTH STUDY FOR ELDERS (TCHS-E). *L.N. Liao[1], C.C. Lin[2,3], C.H. Lin[2,3,4], C.W. Yang[4,5], C.S. Liu[2,3], C.I. Li[3,5], N.H. Meng[3,6], C.K. Chang[3,6], W.Y. Lin[2,3], F.Y. Wu[1], T.C. Li[7,8].* 1) Department of Public Health, China Medical University, Taichung, Taiwan; 2) Department of Family Medicine, China Medical University Hospital, Taichung, Taiwan; 3) School of Medicine, College of Medicine, China Medical University, Taichung, Taiwan; 4) Ph.D. Program for Aging, College of Medicine, China Medical University, Taichung, Taiwan; 5) Department of Medical Research, China Medical University Hospital, Taichung, Taiwan; 6) Department of Physical Medicine and Rehabilitation, China Medical University Hospital, Taichung, Taiwan; 7) Graduate Institute of Biostatistics, China Medical University, Taichung, Taiwan; 8) Department of Healthcare Administration, College of Health Science, Asia University, Taichung, Taiwan.

Purpose: We undertook a cross-sectional study to assess the main effects and possible gene-gene interaction of three inflammatory cytokine gene polymorphisms tumor necrosis factor α (TNF-α), lymphotoxin α (LTA) and interleukin-6 (IL-6) on markers of appendicular skeletal muscle mass by investigating nine single-nucleotide polymorphisms (SNPs) of these three genes in community-dwelling elders in Taiwan. **Methods:** Arm and leg lean muscle mass (LMS, kg) and appendicular skeletal muscle mass (ASM; kg), were measured by Dual-energy X-ray absorptiometry. Height-adjusted skeletal muscle index (hSMI; kg/m^2) was defined as ASM divided by height squared. Weight-adjusted skeletal muscle index (wSMI; %) was tallied as ASM divided by weight. Three SNPs (rs2239704, rs909253, and rs1041981) of LTA gene, three SNPs (rs1799964, rs1800629, and rs3093662) of TNF-α gene, and three SNPs (rs1880243, rs1800796, and rs1554606) of IL-6 gene were utilized to genotype 472 unrelated elderly subjects (221 women and 251 men; mean age, 73.8±6.1 years). **Results:** Significant genotypic association was found for SNP rs2239704 (in LTA; P <0.05) on leg LMS, ASM, hSMI, and wSMI and SNP rs1799964 (in TNF-α; P <0.05) on arm LMS and ASM in women; and for SNP rs1880243 (in IL-6; P <0.05) on leg LMS and ASM in men. We also identified significant two-locus gene-gene effects involving rs909253 in LTA gene and rs1799964 in TNF-α gene for leg LSM and ASM in both men and women and for hSMI and wSMI in women; and rs1041981 in LTA gene and rs1799964 in TNF-α gene for leg LSM, ASM, hSMI and wSMI in women; rs909253 in LTA gene and rs1800796 in IL-6 gene for leg LSM, ASM, hSMI, and wSMI in men; and rs1041981 in LTA gene and rs1800796 in IL-6 gene for arm and leg LMS, ASM, hSMI and wSMI in men. **Conclusions:** These results support the hypothesis that inflammatory genes are involved in lean muscle mass markers, most likely via complex gene-gene interactions. In addition, there exists gender difference in the association between inflammatory genes and lean muscle mass markers.

847S

Genetic risk of rheumatoid arthritis conferred by *HLA-DRB1* in African Americans stratified by local ancestry. *R.J. Reynolds[1], N.M. Pajewski[2], S. Raychaudhuri[3,4], R.M. Plenge[5], S.L. Bridges[1].* 1) University of Alabama at Birmingham, Birmingham, AL; 2) Wake Forest University, Winston Salem, NC; 3) Brigham and Women's Hospital, Boston, MA; 4) Broad Institute, Cambridge, MA; 5) Merck & Co.

HLA-DRB1 Val11 is the residue with highest risk for rheumatoid arthritis in Caucasians. We sought to validate the rheumatoid arthritis (RA) risk of *HLA-DRB1* exon 2 amino acid positions and residue substitutions in African Americans. We inferred local ancestry (copies of ancestral alleles) flanking *HLA-DRB1* to test the hypothesis that RA risk from residues at amino acid position 11 is independent of European ancestry at this locus. 579 seropositive RA patients and 893 healthy controls were four digit *HLA-DRB1* genotyped and all samples were genotyped on the immunochip. Ancestral population reference panels of Africans (Zambia) and Europeans (1200 haplotypes each) were genotyped on the immunochip, and for all samples, the extended major histocompatibility complex of chromosome six was phased using BEAGLE. Local ancestry was estimated with HAPMIX. Association analysis was performed using PLINK. Binomial regression models were fit in R to test the hypothesis that the Val11 and Asp11 residues had equivalent risk with local ancestry. Local European ancestry (P-value = 9.02E-04), but not global European admixture (P-value = 0.07) was associated with increased RA risk. Amino acid 11 was the most strongly associated position (Permutation P-value <1E-05), but positions 71 and 74 were not significantly associated. Within position 11 the valine, OR(CI) = 3.0 (2.3, 4.0), and aspartic acid, OR(CI) = 1.9 (1.3, 2.7), residues conferred RA risk. The aspartic acid residue is specific for *HLA-DRB1* *09:01. There was no joint effect of local European ancestry and Val11 on RA risk (P-value = 0.69). In contrast, there was an interaction effect of local ancestry and Asp11 on RA risk (P-value = 8.59E-03). Stratified by two copies of African ancestry, the observed Asp11 risk increased from OR 1.9 to 2.6 (1.7, 4.1). Asp11 was not a significant risk residue for individuals with either 1 or 2 copies of local European ancestry. These results demonstrate that the major source of RA risk comes from *HLA-DRB1* amino acid position 11 and the valine residue, which does not vary by African or European ancestry. However we also found evidence of heterogeneity in RA risk between Caucasians and African Americans for both *HLA-DRB1* amino acid positions 71 and 74 and the Asp11 residue. Importantly, we document that local ancestry of class II HLA genes may mediate the individual genetic risk for immune system related diseases.

848M

THE MAIN EFFECTS AND GENE-GENE INTERACTIONS AMONG CRP, TNF-α AND LTA IS ASSOCIATED WITH HANDGRIP STRENGTH IN COMMUNITY-DWELLING ELDERS IN TAIWAN-TAICHUNG COMMUNITY HEALTH STUDY FOR ELDERS (TCHS-E). *F.Y. Wu[1], C.S. Liu[2,3], C.C. Lin[2,3], C.H. Lin[2,3,4], W.Y. Lin[2,3], C.I. Li[3,5], C.W. Yang[4,5], N.H. Meng[3,6], C.K. Chang[3,6], T.C. Li[7,8], L.N. Liao[1].* 1) Department of Public Health, China Medical University, Taichung, Taiwan; 2) Department of Family Medicine, China Medical University Hospital, Taichung, Taiwan; 3) School of Medicine, College of Medicine, China Medical University, Taichung, Taiwan; 4) Ph.D. Program for Aging, College of Medicine, China Medical University, Taichung, Taiwan; 5) Department of Medical Research, China Medical University Hospital, Taichung, Taiwan; 6) Department of Physical Medicine and Rehabilitation, China Medical University Hospital, Taichung, Taiwan; 7) Graduate Institute of Biostatistics, China Medical University, Taichung, Taiwan; 8) Department of Healthcare Administration, College of Health Science, Asia University, Taichung, Taiwan.

Purpose: We undertook a cross-sectional study to assess the main effects and possible gene-gene interactions of inflammatory gene polymorphisms C-reactive protein (CRP), tumor necrosis factor α (TNF-α) and lymphotoxin α (LTA) on handgrip strength, by investigating eleven single-nucleotide polymorphisms (SNPs) of three genes in community-dwelling elders in Taiwan. Methods: Eleven SNPs (rs2794520, rs1205, rs1130864, rs1800947, and rs3093059 in CRP gene; rs1799964, rs1800629, and rs3093662 in TNF-α gene; rs2239704, rs909253, and rs1041981 in LTA gene) were utilized to genotype 472 unrelated elderly subjects (221 women and 251 men; mean age, 73.8±6.1 years). Handgrip strength was measured by handgrip dynamometer (TTM-110D, TTM Co. Japan). We systematically explained the main effects and possible gene-gene interactions using linear regression to adjust for potential confounders. Results: All SNPs were consistent with Hardy-Weinberg equilibrium (P >0.05). For the comparison of handgrip strength among the genotype, dominant and recessive models, the linear regression models were adjusted for age, BMI, smoking and physical activity. In the genotype and dominant models, there were significant differences in handgrip strength in the G allele of CRP rs3093059 in woman (P <0.05). In women, CRP rs2794520 and LTA rs2239704 were significant markers for handgrip strength in the genotype and recessive models (P <0.05). We also found that men carrying the AA or AG genotypes of TNF-α rs1800629 had greater handgrip strength; women carrying the GG genotype of CRP rs1205 had lower handgrip strength. We identified significant gene-gene interaction of LTA (rs2239704, rs909253, and rs1041981) and TNF-α (rs1799964) polymorphisms on handgrip strength in women; CRP (rs3093059) and TNF-α (rs1799964) polymorphisms in men. Furthermore, CRP rs2794520 & TNF-α rs1800629, CRP rs1205 & TNF-α rs1800629, and CRP rs1800947& LTA rs2239704 in women, as well as CRP rs3093059 & LTA rs2239704 in men, were significant interactions on handgrip strength. Conclusion: Our study show that an observed main effect and interaction between CRP, TNF-α and LTA polymorphisms on handgrip strength in elders.

849T

A gene-gene interaction in a shared Alzheimer disease/age-related macular degeneration pathway. *M.W. Logue[1,2,3], M. Schu[1], J. Farrell[1], K.L. Lunetta[2], G. Jun[1,2,4], C.T. Baldwin[1], M.M. DeAngelis[5], L.A. Farrer[1,2,4,6,7].* 1) Biomedical Genetics, Boston University School of Medicine, Boston, MA; 2) Biostatistics, Boston University School of Public Health, Boston, MA; 3) Research Service, VA Boston Healthcare System, Boston, MA; 4) Ophthalmology, Boston University School of Medicine, Boston, MA; 5) John A. Moran Eye Center, University of Utah, Salt Lake City, UT; 6) Neurology, Boston University School of Medicine, Boston, MA; 7) Epidemiology, Boston University School of Public Health, Boston, MA.

There is a growing body of biochemical and epidemiological evidence that Alzheimer disease (AD) and age-related macular degeneration (AMD) share pathogenic mechanisms. Recently, we found that risk genes for both disorders are enriched in clathrin-mediated endocytosis (CME), LXR/RXR activation, and atherosclerosis signaling pathways (Logue et al, 2014). Examining these pathways in summary GWAS data, we identified several novel AMD loci including *HGS* and *TNF*. Here, we explored the possibility that AD risk variants modify the effect of AMD risk variants within these three pathways by testing gene-gene interaction models in HapMap 2 imputed GWAS data for 1,336 AMD cases and 1,121 controls of European ancestry from the MMAP Study cohort (Chen, et al., 2010,Fritsche, et al., 2013). We analyzed the peak AD variants (or LD proxies) in *PICALM*, *CD2AP*, *APOE*, and *CLU* for interaction with the peak AMD risk loci in *APOM*, *CSNK2B*, *HGS*, *TNF*, *CETP*, *C3*, *C4A*, and *COL10A1*, restricting AD and AMD risk variant interactions to genes within one of the 3 pathways (22 independent tests; significance determined using robust SE estimates). Using a logistic-GEE model of AMD risk which included an AMD risk variant and AD risk variant, a GxG interaction, and age as a covariate, we found significant interaction (p= 0.00019; exceeding a Bonferroni-corrected threshold) within the CME pathway between the known AMD risk variant rs805262 in *CSNK2B* and the known AD risk variant rs10948363 in *CD2AP*. The rs10948363 G allele (MAF=27%), which is an AD risk allele, mitigates (heterozygous state) and even reverses (rare homozygous state) the risk of AMD conferred by the *CSNK2B* risk variant. In the MMAP dataset, the odds ratio (OR) associated with each minor allele T of rs805262 (MAF=46% in 1000 Genomes EUR) is 1.36 for those with the rs10948363 AA genotype (53% of EUR) and 1.05 among rs10948363 heterozygotes (41% of EUR), whereas rs805262 has an apparent protective effect (OR=0.65) for those who have the uncommon GG rs10948363 genotype (6.1% of EUR). Three other interactions, all of which involved genes in the CME pathway (*APOM* x *CD2AP*, *CSNK2B* X *CLU*, and *APOM* x *CLU*) were nominally significant (p<0.01). These findings reinforce the idea that AD and AMD share genetic mechanisms and that the clathrin-mediated endocytosis pathway may also play an important role in AMD pathogenesis.

850S

Targeted regulome sequencing reveals common, rare, and private regulatory variants associated with fetal adiposity. *C. Guo[1], I. McDowell[1], AA. Pai[2], DM. Scholtens[3], GE. Crawford[1,4], BE. Engelhardt[1,5], MG. Hayes[6], WL. Lowe[6], TE. Reddy[1,4].* 1) Institute for Genome Science and Policy, Duke University, Durham, NC; 2) Department of Biology, Massachusetts Institute of Technology, Cambridge, MA; 3) Department of Preventive Medicine, Division of Biostatistics, Northwestern University Feinberg School of Medicine, Chicago, IL; 4) Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC; 5) Department of Biostatistics & Bioinformatics, Duke University Medical School, Durham, NC; 6) Division of Endocrinology, Metabolism and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL.

Fetal growth has immediate and long-term effects for the health of the newborn. In particular, small and large for gestational age babies are at risk for later life obesity and type-2 diabetes. In our previous genome-wide association study of 4281 newborns, we demonstrated association of variation between CCNL1 and LEKR1 in 3q25.31 with fetal adiposity. As the associated variants were not in protein coding exons, we hypothesized that they alter the expression of nearby genes. To test this hypothesis, we sequenced regulatory elements within 250kb of the lead GWAS SNP, encompassing all variants in strong LD (r2 >0.8) with rs900400, across 760 babies in the upper and lower deciles of fetal adiposity in the Hyperglycemia and Adverse Pregnancy Outcome study. The babies represented four ancestry groups: European, Afro Caribbean, Mexican-American, and Thai. To define regulatory elements for sequencing, we used genome-wide measurements of open chromatin across ~50 different cell types including pre-adipocytes, liver, and islet β cells. Open chromatin regions are a general indicator of transcription factor binding and gene regulation, and several studies now demonstrate that a substantial fraction of complex phenotypes involve genetic variation in those regions. In total, we sequenced ~100 regulatory elements at a median coverage of 1100x for each sample. We found 995 variants (921 SNPs and 74 indels), 560 of which are not found in dbSNP or the 1,000 Genomes Project. To assess the regulatory function of the identified variants, we mapped eQTLs for protein coding genes and long-noncoding RNAs in four different cell types using data from the GTEx project, and in lymphoblastoid cell lines from the gEUVADIS dataset. Several variants in our sequencing are eQTLs for nearby genes, including two long-noncoding RNAs. Together, by combining targeted regulome sequencing and genome-wide eQTL analyses, these results suggest a potential role for novel genes including long-noncoding RNAs in fetal adiposity.

851M

Genome-wide Sequencing to Identify Novel Variants for Obesity in Pima Indians. *K. Huang, P. Piaggi, S. Kobes, R. Hanson, C. Bogardus, L. Baier.* Phoenix Epidemiology and Clinical Research Branch, NIDDK, NIH, Phoenix, AZ.

To identify genetic variation that increases risk for type 2 diabetes (T2D) or obesity, we obtained whole genome sequence data on 335 Pima Indians (51%; male, age: 25.1±5.8 years, BMI: 33.5±7.3 kg/m2). Sequencing was performed by Illumina (N=301) and Complete Genomics, Inc (N=34). ~13 million variants were found, including ~11 million SNPs, ~1.6 million Indels and 255,802 substitutions. Among all SNPs, 2.7 million were novel. Individuals who were sequenced were part of a study of health among Pima Indians living in the Gila River Indian Community, and were informative for longitudinal measures of BMI and type 2 diabetes status. As a preliminary screen for selecting variants for follow-up genotyping, we performed association analyses between all novel common SNPs (N=71,021 with a MAF≥0.05 and a genotype calling rate≥0.85) and maximum BMI childhood z-score (age 5-20 years; N=287) and maximum BMI in adulthood (after age 15 years from an exam where the individual was non-diabetic, N=289). The most significant preliminary association with childhood z-score was a novel exonic SNP in ZNF595 (chr4:86248, A/C, MAF=0.15; $p=3.2\times10^{-12}$ after adjusted for birth year, β=1.2 units per copy of the risk allele). In addition, variants in CHRNA4 and OR4A15/OR4C15 showed genome-wide significant association with childhood z-score (smallest $p=1.7\times10^{-10}$ after adjusted for birth year, β=1.1 units per copy of the risk allele) and adulthood BMI (chr11:55226415, G/A, MAF=0.15; $p=2.0\times10^{-9}$ after adjusted for age, sex and birth year, β=1.0kg/m^2 per copy of the risk allele), respectively. These novel variants and others selected for having the lowest p values or being positioned within excellent biologic candidates will be genotyped in 5,880 American Indians for validation of the preliminary associations with z-score during childhood and BMI during adulthood.

852T

Fine-mapping eGFR susceptibility loci through trans-ethnic meta-analysis. *A. Mahajan[1], J. Haessler[2], N. Franceschini[3], A. Morris[1,4,5].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 2) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; 3) University of North Carolina, Chapel Hill, NC, USA; 4) Department of Biostatistics, University of Liverpool, Liverpool, UK; 5) Estonian Genome Center, University of Tartu, Tartu, Estonia.

Chronic kidney disease (CKD) is a major public health problem and affects nearly 10% of the global population. Reduced estimated glomerular filtration rate (eGFR), a measure of kidney function used to define CKD, is associated with cardiovascular disease morbidity and mortality, acute kidney injury, and progression to end stage renal disease. Genome-wide association studies (GWAS) have been successful in identifying loci for eGFR. However, these loci are typically characterised by common lead SNPs with association signals extending over large genomic intervals containing multiple transcripts. As a result, limited progress has been made in identifying causal variants for eGFR and understanding the downstream pathogenesis of CKD. To address these drawbacks, we performed trans-ethnic meta-analysis to fine-map known eGFR loci by leveraging differences in distribution of linkage disequilibrium between diverse populations. We considered six GWAS comprising of 23,568 individuals of European, African American, and Hispanic ancestry, each supplemented by imputation up to the 1000 Genomes Project reference panel (March 2012 release). Within each study, association with eGFR (MDRD equation) was tested under an additive model. We then combined association summary statistics across studies with MANTRA, 500kb up and down of the lead SNP at known eGFR loci, and constructed "credible sets" of SNPs that encompass 99% of the posterior probability of being causal. We resolved fine-mapping of potential causal variants to less than 20 variants at three loci: *GCKR* (3 SNPs, 144.5kb), *UMOD/PDILT* (4 SNPs, 39.3kb), and *SHROOM3* (19 SNPs, 74kb). At *GCKR*, the credible set covers three SNPs including *GCKR* P446L, which is predicted to be the functional variant at this locus. Variants in the 99% credible set for *SHROOM3*, include intronic variants in the gene and overlap regulatory elements from ENCODE, thereby highlighting a potential mechanism for the action of this locus on eGFR. These findings provide evidence that trans-ethnic GWAS can be used to fine-map potentially causal variants at complex traits loci that can be taken forward for experimental validation and could help to further our understanding of the biological mechanisms underlying disease.

853S

Fine-mapping major histocompatibility complex associations in psoriasis and its clinical subtypes by HLA/MICA variant imputation. *Y. Okada[1-5], B. Han[3-5], L.C. Tsoi[6], P.E. Stuart[7], E. Ellinghaus[8], T. Tejasvi[7], V. Chandran[9,10], F. Pellett[10], R. Pollock[10], A.M. Bowcock[11], G.G. Krueger[12], M. Weichenthal[13], J. Voorhees[7], P. Rahman[14], P.K. Gregersen[15], A. Franke[16], R.P. Nair[7], G.R. Abecasis[6], D.D. Gladman[9,10,17], J.T. Elder[7,18], P.IW. de Bakker[4,5,19,20], S. Raychaudhuri[3-5,21].* 1) 1.Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan; 2) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 3) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 4) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 5) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 6) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 7) Department of Dermatology, University of Michigan Medical School, Ann Arbor, MI, USA; 8) Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany; 9) Department of Medicine, Division of Rheumatology, University of Toronto, Toronto, Ontario, Canada; 10) Centre for Prognosis Studies in the Rheumatic Diseases, Toronto Western Research Institute, University of Toronto, Toronto, Ontario, Canada; 11) National Heart and Lung Institute, Imperial College, London, UK; 12) Department of Dermatology, University of Utah, Salt Lake City, UT, USA; 13) University Department of Dermatology, Christian-Albrechts-University of Kiel, Kiel, Germany; 14) Memorial University, St. John's, Newfoundland, Canada; 15) The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, NY, USA; 16) Christian-Albrechts-University of Kiel, Kiel, Germany; 17) Toronto Western Research Institute, Toronto, Ontario, Canada; 18) Ann Arbor Veterans Affairs Hospital, Ann Arbor, MI, USA; 19) Department of Medical Genetics, Division of Biomedical Genetics, University Medical Center, Utrecht, The Netherlands; 20) Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands; 21) NIHR Manchester Musculoskeletal Biomedical, Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK.

Backgrounds: Psoriasis vulgaris (PsV) risk is strongly associated with genetic variation within the major histocompatibility complex (MHC) region, although its fine genetic architecture has not been elucidated. **Methods:** To fully characterize and fine-map the MHC associations of PsV, we conducted a large-scale fine-mapping study of PsV risk in the MHC region in 9,247 PsV cases and 13,589 controls of European descent. We also evaluated risk of two major clinical subtypes of PsV, psoriatic arthritis (PsA; n = 3,038) and purely cutaneous psoriasis (PsC, defined as 10 years of psoriasis without developing PsA; n = 3,098). We imputed class I and II HLA genes variants by applying SNP2HLA software to the SNP genotype data. In addition, we newly constructed an imputation reference panel of sequence variants for *MICA*, an HLA-like gene within the MHC region that has been implicated for PsA risk. We applied *MICA* variant imputation to the SNP genotype data and evaluated their risk as well. **Results:** As previously described, we observed that *HLA-C*06:02* demonstrated the most significant impact on overall PsV risk (odds ratio [OR] = 3.38, 95% confidence interval [95%CI]: 3.18-3.60, $P = 1.7\times10^{-364}$). Stepwise conditional analysis revealed multiple independent risk variants of both class I and class II HLA genes for PsV susceptibility independent of *HLA-C*06:02* (*HLA-C*12:03*, HLA-B amino acid positions 67 and 9, HLA-A amino acid position 95, and HLA-DQα1 amino acid position 53; $P < 5.0\times10^{-8}$), but no apparent independent risk conferred by *MICA*. Strikingly, we found that risk heterogeneity between PsA and PsC may be driven by one amino acid position at HLA-B (Glu at HLA-B amino acid position 45; OR = 1.46, 95%CI: 1.31-1.62, $P = 2.9\times10^{-12}$), which demonstrated much more significant association signals compared to classical *HLA-B* alleles including *HLA-B*27* and *HLA-B*39:01* ($P > 1.0\times10^{-4}$). **Conclusion:** These results indicate that multiple class I and II HLA genes (*HLA-C*, *HLA-B*, *HLA-A*, and *HLA-DQA1*) contribute to development of PsV, and suggest that different genetic factors, most evident at *HLA-B*, underlie for the differential risk for of specific PsV sub-phenotypes. Our study illustrates the value of high-resolution HLA and *MICA* imputation for fine-mapping causal variants in the MHC.

854M

Meta-analysis on the 22q11.21 region identifies an autoimmune disease risk allele as associated with systemic lupus erythematosus. *Y. Zhang[1], Y. Wang[1], J. Yang[1], N. Hirankarn[2], XJ. Zhang[3], YL. Lau[1], W. Yang[1,4].* 1) Paediatrics, The University of Hong Kong, Hong Kong, Hong Kong; 2) 2.Lupus Research Unit, Department of Microbiology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand; 3) State Key Laboratory Incubation Base of Dermatology, Key Laboratory of Dermatology, Anhui Medical University, Ministry of Education, China, Hefei, Anhui, 230032, China; 4) Centre for Genomic Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong.

Systemic lupus erythematosus (SLE) is an autoimmune disease with unclear etiology. It usually presents with a diverse spectrum of clinical manifestations. Genetic predisposition is responsible for 50 to 60 percent of the likelihood of developing the disease. Genome-wide association studies (GWASs) on SLE have identified more than 40 loci with genome-wide significance. The 22q11.21 genomic region is a shared locus for several autoimmune diseases including SLE, systemic sclerosis (SSc), Crohn's disease (CD), celiac Disease (CeD), Rheumatoid Arthritis (RA), Psoriasis(PS) and Inflammatory bowel disease(IBD). Here we examined the association for SLE in this region, and emphasized on the relationship of all the reported variants with various autoimmune diseases. Through meta-analysis of two existing GWASs on Chinese Han populations with a total of 1,659 cases and 3,398 controls matched geographically, we have identified SNP rs2298428 as the most significant SNP for SLE in the 22q11.21 region (P= 2.17E-09). The association of rs2298428, a missense mutation in YDJC gene, was further replicated in three cohorts from Hong Kong, Anhui and Thailand, independent of the GWAS samples used (P=1.31E-11, OR=1.23). This SNP also showed consistent correlation with the expression of UBE2L3 in multiple cell types, highlighting the functional mechanism of the abnormal expression of UBE2L3 to SLE.

855T

Involvement of GTF2IRD1 in the complex hearing phenotype of Williams-Beuren Syndrome. *C.P. Canales[1], A.C.Y. Wong[2], G.D. Housley[2], E.C. Hardeman[1], S.J. Palmer[1].* 1) Cellular and Genetic Medicine Unit, School of Medical Sciences (SOMS), UNSW Australia, Sydney, Australia; 2) Translational Neuroscience Facility & Department of Physiology, School of Medical Sciences (SOMS), UNSW Australia, Sydney, Australia.

Amongst the spectrum of physical and neurological defects in Williams Beuren Syndrome (WBS), it is common to find a distinctive response to sound stimuli that includes extreme adverse reactions to loud or sudden sounds (usually called 'hyperacusis'), and a fascination with certain sounds that may manifest as strengths in musical ability. However, hearing tests have indicated that sensorineural hearing loss (SNHL) is frequently found in WBS patients and the genetic origins of this unusual auditory phenotype are currently unknown. However, the use of the term 'hyperacusis' is inaccurate since hyperacuity implies increased functional sensitivity leading to detectably lower hearing thresholds, and 'auditory allodynia', meaning aversion to, or fear of certain sounds that are usually acceptable to others, is a more accurate description of this condition. Here, we investigated the involvement of *GTF2IRD1*, a gene located within the WBS deletion that has been implicated as a contributor to the WBS neurocognitive profile and craniofacial abnormalities. Using knockout mice, we analysed the expression of the gene in the inner ear and examined hearing capacity. Rather than using acoustic tests in mice that rely on behavioural responses, which could be confounded by the altered anxiety responses previously reported in *Gtf2ird1* knockouts, we used objective physiological assays under general anaesthesia to evaluate hearing capacity, minimizing the behavioural component. Using auditory brainstem response (ABR), which measures sound-evoked auditory neurotransmission from the cochlear nerve to the auditory midbrain, and distortion product of otoacoustic emissions (DPOAE), which measures sound-evoked electromechanical amplification via the outer hair cell 'cochlear amplifier', we demonstrated that *Gtf2ird1* null mice have hypoacusis (higher hearing thresholds) in both assessments. These data, together with the *Gtf2ird1* expression pattern in diverse cell types within the inner ear, indicate that the principal hearing deficit in the mice can be traced to impairments in cochlear amplifier, suggesting that similar mechanisms may underpin the SNHL experienced by WBS patients. The origin of the auditory allodynia remains unexplained but may form part of a central defect in the processing of sensory input and/or emotional control mechanisms, which *GTF2IRD1* may also contribute to.

856S

Genetic variant at ETS1 locus increases lupus risk and affects Stat1 binding. *L. Kottyan[1,2], X. Lu[1,3], E. Zoller[1], M. Weirauch[1,3], B. Namjou[1], K. Greis[4], N. Shen[1,5], K. Kaufman[1,2], J. Harley[1,2,3], The International Consortium on the Genetics of Systemic Lupus Erythematosus (SLEGEN).* 1) Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) United States Department of Veterans Affairs Medical Center, Cincinnati, Ohio, USA; 3) Immunology Graduate Program, University of Cincinnati College of Medicine, Cincinnati, OH; 4) Cancer Cell Biology, University of Cincinnati Medical Center, Cincinnati, OH; 5) Institute of Rheumatology, Shanghai Jiao Tong University School of Medicine, Shanghai, China.

Over the past 10 years, genetic association studies have identified over 55 systemic lupus erythematous (SLE)-risk loci. Genetic variants in ETS1 have been associated with SLE in several independent studies in populations of Asian ancestry. Several recent studies have implicated ETS1 as a critical driver of immune cell function; indeed, mice deficient in ETS1 develop an SLE-like autoimmunity. rs1128334, in the 3' UTR of ETS1, has been associated with the decreased expression of ETS1. We performed a fine-mapping study of over 20,000 subjects of two multi-ancestral cohorts using genotyped and imputed variants spanning the ETS1 locus. By constructing genetic models using frequentist and Bayesian association methods, we identified a set of variants that are most likely to be causal. Of the three variants that we tested, only rs6590330 differentially binds lysate from B cell lines as assessed by electrophoretic mobility shift assays and DNA affinity precipitation assays. We found that the active form of the transcription factor STAT1 binds the risk allele of rs6590330 but not the non-risk allele using mass spectrometry, a finding confirmed by Western blot analysis. rs6590330 is in strong linkage disequilibrium with rs1128334 and is predicted to also be associated with decreased ETS1 expression. We propose a model in which the risk allele of rs6590330 increases SLE risk by binding pSTAT1 and depressing the expression of ETS1.

857M

Rs738409 Polymorphism in PNPLA3 Gene is Associated with Lower Insulin Resistance in Korean Men. *JH. Park[1], JM. Yun[1], HT. Kwon[2].* 1) Family medicine, Seoul National University Hospital, Seoul, South Korea, MD; 2) Family Medicine, Healthcare Research Institute, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, South Korea, MD.

Background & Aims: The rs738409 polymorphism in adiponutrin is a well-known genetic risk factor for nonalcoholic fatty liver disease (NAFLD) development. When we consider that NAFLD is closely associated with insulin resistance, the association between rs738409 variant and insulin resistance is highly suspected. Methods: We enrolled 1,189 Korean men who visited the Health Promotion Center of Seoul National University in Seoul, South Korea from December 2009 to June 2012 and 1,189. Hepatic steatosis was evaluated by abdominal ultrasound and subjects with secondary causes of NAFLD or alcohol consumption of more than 30g/day were excluded from NAFLD classification. Serum glucose and insulin levels were measured using overnight fasting blood sample. The homeostasis model assessment estimated insulin resistance (HOMA-IR) index was calculated by multiplying glucose in mg/dl by insulin in μU/ml/405. The cutoff value of 2.56 was used for insulin resistance diagnosis as previously reported in Korean men. Rs738409 genotyping was performed using TaqMan assay on a ViiATM 7 Real-Time PCR System. Results: The participants were predominantly middle-aged men (49.1 ±7.0 years; range, 30-60 years), and the frequencies of NAFLD and insulin resistance were 43.7% and 17.5%. The minor allele frequency of the rs738409 G allele was 0.43.The rs738409 G allele was not associated with insulin resistance status (Fisher's exact P = 0.653) and HOMA-IR index level (one-way ANOVA P = 0.068) in the analysis without adjustment for NAFLD status by genotypes. We performed multivariate regression analysis after adjustment for NAFLD status. The rs738409 CG and GG genotype showed lower mean levels of HOMA-IR index compared to the CC genotype (P = 0.031 and < 0.001, respectively. Then, we checked the effect of the rs738409 G allele on HOMA-IR index levels in each NAFLD and non-NAFLD subgroup stratified by NAFLD status using one-way ANOVA. In the non-NAFLD group, the CG and GG genotypes compared to the CC genotype showed significantly lower mean levels of HOMA-IR index in an additive manner (Sidak-corrected P = 0.018 and < 0.001, respectively); however, in the NAFLD group, those inverse associations were not shown. Conclusion: The adiponutrin polymorphism rs738409 known as a NAFLD genetic risk factor was also associated with decreased levels of insulin resistance assessed by HOMA-IR index, and the inverse association was prominent in subjects without NAFLD.

858T

Extended Haplotypes of controls regions of the HLA-G are associated with type 1 diabetes mellitus. *R. De Albuquerque[1, 2, 3], N. Lucena-Silva[2], D. Meire Rassi[4], E. Cruz Castelli[5], S. Gregori[3], E. Donadi[4], C. Teixeira Mendes-Junior[1].* 1) Faculty of Medicine of Ribeirão Preto, University of São Paulo, Departament of Genetic, Ribeirão Preto-SP, Brazil; 2) Aggeu Magalhães Research Center, Department of Immunology, Oswaldo Cruz Fundation (Fiocruz), Recife-PE, Brazil; 3) Division of Regenerative Medicine, Stem Cells and Gene Therapy, San Raffaele Telethon Institute for Gene Therapy (HSR-TIGET), San Raffaele Scientific Institute, Milan, Italy; 4) Faculty of Medicine of Ribeirão Preto, University of São Paulo, Department of Medicine, Ribeirão Preto-SP, Brazil; 5) Departamento de Patologia, Faculdade de Medicina de Botucatu, Universidade Estadual Paulista Júlio de Mesquita Filho, Botucatu, Brazil.

The TD1 (Type 1 Diabetes) is a multifactorial disease in which the autoimmune destruction of pancreatic cells is mediated by humoral and cellular mechanisms. Various subpopulations of lymphocytes are involved in this process, however, the CD8 + T lymphocytes appear to be the cells responsible for early insulitis. Genetic susceptibility to TD1 has been primarily attributed to MHC genes: a study mapping the MHC region identified the HLA-G region as an independent susceptibility gene for the disease. A major function of HLA-G is the inhibition of Natural Killer cells (NK) and cytotoxic T cells (CTL), by interaction with inhibitory receptors found in these cells. The HLA-G gene presents a 3' untranslated region (3'UTR) that contains some polymorphisms, in this study, we analyzed the 9 most studied, 3 of them related with several posttranscriptional regulatory elements. On the other hand, the transcription level is driven by the 5' regulatory region through of the transcription factor, which are target for some of the 28 SNPs that we studied. We analized extended Haplotypes of those both controls regions of the HLA-G in 120 patients with TD1 and 100 controls from the same geographical region of the patients. DNA was extracted by the salting out. The 5' regulatory region and 3'UTR fragments were amplified using the polymerase chain reaction (PCR) and were directly sequenced in an ABI310 Genetic Analyzer. The presence of significant associations between the SNPs detected of the HLA-G gene was evaluated by means of a likelihood ratio test of linkage disequilibrium (LD), using the ARLEQUIN. The PHASE method and expectation-maximization (EM) algorithm were used to infer the haplotypes. Regarding 3'UTR haplotypes, we observed two significant haplotypes, a conferring susceptibility to TD1 (UTR17), and another suggesting a protective role to TD1 (UTR-3). On 5' URR, we found four significant haplotypes, two for susceptibility (G0104a, G010101a) and two for protective role(G0104a, G010101c). Thus, with relation extended haplotypes, suggesting a balance between these regions, which were according to the previous papers, but we found two new haplotypes (PG0104b + UTR-13, G0103f + UTR-17), on the total two conferring protection (G010101c + UTR-4, G0104a + UTR-3), and one of the new (G0103f + UTR-17) conferring susceptibility. The present study revealed significant associations between and haplotypes of controls regions of the HLA-G with TD1 susceptibility.

859S

Sequencing of the *TBX6* Gene in Families with Familial Idiopathic Scoliosis. *E.E. Baschal[1], K. Swindle[1], C.M. Justice[2], A. Perera[3], R.M. Baschal[1], A. Poole[1], O. Pourquié[4], O. Tassy[4], N.H. Miller[1].* 1) Department of Orthopedics, University of Colorado Denver Anschutz Medical Campus, Aurora, CO; 2) Genometrics Section, Computational and Statistical Genomics Branch, NHGRI, NIH, Baltimore, Maryland, USA; 3) Molecular Biology Facility, Stowers Institute, Kansas City, USA; 4) d)Department of Cell Biology and Development, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Strasbourg, France.

Idiopathic scoliosis, which affects 3% of the population, is a structural curvature of the spine whose underlying genetic etiology has not been established. Congenital scoliosis is the presence of vertebral malformations as identified on radiographs. Idiopathic scoliosis has been reported to occur at a higher rate than expected by chance alone in family members of individuals with congenital scoliosis, indicating that the two diseases might have a shared etiology. The *TBX6* gene on chromosome 16p, essential to somite development, has been associated with congenital scoliosis in a Chinese population. The 16p11.2 locus encompasses several unique genomic structural variants and several reports identify an increased incidence of scoliosis and vertebral anomalies in patients with 16p11.2 rearrangements. Previous studies have identified linkage to this locus on 16p in a sample of families with familial idiopathic scoliosis (FIS) segregating with an autosomal dominant pattern of inheritance. Given these factors, we sequenced the *TBX6* gene in a sample of families with FIS to determine the potential association between *TBX6* variants and FIS. We selected parent-offspring trios from 13 families (42 individuals) with FIS for Sanger sequencing of the *TBX6* exons. The trios were selected from a large population of families with FIS in which a genome-wide scan had resulted in linkage to 16p. Selection criteria required that each trio contain both an affected proband and an affected parent, each with a scoliotic curvature of ≥25°. Sequencing analyses of this familial subset resulted in the identification of five coding variants. All 13 families had at least one variant. Three of the five variants were novel; the remaining two variants were previously characterized and accounted for 90% of the observed variants in these families. In all cases, there was no correlation between transmission of the *TBX6* variant allele and FIS phenotype. Transmission of sequence variants within the *TBX6* gene was not correlated with the FIS phenotype in this select study population. However, these findings do not rule out the *TBX6* gene as a whole, as a causative variant could be located in a region of the gene that was not sequenced in this study. There are also several other candidate genes in the region. Further investigation of the *TBX6* gene and additional genes within this region is required in order to delineate their potential roles in the etiology of FIS and axial skeletal development.

860M

Association of variants in *GALNT10* and related pathway genes with body mass index in African Americans. *M. Stromberg[1, 2, 3], P. Mudgal[3], B.I. Freedman[4], D.W. Bowden[2, 3, 5, 6], M.C.Y. Ng[2, 3].* 1) Molecular Genetics & Genomics, Wake Forest University, Winston Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Internal Medicine - Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC; 5) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 6) Internal Medicine-Endocrinology, Wake Forest School of Medicine, Winston-Salem, NC.

Obesity is a chronic health problem which has a higher prevalence in African Americans (AAs) (35.7%) compared to non-Hispanic whites (23.7%) in the United States. Genetic contributors to obesity in AAs remain poorly understood, although recent genome-wide association studies (GWAS) identified several regions including the *GALNT10* locus as being associated with body mass index (BMI) in AAs. *GALNT10* is a member of the N-acetylgalactosaminyltransferase family which is located in the Golgi apparatus and is expressed in the hypothalamus of the brain, which regulates hunger. There are 20 members in the GALNT family, each with a unique expression pattern. Previous studies have not identified the causal variant(s)at *GALNT10* and have not examined biological candidate genes from the relevant pathways in detail. As such, this study assessed 71 candidate genes for association with BMI in 2,293 African American subjects. Genotyping was performed using an Affymetrix Biobank array customized for fine-mapping of GWAS and candidate loci related to metabolic traits. Single SNP association was performed in two cohorts using linear regression under an additive model with adjustment for age, sex and the first principal component followed by meta-analysis. Nominal significance (P<0.0005) was observed at *ATF4* (p=0.000337), *GALNT18* (p=7.91E-06), *ST3GAL1* (p= 0.000191), *WBSCR17* (p=0.000215), with the strongest association at *GALNT18*. *GALNT18*, or *GALNTL4*, has been previously associated with Type 2 Diabetes in AAs. The most significant association observed at the *GALNT10* locus is rs1346482 (p=0.001015). Previously associated SNPs in this locus (rs815611, rs2033195, rs7708584) are shown to have pairwise r^2 values ranging from 0.32-0.84 (0.84, 0.68, 0.32 respectively) in the YRI population. These findings suggest that several members of the GALNT family and related pathway members may play roles in modulating adiposity in AAs, although further replication is needed.

861T

Targeted Sequencing of an Admixture Mapping Peak in Latinos Implicates Rare Non-coding Variation in Asthma Susceptibility. *D.G. Torgerson[1], M. Pino-Yanes[1], C.R. Gignoux[1], C. Eng[1], E.G. Burchard[1,2] on behalf of the GALA II Investigators.* 1) Department of Medicine, University of California San Francisco, San Francisco, CA; 2) Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA.

The prevalence and morbidity of asthma varies widely among global populations and racial/ethnic groups within the U.S. We previously identified a genome-wide significant admixture mapping peak at 18q21.1, whereby local Native American ancestry was associated with an increased risk of asthma in a meta-analysis of 3,902 Hispanic/Latino individuals. The ancestry association was replicated in a large independent study of 3,774 Latinos; however, the variation behind the ancestry association remains to be identified. Towards this goal, we sequenced a 342 Kb contiguous region across the top of the admixture mapping peak including all coding sequences (including *SMAD2* and *ZBTB7C*) and complete non-coding/intergenic sequences in 1,978 Mexicans and Puerto Ricans with and without asthma from the Genes-environments & Admixture in Latino Americans study (GALA II study). No new individual variant was found to be associated with asthma in either Mexicans or Puerto Ricans adjusting for global and local ancestry, or in a meta-analysis across populations. Collapsing tests showed no significant contribution of coding variation in either *SMAD2* or *ZBTB7C*. However, we identified a significant cumulative contribution of rare, non-coding variants 5' upstream of *SMAD2* in Mexicans using the sequence kernel association test (SKAT test, permutation $p<10^{-4}$) by employing a sliding window approach to grouping variants. The association was specific to individuals of Mexican ethnicity, and was consistent in location to the most significant allelic associations at common variants identified through traditional genome-wide association testing. Using rtPCR in a subset of individuals, we also tested for the presence of individual and cumulative rare variant eQTLs for *SMAD2* and *ZBTB7C*. Overall we identified a significant contribution of rare, non-coding variation to asthma susceptibility that are likely to effect the transcription of *SMAD2*. Furthermore, the association appears to be population-specific, consistent with the hypothesis that admixture mapping can accelerate the search for rare, population-specific variation that contributes to complex disease.

862S

Replication and fine-mapping of trait-stratified genome-wide association study identifies novel genetic associations with cytokine phenotypes in systemic lupus erythematosus. *T.B. Niewold[1], M. Imgruet[1], Y. Ghodke-Puranik[1], J.M. Dorschner[1], J.A. Kelly[2], M. Marion[3], J.M. Guthridge[2], C.D. Langefeld[3], J.B. Harley[4], J.A. James[2], K.L. Sivils[2].* 1) Rheumatology and Immunology, Mayo Clinic, Rochester, MN; 2) Arthritis & Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK; 3) Department of Biostatistical Sciences, Wake Forest University, Winston-Salem, NC; 4) Cincinnati Children's Hospital Medical Center and Cincinnati VA Medical Center, Cincinnati, OH.

Introduction: Systemic lupus erythematosus (SLE) is a highly heterogeneous disorder, characterized by differences in serum cytokines and clinical manifestations. High serum interferon alpha (IFN-α) is an important heritable phenotype in SLE which is involved in primary disease pathogenesis, and this phenotype accounts for some of the biological heterogeneity between SLE patients. The aim of the study was to replicate and fine-map previously detected genetic associations with serum IFN-α in SLE. Methods: We previously undertook a case-case genome-wide association study of SLE patients stratified by ancestry and extremes of phenotype in serum IFN-α. Single nucleotide polymorphisms (SNPs) in seven loci identified in this screen were selected for follow up in a large independent cohort of more than 1300 SLE patients (672 European-ancestry, 420 African ancestry, and 231 Hispanic/Amerindian ancestry). Each ancestral background was analyzed separately, and a panel of 334 ancestry-informative markers was used to control for ancestry and admixture. Results: SNPs in the *PPM1H* and *EFNA5* loci demonstrated strong association with serum IFN-α in eQTL analyses (p<10-4 for each association) in European ancestry. In African ancestry subjects, SNPs in both *LPAR1* and *EFNA5* were associated with serum IFN-α, and one of the SNPs was common between European- and African- ancestry subjects. In Hispanic/Amerindian ancestry patients, *LRRC20* demonstrated nominal evidence for association with serum IFN-α. Conclusions: This study demonstrates the power of using a serologic subphenotype to elucidate genetic factors involved in complex autoimmune disease. The distinct associations observed in different ancestral backgrounds emphasize the heterogeneity of molecular pathogenesis in SLE, and the need for stratification by subphenotypes in genetic studies. We hypothesize that these genetic variants play a role in disease manifestations and severity in SLE.

863M

Refinement of association signals and residual heritability in host control of HIV viral load. *P.J. McLaren[1,2], J. Fellay[1,2], the International Collaboration for the Genomics of HIV.* 1) School of Life Science, EPFL, Lausanne, Vaud, Switzerland; 2) Swiss Institute of bioinformatics, Lausanne, Vaud, Switzerland.

Background: Genome-wide association studies of HIV outcome consistently identify the MHC region as the major genetic influence on disease progression. Through establishment of the International Collaboration for the Genomics of HIV we sought to bring together all existing GWAS data in HIV patients to maximize power to uncover further association signals.

Methods: Genome-wide SNP data were collected from 25 clinical centers. Plasma viral load measurements obtained during the chronic phase of untreated infection were available for 6,315 individuals of European ancestry. Missing genotypes were imputed using the 1,000 Genomes Project reference. Single marker association testing was performed per study using linear regression and combined across studies by meta-analysis. Classical HLA alleles and amino acid variants were imputed using the SNP2HLA pipeline. Heritability estimates were calculated using GCTA.

Results: Consistent with previous studies, the top association signal was amino acid position 97 in HLA-B (p=4e-143), with independent associations at positions 67 (p=4e-112) and 45 (p=8e-49). Controlling for these positions fully accounted for classical HLA-B allele associations. Residual association at rs9264942, a marker of HLA-C expression level, was retained after controlling for amino acid effects (conditional p=1e-11). Outside of the MHC, a second peak of association in the CCR5 region (top SNP rs4317138 p=8e-19) was observed. Conditioning for the known effects of CCR5Δ32 and CCR2V64I we found evidence for residual association at rs4683217 (p=5e-10) upstream of CCR2, suggesting the presence of additional functional variants in this region. Considering all SNPs and assuming an additive genetic model, we obtained a narrow-sense heritability (h2) estimate of 26% (p<2e-16) for viral load. Removing the HLA and CCR5 regions, we observed a reduced, yet still significant h2 estimate of 16% (p=5e-8).

Conclusions: By combining available GWAS data in HIV infected individuals we refined the known association signals in HLA-B and CCR5. Controlling for these main effects, we uncovered additional, independent associations in known regions. Heritability analysis suggested that variation outside known regions, captured through common SNPs, also contribute to HIV control.

864T

Study of genetic risk factors for susceptibility to leprosy - the chromosomal region 6q25-q27 revisited. *G.B. RAMOS[1], H. SALOMAO[1], A.S. FRANCIO[1], C.C. CARDOSO[4,5], M.O. MORAES[4], A.C. PEREIRA[3], R.I. WERNECK[2], M.T. MIRA[1,2].* 1) School of Medicine, Pontifícia Universidade Católica do Paraná, Curitiba, Paraná, Brazil; 2) Health and Biosciencies School, Pontificia Universidade Católica do Paraná, Curitiba, Paraná, Brazil; 3) Instituto Lauro de Souza Lima, Bauru, São Paulo, Brazil; 4) Laboratório de Hanseníase, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Rio de Janeiro,; 5) Laboratório de Virologia Molecular, Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

Leprosy, a chronic infectious disease, affects approximately 220,000 new individuals worldwide every year. Genetic studies have identified several genes and genomic regions contributing to the control of host susceptibility to leprosy. In 2004, a study identified variants of genes *PARK2/PACRG* associated with the disease, but the association signal observed does not seem to entirely explain the genetic effect observed for this locus, suggesting the presence of additional candidates genes. Here, we test variants of *SOD2*, located near to *PARK2/PACRG* at the 6q25-q27 region, for association with leprosy in four Brazilian population samples of distinct design and ethnic background. The primary population included 186 volunteers distributed in 71 trios (55 independent) composed by one affected individual and their parents, recruited at the Prata Village, located at the Amazonic state of Pará; the replication study involved three case-control samples from the southern, southeastern and central-western regions of Brazil: Curitiba, Paraná (198 cases and 183 controls); Rio de Janeiro, Rio de Janeiro (507 cases and 236 controls) and Rondonópolis, Mato Grosso (411 cases and 424 controls), respectively. SNP markers covering the entire *SOD2* locus were genotyped by fluorescence-based TaqMan and/or Sequenon technology. Family-based association analysis was performed as implemented in the software FBAT, version 2.0.4. Logistic regression was applied to the case-control samples, as implemented in the software SAS, version 9.1. Family-based analysis revealed borderline evidence of association, after adjusting for age and gender (p_c), between leprosy and alleles of markers rs5746136 (p_c = 0.07; OR = 2.084), rs4880 (p_c = 0.07; OR = 0.256) and rs295340 (p_c = 0.042; OR = 2.637). Findings were similar in the population sample of Curitiba: rs5746136 (p_c = 0.15; OR = 1.394), rs4880 (p_c = 0.03; OR = 0.578) and rs295340 (p_c = 0.04; OR = 1.855). No association was found in the population samples of Rondonópolis and Rio de Janeiro. These preliminary results suggest evidence of association between *SOD2* with leprosy in at least two independent, ethnically distinct Brazilian population samples; further analysis of LD patterns and association using a combined population sample is currently ongoing.

865S

Next-generation sequencing and targeted fine linkage disequilibrium mapping reveal FREM1 mutations associated with HIV acquisition in a sub-Saharan African cohort of female sex workers. *J.F. Tuff[1], D. La[1], N. Klaponski[2], V. Ly[2], A. Yuen[1], B. Liang[2], S. Tyler[2], F.A. Plummer[2,3], M. Luo[1,3].* 1) JC Wilt Infectious Diseases Research Centre, Winnipeg, MB, Canada; 2) National Microbiology Laboratory, Winnipeg, MB, Canada; 3) University of Manitoba, Department of Medical Microbiology, Winnipeg, MB, Canada.

There is considerable heterogeneity in susceptibility to HIV-1 (HIV) infection among individuals. This variability in risk can be attributed to a combination of environmental, viral, and host genetics. A group of women in the Pumwani Sexworker (ML) Cohort, established in Nairobi, Kenya, exhibit a HIV-exposed seronegative (HESN) phenotype, despite repeated exposure through high-risk sex work. We previously described the identification of an intronic single nucleotide polymorphism (SNP) (rs1552896; NM_144966.5:c.1881+58G>C) in FRAS1 related extracellular matrix 1 (*FREM1* [MIM: 608944]) as the highest-ranking correlate of this protective phenotype in a low-resolution genome-wide association study. To further investigate this association, we herein sought to identify *FREM1* mutations in linkage disequilibrium (LD) with rs1552896, followed by fine mapping of polymorphic loci deemed most likely to be functionally causal as determined by SIFT/PolyPhen-2 or ENCODE via RegulomeDB. We conducted pyrosequencing of *FREM1* (n=69) with Roche 454 GS FLX high throughput sequencing technology, and identified 1576 polymorphisms, 966 of which were newly discovered. Four of the 69 polymorphisms showing putative LD with rs1552896 (LOD ≥2) were chosen for fine mapping via PCR and Sanger sequencing: (i) rs2779500 (NM_144966.5:c.1315G>C, NP_659403.4:p.Val-439Leu (n=541)), (ii) rs1353223 (NM_144966.5:c.1495A>G, NP_659403.4:p.Ile499Val (n=633)), (iii) rs10810271 (NM_144966.5:c.1262-1086A>G, (n=347)) and (iv) a novel microsatellite allele (c.1261+1207CCCT[8]TCCT[7]; Mi8*7, (n=1956)) found within a newly characterized complex tandem tetranucleotide repeat (CTTR). Fine mapping confirmed significant LD between rs1552896 and each of rs2779500 (LOD 6.94, D' 0.993, r^2 0.048), rs1353223 (LOD 3.92, D' 1.0, r^2 0.019), rs10810271 (LOD 5.29, D' 1.0, r^2 0.049) and Mi8*7 (LOD 85.37, D' 0.59, r^2 0.348) but only Mi8*7 was shown to associate with HESN ($p=7.50 \times 10^{-3}$, OR=2.17, 95% CI 1.21-3.88). Analysis of samples successfully genotyped at both the rs1552896 and CTTR loci (n=1046) revealed that Mi8*7 is more strongly associated with HESN ($p=7.67 \times 10^{-3}$, OR=2.18, 95% CI 1.21-3.90) than is rs1552896 (p=8.77 x 10-3, OR=2.15, 95% CI 1.20-3.84) suggesting that Mi8*7 is a better marker for HESN in FREM1. We further show that Mi8*7 belongs to a subgroup of *FREM1* CTTR ($(CCCT)_{any}(TCCT)_{7-12}$) that, when homozygous, is also correlated with the HESN phenotype ($p=1.99 \times 10^{-4}$, OR=2.22, 95% CI 1.45-3.40).

866M

The Role of *ALDH7A1* in Body Composition among West Africans. *A.R. Bentley, G.J. Chen, D. Shriner, A.P. Doumatey, A.A. Adebowale, C.N. Rotimi.* Center for Research in Genomics and Global Health, Natl Human Genome Research Institute, Bethesda, MD.

West Africa is experiencing some of the fastest growth in Type 2 Diabetes (T2D) prevalence in the world. It is believed that much of this increase is secondary to ongoing changes in body composition (BC). In this study, we investigated the association of BC indices (BMI, Fat Mass [FM], Fat Free Mass [FFM], and % Fat Mass [PFM]) with a locus (5 q22-q31) previously identified in a T2D linkage analysis in West Africans (WA). Previously we performed fine-mapping of this region using 1405 SNPs genotyped in 931 WA. Seven SNPs were associated with at least one BC index, with the strongest association for rs2306617 (*ALDH7A1*) and BMI (β 1.3, $p=1.0 \times 10^{-6}$). Based on these prior results, we conducted targeted resequencing of *ALDH7A1* in 142 WA (48 obese, T2D; 48 non-obese, T2D; 48 obese, non-T2D). Identified SNPs were genotyped in 1726 WA, and the 37 SNPs that met quality control standards were evaluated in linear regression models. Three variants were associated with the BC indices, with the most striking association for rs3736174 with FFM (β 2.1 kg, $p=9.9 \times 10^{-11}$) and BMI (β 0.9, $p=2.8 \times 10^{-5}$). Other associated variants include rs1138005 with FFM (β -2.2 kg, $p=2.6 \times 10^{-4}$) and rs7447380 with BMI (β 2.7, p=0.001). Of note, these variants have been annotated for promoter and enhancer marks, with rs3736174, in particular, influencing gene expression in cell types of relevance to BC (chondrocytes, adipocytes, and mesenchymal stem cells). The stronger results with FFM, and the previous identification of *ALDH7A1* in a GWAS of osteoporosis, led us to investigate the association of this gene on bone phenotypes. We evaluated phenotype data for Aldh7a1 knock-out mice (Jackson Laboratory, Project NIH-0076). Among the 16 mice on which DEXA scans had been performed, the null genotype was associated with a higher ratio of bone mineral content to lean body mass (p=0.04). The null genotype was not associated with markers of adiposity. The associations of this locus with FFM in WA and bone mineral content in mice suggest that BMI associations with *ALDH7A1* may, in fact, reflect an influence of this locus on body weight through bone mineralization. *ALDH7A1* encodes an aldehyde dehydrogenase, and reactive aldehydes have been shown to affect bone remodeling. Taken together, this work suggests that *ALDH7A1* influences body composition, which may result from perturbations in bone mineralization.

867T

Targeted sequencing of GWAS loci: insight into genetic etiology of cleft lip and palate. *E.J. Leslie[1], M.A. Taub[2], Q. Zhang[3], K. Meltz Steinberg[3], D.C. Koboldt[3], Y. Kousa[4], D.E. Larson[3], J. Bidinger[5], H. Wang[5], R.S. Fulton[3], R.A. Cornell[6], B.C. Schutte[4], W.D. Fakhouri[7], G.L. Wehby[8], L.M. Moreno[9], I. Ruczinski[2], J.T. Hecht[10], A. Scott[5], K. Christensen[11], A.C. Lidral[9], T.H. Beaty[5], G.W. Weinstock[3], J.C. Murray[12], M.L. Marazita[1].* 1) Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Dept. Biostatistics, Johns Hopkins University; 3) The Genome Institute, Washington University-St. Louis; 4) Microbiology and Molecular Genetics, Michigan State University; 5) Dept. Epidemiology, Johns Hopkins University; 6) Dept. Anatomy and Cell Biology, University of Iowa; 7) Dept. Diagnostic and Biomedical Sciences, University of Texas Health Science Center at Houston; 8) Dept. of Health Management and Policy, University of Iowa; 9) Dept. Orthodontics, University of Iowa; 10) Dept. Pediatrics, University of Texas Health Science Center at Houston; 11) University of Southern Denmark; 12) Dept. Pediatrics, University of Iowa.

Nonsyndromic cleft lip with or without cleft palate (CL/P) is a common birth defect with complex inheritance. Despite substantial progress in identifying risk loci by multiple genome-wide association studies (GWAS), the causal variants at each locus remain unidentified. To this end, we performed the first targeted sequencing of GWAS loci for CL/P. We sequenced thirteen regions from GWAS and candidate gene studies, totaling 6.3Mb, in 1521 Asian and Caucasian trios with CL/P and performed statistical analyses on *de novo* variants (DNVs) and single nucleotide variants (SNVs). We identified 123 novel DNVs, including a missense DNV in the *PAX7* DNA-binding domain disrupting DNA-binding by EMSA. This *PAX7* DNV is only the third CL/P coding DNV ever reported. Multiple DNVs occurred in putative regulatory elements or confirmed craniofacial enhancers. Specifically, one DNV occurred in a human neural crest enhancer downstream of *FGFR2* and two DNVs occurred in another enhancer near *HHAT*. To determine if rare coding SNVs were over-transmitted to affected offspring we performed gene-based burden tests. Although there were no significant results after correcting for multiple testing, there are many compelling SNVs worthy of follow-up in functional studies. For example, we identified four nonsense SNVs in *ARHGAP29*, the gene previously identified as the best CL/P candidate at the 1p22 region. We also used ScanTrio to identify windows containing significant over- or under-transmission of rare SNVs and assessed significance through permutations of transmitted and untransmitted haplotypes. This analysis revealed promising clusters of rare SNVs near GWAS hits on 8q24 and *NOG*. Further, our analysis of common SNVs recapitulated the GWAS results and identified multiple highly associated SNVs in craniofacial regulatory elements on 8q24 and near *NOG* and *NTN1*. While cleft palate has not been reported in *Ntn1*$^{-/-}$ mice, we note that such mice die perinatally from inability to feed, a common issue in mice with a cleft. To support the role of *Nog* and *Ntn1* in CL/P, we analyzed the expression patterns of these signaling molecules. Nog is expressed in oral periderm, a critical cell type for palatal elevation and fusion, while Ntn1 is expressed in palatal mesenchyme, localizing to the basement membrane of the medial edge epithelium and the oral surface of the palatal shelves. In aggregate this targeted sequencing study has identified several regions and variants that may be causal for CL/P.

868S

Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *G. Kichaev[1], WY. Yang[2], S. Lindstrom[3], F. Hormozdiari[2], E. Eskin[1,2,5], AL. Price[3,4], P. Kraft[3,4], B. Pasaniuc[1,5,6].* 1) Bioinformatics Inter-departmental Program, UCLA, Los Angeles, CA., USA; 2) Dept of Computer Science, UCLA, Los Angeles, CA., USA; 3) Program in Genetic Epidemiology and Statistical Genetics, Harvard School of Public Health, Boston MA., USA; 4) Dept of Biostatistics, Harvard School of Public Healthm Boston MA., USA; 5) Dept of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA., USA; 6) Dept of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA., USA.

Standard statistical approaches for prioritization of variants for functional testing in fine-mapping studies either use marginal association statistics or estimate posterior probabilities for variants to be causal under simplifying assumptions. Here we present a probabilistic framework that integrates association strength with genomic functional annotation data (e.g. ENCODE) to improve accuracy in selecting plausible causal variants for functional validation. A key feature of our approach is that it empirically estimates the contribution of each functional annotation to the trait of interest while allowing for multiple causal variants at any risk locus. We devise efficient algorithms that estimate the parameters of our model across all risk loci to further increase performance. In fine-mapping simulations based on 1000 Genomes data our approach reduces the average number of SNPs to be tested in functional assays to identify 90% of all causal SNPs from an average of 12.3 variants per locus when ranking by marginal association statistics (or 25.0 when ranking by posterior probabilities under single causal variant assumption [Maller et al 2012]) to 9.7 variants per locus. We validate our findings using data from a large scale meta-analysis of four blood lipids traits [Teslovich et al 2010] and find that the relative probability for causality is increased for variants in exons and transcription start sites and decreased in repressed genomic regions at the risk loci of these traits. Using these highly predictive trait-specific functional annotations, we estimate causality probabilities across all traits and variants, reducing the size of the 90% confidence set from an average of 17.5 to 13.5 variants per locus in this data.

869M

Analysis of genes involved in carnitine metabolism and functions in autistic patients by targeted sequencing. *J. Ge[1], X. Wang[1], H. Cui[1], B. Zhang[1], H. Xu[2], P. Fang[1], A. Beaudet[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Fudan Unviersity, Shanghai, China.

A recent study reported the link between the deficiency of the trimethyllysine hydroxylase epsilon (TMLHE), the first enzyme on the carnitine biosynthesis pathway, and the risk of autism, raising the possibility that carnitine might be important in the autism etiology. The well-established function of carnitine is its role in fatty acid metabolism as a molecule transporting the fatty acids and acetate across the mitochondrial membrane. The carnitine homeostasis in humans is maintained through intake from dietary sources, endogenous synthesis, and reabsorption in the kidney. To understand whether carnitine has influence on the risk of autism, we analyzed the genes functioning in carnitine biosynthesis, carnitine transporters, carnitine-acylcarnitine acyltransferases, carnitine acylcarnitine translocases, as well as several other genes that can influence carnitine metabolism, such as genes that can cause secondary carnitine deficiency. Using the Fluidigm Access Array™ System for high-throughput PCR amplification and Illumina MiSeq next-generation sequencing, we have sequenced the exons of these 23 genes in 384 autistic probands from the Simons Simplex Collection (SSC). Totals of 108 heterozygous and 2 hemizygous protein changing rare mutations were found as follows: ACADL: c.928_9294 times; CPT1C:p.M798V and ALDH9A1: p.A9G 3times; and CPT2:p.T150A, SLC22A5:p.D122Y, ACADM: p.D266G and ACADM: p.D270G 2 times. The two hemizygous mutations are SLC6A14: p.T418S and SLC25A43: p.R323Q. Interestingly, we found that 2 patients carrying 3 heterzygous mutations in different genes, and 18 patients carrying 2 heterzygous mutations. We are now checking the inheritance of these mutations and expanding our study to the full SSC cohort.

870T

Genetic contributions to obesity and metabolic risk in Mexican-American children. *R.J. Mudgway, S.L. Spilman, A.P. Mallya, J.T. McCracken, E.L. Nurmi.* Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA.

Background: Hispanic populations have among the highest rates of obesity compared to other racial/ethnic groups. Genetic moderators of weight and metabolic factors may help explain vulnerability to obesity in these higher risk populations. Previous studies have identified signaling networks involved in obesity risk, including energy balance, monoaminergic, and growth factor systems. We examined common genetic variants in these candidate pathways for association with BMI and metabolic factors in a sample of Mexican-American children. **Methods:** Baseline data from 356 Mexican-American children 6-12 years old from the Los Angeles area were collected from matched overweight and lean groups. Physiological measures were recorded for each subject, including BMI, fat mass, physical measurements, serum chemistry (such as lipid and glucose levels), respiratory function, and resting energy expenditure. Families underwent a month-long nutrition and diet intervention and returned for follow-up measures. Association between metabolic phenotypes and complete common genetic variantion in energy balance candidates (MC4R, NPY), monoamine receptors (ADRA2A, HTR2C, DRD1-5), and brain-derived neurotrophic factor (BDNF). Key variants showing association with obesity risk in prior literature (FTO and PRKAR2B) were also included. **Results:** In agreement with published reports in Caucasians, the minor (C) allele of rs1421085 in FTO, was associated with increased BMI in Mexican-American children (p=0.027) while the minor allele (G) of rs13224682 in PRKAR2B was associated with leanness (p=0.04). We found novel effects at DRD1, where rs4867798 was associated with baseline BMI (p=0.038) and rs5326 with BMI change over 1 month ($p=2.62\times10^{-4}$). In the X-linked HTR2C gene, rs518147 demonstrated a gender specific effect, where the common allele (C) was associated with higher BMI in girls (p=0.003). Additionally, exploratory analyses revealed a potential role for ADRA2A in resting energy expenditure, MC4R in baseline lipid levels, and both FTO and BDNF in lipid and glucose changes over time. **Conclusions:** Common genetic variation in energy balance, monoamine, and growth factor candidate pathways for obesity were nominally associated with weight and metabolic profile in Mexican-American children. These results warrant replication in large, independent samples and could help target for intervention children at risk for obesity and suggest personalized treatment strategies.

871S

The analysis of *MC1R* polymorphisms can be used as a tool to predict complex phenotypes, such as skin and hair color in Brazilian population? *F. Goncalves, C. Fridman.* Legal Medicine, Ethics and Occupational Health, Medical School, University of Sao Paulo, Sao Paulo, Brazil.

Human pigmentation traits, including color variation in skin, eye, and hair, belong to the most visible and differentiating human traits. The genetic basis underlying variation in human pigmentation has been the subject of intensive research by investigators in a variety of life science communities, including forensic purposes, in which the use of these trials can help identifying missing person or guide some police investigations. The two types of melanin synthesized in well defined chemical reactions are the protective dark colored eumelanin and the sulphur containing light red-yellow pheomelanin. Eumelanin together with pheomelanin constitute the two main pigments of the skin and hair. The events leading to the melanogenesis are controlled by different genes, that we can highlight the melanocortin 1 receptor gene (*MC1R*) which encodes a protein in melanocytes responsible for melanin synthesis regulation. Polymorphisms in *MC1R*, which result in a loss of function of the receptor, are associated with increased pheomelanin production, which leads to lighter skin and hair color in European population. The aim of this study was evaluated the association between the polymorphisms in the only one exon of *MC1R* gene and hair and skin color in a sample of 401 individuals of admixed population from Brazil, intending to use the data in forensic genetics casework in several situations. No deviation in Hardy-Weinberg equilibrium was observed for all the polymorphisms analyzed. We found a strong association between the SNP rs885479 (G>A) and yellow skin color (OR:148.14; CI: 6.95-315.2 to polymorphic homozygous genotype and OR: 14.0; IC: 2.68-72.99 to heterozygous genotype). In the analyses of hair color, we found associations between the red hair color and the heterozygous genotype to the SNP rs1805007 (C>T - OR: 73.89; CI: 3.27-166.4) and rs1110400 (T>C OR: 43.18; CI: 1.83-1020.8). Our data corroborate the findings of other studies in homogeneous populations, suggesting that the analysis of *MC1R* polymorphisms can be used as a tool to access some phenotypic traits and use the obtained in attempt to help forensic investigations. We are increasing our sample and additional analyzes are ongoing to confirm these results. These results are part of a major project which the aim is to study the correlation of several pigmentation genes and skin, hair and eye color. Financial Support: FAPESP2012/02043-6, LIM40-HCFMUSP.

872M

Genetic risk variants for body mass index are associated with decreased excessive daytime sleepiness in 9,832 individuals of European ancestry from NHLBI cohorts. *J.M. Lane[1,2,3], R. Saxena[1,2,3], A.C. Bjonnes[1,2,3], B. Cade[3], S. Redline[3], F.A.J.L. Scheer[3], N. Punjab[4], D. Gottlieb[3].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Division of Sleep Medicine, Brigham and Women's Hospital and Harvard Medical School Boston, MA; 4) Division of Pulmonary and Critical Medicine, Johns Hopkins University, Baltimore, MD.

Nearly half of US adults report suffering from excessive daytime sleepiness (EDS) at least a few days a month, leading to cognitive, psychological and metabolic issues, such as obesity. We therefore asked if there is a common genetic basis for both mass index (BMI) and EDS. We evaluated the relationship between EDS and a genetic risk score for BMI comprised of genetic variants identified in genome-wide association studies. We used data from NHLBI cohorts with genotype and phenotype data in subjects of European ancestry (self-report dichotomous trait) (n=11,737). Analysis was adjusted for age, gender, and ancestry principal components. A fixed effects, inverse-variance meta-analysis was performed. Risk scores were calculated using 18 SNPs previously associated with BMI. We find increased odds of EDS in overweight (OR 1.020, p=0.041), obese (OR 1.083, p less than 0.0001), and morbidly obese (OR 1.195, p less than 0.0001) individuals of European ancestry versus normal BMI (N=11,737). We find a significant association between EDS and genetic variants in BDNF and FTO. In total 14/18 SNPs demonstrate decreased risk of EDS with BMI risk alleles, even after adjustment for BMI. A weighted BMI risk score is significantly associated with EDS (OR [95%CI] 0.96 [0.95-0.98], p=1.74E-05) and is independent of BMI (0.96 [0.94-0.97], p=5.62E-07). We also tested for effect modification by the covariates age, gender, and season and the comorbidities sleep apnea and mood disorder. The prevalence of EDS varies significantly by age (p less than 0.001), peaking at 55 years. Age significantly modifies the relationship between EDS and BMI genetic risk score (Pint=0.017). There was no significant interaction with gender, sleep apnea, mood disorder, or season of measurement. Our association results capture an effect on EDS beyond BMI. Paradoxically, BMI raising alleles individually or in aggregate are associated with a decreased odds of EDS, indicating divergent physiological roles for underlying pathways. Mechanistic understanding of BMI risk variants may provide parallel insights into excessive daytime sleepiness.

873T

Promoter Polymorphism and low Serum Levels of Mannose Binding Lectin as risk factor for Rheumatoid Arthritis in Indian population. *A. Sodhi[1], J. Singh[2], S. Singh[2], S. Arora[3], M. Kaur[1].* 1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, India; 2) Department of Molecular Biology and Biochemistry, Guru Nanak Dev University, Amritsar, India; 3) Rheumatology clinic, Amritsar, India.

Mannose Binding Lectin-2 (MBL-2) is a C-type serum lectin synthesized by the liver as an acute phase protein. MBL-2 gene is located on chromosome 10q11.2- q21and various single nucleotide polymorphism (SNPs) in this gene has been reported to be associated with infection as well as autoimmune diseases. Promoterpolymorphism in MBL-2 gene is reported to be associated withaltered serum MBL levels which are associated with increased risk of various disease conditions. Rheumatoid arthritis (RA) is an autoimmune disease characterized by chronic inflammation and destruction of articular cartilage as well as synovial hyperplasia. Prevalence of RA is 0.5% - 1.0% of the global population and is increasing dramatically.The present study aimed to investigate the frequency of the mutations at positions -550 and -221 of promoter region inMBL-2 genealong with serum MBL and hs-CRP serum levels. The study was approved by institutional ethical committee as per declaration of Helsinki and informed consent was obtained from each individual. Blood samples were collected from 202 RA patients and 200 age, gender and ethnicity matched healthy subjects. Information regarding anthropometric variables, Demographic features, disease duration, and disease activity score was also recorded. Single nucleotide polymorphism in the MBL promoter (-550nt, -221nt) were analyzed by Amplification refectory mutation system-PCR (ARMS-PCR). Functional serum MBL-2 and hs-CRP levels were analyzed using commercially available kits. Out of two SNPs studied, significant difference was observed in genotypic as well as allelic frequencies of H/L (-550nt) polymorphism ($p<0.0001$). L allele of -550 is at risk for RA patients (odd ratio: 1.079 CI 95%: 0.021-2.021; $p<0.0001$). Combination of two promoter polymorphism YXLL is risk over YXHH (odd ratio: 12.12 CI95%: 3.71-39.19; $p<0.00001$) and H allele is protective over L allele (Odd ratio: 0.132; CI 95%: 0.04-0.388; $p<0.0002$). Serum MBL-2 levels were found to be significantly ($p<0.001$) lower in RA patients as compare to healthy controls and hs-CRP serum levels were found significantly ($p<0.01$) higher in RA as compared to healthy controls.The present study showed that the polymorphism in the promoter region of the MBL gene may be a genetic marker associated with RA and presence of mutation in the -550 H/L promoter region of the MBL-2 gene correlated to low MBL serum levels.

874S

Carboxypeptidase E and dopamine transporter SNPs are associated with percent weight change in kidney transplant recipients. *A.G. Stanfill[1, 3], A.K. Cashion[2], D.K. Hathaway[3], Y.P. Conley[1].* 1) Health Promotion and Development, University of Pittsburgh, Pittsburgh, PA., USA; 2) National Institute of Nursing Research, National Institutes of Health, Bethesda, MD; 3) University of Tennessee Health Science Center, Memphis, TN.

Variations in dopaminergic pathway genes have recently gained attention for being associated with weight gain and obesity. These genes include the dopamine active transporter gene (SLC6A3 or DAT1), which codes for the transporter responsible for dopamine reuptake from the synapse, and the carboxypeptidase E (CPE) gene, which has been shown to have a regulatory effect on that transporter. While the associations are exciting, there are often logistical difficulties in the longitudinal study of the genetics of weight gain. Kidney transplant recipients make an ideal population for these types of genetic association studies, as approximately 30% of recipients gain a significant amount of weight (>10 kg) in the first year after surgery. Although many clinicians attribute this weight gain to immunosuppressant therapies, this assumption has not been supported in the literature. Furthermore, there is no known relationship between dopamine function and renal disease, although gene expression work in this population has shown a significant relationship between weight gain and dopaminergic pathway genes. The purpose of this study was to test variations in SLC6A3/DAT1 and CPE for associations with weight gain, using kidney transplant recipients as a model population. Blood samples were previously collected from 70 kidney transplant recipients (43% female, 57% African American, aged 50.7 ±13.2 years, baseline weight 181.9 ±39.4 pounds) at the time of surgery. Two SNPs in SLC6A3 (rs6347, rs6350) and three SNPs in CPE (rs1583645, rs1946816, rs34516004) were genotyped using Taqman assays. Associations with percent weight change at 12 months post-transplant were done using dose dependent and risk allele analyses. ANOVA showed a dose dependent effect for rs6347 ($p=0.05$) and rs1946816 ($p=0.019$). Risk allele chi squares showed a statistically significant effect for the A risk allele rs6347 ($p=0.0003$), and for the C risk allele of rs1946816 ($p=0.00004$). The C allele of rs1946816 also showed a negative correlation with weight change ($r=-0.22$, $p=0.05$). Multiple regression modeling supported these associations (rs6347 TT genotype $p=0.053$, rs1946816 CC $p=0.048$). While this was done as a pilot study, these results are promising for repetition with a larger sample size. If confirmed in a larger sample, the results could be translated into potential clinical applications to identify individuals most genetically at risk for weight gain post-transplant.

875M

Candidate gene association study of chronic obstructive pulmonary disease using a targeted high throughput sequencing approach. *J. Klar[1], H. Matsson[2], C. Söderhäll[2], H. Backman[3], A. Lindberg[3], E. Rönmark[3], J. Kere[2], B. Lundbäck[4], N. Dahl[1].* 1) Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 2) Department of Biosciences and Nutrition and Center for Innovative Medicine (CIMED), Karolinska Institutet, Huddinge, Sweden; 3) The OLIN studies, Sunderby Hospital of Norrbotten, Luleå, Sweden; 4) Krefting Research Centre, Institute of Medicine, University of Gothenburg, Sweden.

Background: Chronic obstructive pulmonary disease (COPD) is a common disease in Sweden, affecting approximately one in six in age over 40 years, with severe impact on health and quality of life. COPD results from environmental factors, especially cigarette smoking, with contributions from yet unknown genetic background factors and by gene-environment interactions. Not all smokers develop COPD suggesting that genetic factors modulate the life-time risk and conversely, non-smokers can be at elevated risk for COPD. The general objective of this project is to get insight into the mechanisms and interactions of specific genes and genetic pathways that contribute to the development of COPD.

Aim: To investigate candidate genes, with emphasis on genes important for lung development and homeostasis, in defined populations of smokers and non-smokers in search for genetic variations associated with development of COPD.

Methods: : We conducted a candidate gene analysis on 200 kb of enriched sequences, including 22 genes implicated in lung development and 71 genes and regions previously associated with COPD. Targeted enrichment (HaloPlex; Agilent) and high throughput sequencing (Illumina) was performed on 96 patients and 96 healthy controls retrieved from the Swedish Obstructive Lung Disease in Norrbotten (OLIN) Studies sample set.

Results: We identified a total of 2,151 SNPs of which 78, distributed in 45 gene regions, have significantly different allele frequencies in COPD cases than controls. All COPD associated SNPs showed a strong effect on the development of COPD, as indicated either by low or high odds ratios (OR). The SNPs are mainly located in genes that cluster in pathways associated with cell proliferation, including genes involved in both development and damage repair of the lung.

Conclusion: Our preliminary results confirm previous findings and, in addition, high effect sizes of SNP variants associated with COPD compared to other studies. The strong effect sizes are possibly attributed to the genetic background of our study populations. The results support the idea that variants in genes for lung development are important determinants of adult lung function that may ultimately contribute to COPD.

876T

SNP variants in MHC are associated with sarcoidosis susceptibility and subgroups - a joint case-control association study in four European populations. *A. Wennerström[1,2,3], E. Lahtela[2], V. Anttila[4,12], J. Grunewald[5], C. van Moorsel[6], M. Petrek[7], A. Eklund[5], J. Grutters[6], V. Kolek[7], L. Padyukov[10], A. Pietinalho[8], M. Ronninger[5], M. Seppänen[9], O. Selroos[11], M-L. Lokki[2].* 1) National Institute for Health and Welfare (THL) Public Health Genomics Unit, Helsinki, Finland,; 2) Transplantation Laboratory, Haartman Institute, University of Helsinki, Finland, Helsinki, Finland,; 3) University of Helsinki The Institute for Molecular Medicine Finland (FIMM) Biomedicum, Helsinki, Finland,; 4) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, United States; 5) Respiratory Medicine Unit, Department of Medicine Solna and CMM Karolinska Institutet and Karolinska University Hospital, Solna, Sweden; 6) Department of Pulmonology, St Antonius Hospital Nieuwegein, and Heart and Lung Center University Medical Center Utrecht, Utrecht, Netherlands; 7) Laboratory of Immunogenomics and Immunoproteomics, Faculty of Medicine and Dentistry, Palacky University Olomouc, Olomouc, Czech Republic,; 8) Raasepori Health Care Centre, Raasepori, Finland,; 9) Immunodeficiency Unit, Division of Infectious Diseases, Department of Medicine, Helsinki University Central Hospital, Helsinki, Helsinki, Finland Finland; 10) Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden Finland; 11) Semeco AB, Vejbystrand, Sweden; 12) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, United States.

Sarcoidosis is a multiorgan inflammatory disorder of unknown aetiology. The most probable pathophysiology of sarcoidosis, the dysregulation of the immune response strongly suggests benefits from a better understanding of the role of the immune mediating genes (e.g. MHC genes) in sarcoidosis susceptibility. We present results from a Finnish case-control discovery sample as well as three independent replication studies from the Swedish, Dutch and Czech populations. We studied four genes in the MHC Class III region (LTA, TNF, AGER, BTNL2) and HLA-DRA in relation to HLA-DRB1 alleles to detect variants predisposing to sarcoidosis and to identify genetic differences between patient subgroups. Patients with sarcoidosis (n=805) were further subdivided based on the disease activity and the presence of Löfgren syndrome. In a meta-analysis, seven SNPs were associated with non-Löfgren sarcoidosis (NL; the strongest association with rs3177928 in HLA-DRA, P=1.79E-07, OR=1.9) and eight with Löfgren syndrome (LS; the strongest association with rs3129843 in BTNL2/HLA-DRA region, P= 3.44E-12, OR=3.4) when compared with healthy controls (n=870). The high LD between SNPs and an HLA-DRB1 challenged the result interpretation. In addition to these SNPs, population-specific associations for sarcoidosis were observed. In conclusion, there is clear evidence that polymorphisms in the BTNL2 and HLA-DRA have a role insarcoidosis susceptibility. Most importantly, our study revealed sarcoidosis-related variants that were shared across ethnicities as well as ethnicity-specific genetic markers. Future functional studies are required to reveal the causal variants of these associations and the immunogenetic basis related to sarcoidosis.

877S

Assessment of *LGALS3* genetic variants rs4644, rs4652, rs2075601 and galectin-3 levels as risk factor in Rheumatoid Arthritis. *T. Kaur[1], M. Kaur[1], S. Singh[2], S. Arora[3], J. Singh[2].* 1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, India; 2) Department of Molecular Biology and Biochemistry, Guru Nanak Dev University, Amritsar; 3) Rheumatology Clinic, Amritsar.

Galectin-3 is a pro-inflammatory molecule family encoded by *LGALS3* gene on 14q21 chromosome. It can act as a key player in various inflammatory diseases by activation of macrophages and neutrophils. Rheumatoid arthritis (RA)is a chronic, inflammatory, autoimmune disease characterized by destruction of peripheral joints leading to deformity and disability. Worldwide, 1% of population issuffering from RA. Elevated levels of galectin-3 have been implicated in serum and synovial fluid of RA patients. A single study involving *LGALS3* genetic variants have indicated its association with RA. The present study has been proposed to evaluate *LGALS3* variants and its protein serum levels in RA in Indian population.The present case-control study included 200 RA patients, diagnosed according to 1987 revised criteria of American college of Rheumatology and acohort of 200 age, genderand ethnicity matched controls.The study was approved by institutional ethical committee in accordance with declaration of Helsinki and written informed consent was obtained from each participant. Genetic typing for rs4644 and rs4652 was done by PCR-RFLP method while for rs2075601 by tetra primer PCR method and by Sanger Sequencing. Serum galectin-3 and serum hs-CRP levels were assessed in both patients and controls using commercial available ELISA kits (Abcam, UK and MyBiosource, USA respectively).Suitable statistical analysis was performed using SPSS version18.0.Significant difference in allelic distribution has been observed in RA patients as compared to controls ($p<0.05$) and there was suggestive evidence of an association in a Co-dominant model (AA vs AC = AC vs CC; OR= 1.37, 95% CI 1.00-1.88, p = 0.048) for rs4644 SNP. For rs4652 SNP, present study found a dominant mode of association with RA (AC/CC vs AA; OR- 3.947, 95% CI 0.86 -19.67, p= 0.047) and AA genotype was found as a risk factor. For rs2075601 SNP, the frequency of TT genotype was found to be prevalent in RA patients (11%) than controls (7%). Different genetic combinations for three SNPs showed that heterozygosity at three loci was risk factor for our RA population.Significantly elevated serum galectin-3 levels and hs-CRP levels have been observed in RA patients than controls ($p<0.001$, $p<0.001$ respectively). In present study, association of *LGALS3* variant and its high protein serum levels with RA suggested the role of galectin-3 as a genetic marker with disease.

878M

Genetic association study between 39 genes and nonsyndromic cleft lip and cleft palate in Brazilian population. *TK. Araujo[1], R. Secolin[1], TM. Félix[2], MIB. Fontes[1,3], IL. Monlleó[4], J. Souza[5], AC. Fett-Conte[6], EM. Ribeiro[7], AC. Xavier[8], A. Rezende[9], M. Simioni[1], VL. Gil-da-Silva-Lopes[1].* 1) Department of Medical Genetics, Faculty of Medical Sciences, University of Campinas (UNICAMP), Brazil; 2) Medical Genetics Service, Hospital de Clínicas de Porto Alegre (HCPA), Brazil; 3) Medical Genetics Sector, State University of Alagoas (UNCISAL), Brazil; 4) Clinical Genetics Service, Federal University of Alagoas (UFAL), Brazil; 5) Medical Genetics Sector, Assistance Center for Cleft Lip and Palate (CAIF), Brazil; 6) Molecular Biology Department, Medicine School of Sao José do Rio Preto (FAMERP/FUNFARME), Brazil; 7) Medical Genetics Sector, Hospital Infantil Albert Sabin (HIAS), Brazil; 8) Center for Research and Rehabilitation of Lip and Palate Lesions (CRRLPL), Centrinho Prefeito Luiz Gomes, Brazil; 9) Center of Health Sciences, Department of Clinical and Toxicological Analysis, University Federal of Rio Grande do Norte (UFRN), Brazil.

The objective of this study was to evaluate genetic association of 39 candidate genes and Nonsyndromic cleft lip with cleft palate (NSCLP) in a sample from Brazilian population. It was performed a case-control association study composed of 537 individuals. The case group was compound of 182 cases with NSCLP (99 males and 83 females) enrolled in Brazilian Database on Orofacial Clefts and previously evaluated by clinical geneticist. A control group of healthy individuals with no history of orofacial cleft in three generations was included. Statistical power of the sample was evaluated using Gpower software. Two hundred fifty-three tagging single nucleotide polymorphism (tagSNPs) were genotyped by the OpenArray™TaqMan system (Applied Biosystems). The association analysis was performed by logistic regression using Plink software. A stepwise regression analysis was performed using the R program. The statistical results were corrected for multiple testing using the Bonferroni correction for each polymorphism evaluated. The sample showed 80.11% of statistical power to detect genetic association. Twenty-four SNPs spanning 16 genes were statistically associated with the etiology of NSCLP in the sample. The associated genes were: *TCBE3* rs2235541*T (p=0.032); *MSX1* rs3775261*T (p=0.009); *SPRY1* rs300566*A (p=0.017); *MSX2* rs4868442*T (p=0.020); *PRSS35* rs4706180*C (p=0.032); rs512140*T (p=0.035); *TFAP2A* rs537112*T (p=0.015); rs533558*G (p=0.001), rs303048*T (p=0.006); rs1675414*C (p=0.042); *SHH* rs1233556*T (p=0.047); *VAX1* rs10787760*G (p=0.011); rs7086344*T (p=0.000); rs6585429*T (p=0.013); *TBx10* rs3758938*G (p=0.015); *WNT11* rs7936750*G (p=0.018); *PAX9* rs1955734*T (p=0.001); *BMP4* rs17563*C (p=0.016); *JAG2* rs11621316*G (p=0.003); *KIF7* rs4932238*C (p=0.007); rs4932240*T (p=0.029); *AXIN2* rs11655966*T (p=0.030); *DVL2* rs2074222*G (p=0.002); rs222850*C (p=0.016). The stepwise regression analysis showed that SNPs associated contribute together to 15.5% of the determinants of etiology of NSCLP in this sample. In conclusion, these data suggest that these genes are associated with NSCLP in the Brazilian population. In addition, this is the first study to suggest association between *KIF7* and *TCEB3* genes and etiology of NSCLP. Supported by: FAPESP, CNPq.

879T

Exome sequencing and targeted DNA resequencing reveals association of the MYO5B SNP rs183559995 with risk of Familial Nonsyndromic Cleft Lip and Palate. *S. Beiraghi[1], H.A. Stessman[2], A.K. Mitra[2], R.J. Schaefer[3], W. Wang[3], C.L. Myers[3], B.G. Van Ness[2].* 1) Division of Pediatric Dentistry, University of Minnesota, Minneapolis, MN; 2) Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN; 3) Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN.

Non-syndromic cleft lip with or without palate (NSCL/P) is one of the most common congenital birth defects with an incidence of 1/500-1/1000. Genetic factors have been shown to play significant role in the development of NSCL/P. Using non-parametric linkage analysis study on a six-generation family (n=27) with probable autosomal dominant, low penetrance inheritance of NSCL/P, we have previously reported that a 5.7-Mb genomic region on 18q21.1 that potentially contains a pathogenic, high-risk variant for NSCL/P. In the current study, we performed exome sequencing on 6 affected individuals, 2 obligate carriers, and 4 unaffected individuals from the NSCL/P family using Illumina HiSeq with TruSeq Exome Enrichment. 100 Western European (CEPH) genomes from the 1000 Genomes Project were utilized as unaffected controls owing to low penetrance for NSCL/P within this family. All sequences were mapped to hg19 human reference genome using Burrows Wheeler Aligner 0.5.9 (BWA) and variants were called using SAMtools and GATK Unified Genotyper for all sites with greater than 8 reads. High quality variants were used as markers in genome wide association analysis linked to the affected phenotype (NSCL/P) using PLINK whole genome association analysis toolset. Four (4) candidate SNPs within the same gene, MYO5B, a myosin family member involved in protein trafficking, as well as additional variants (SNVs and INDELs) were identified that may contribute to NSCL/P disease etiology. Subsequent targeted Sanger re-sequencing in 33 family members demonstrated that the MYO5B intronic SNP rs183559995 (G/A) was significantly associated with the NSCL/P trait (p= 0.001, 95% CI=2.25-185.3, OR=15). Our results indicate that this SNP is a strong candidate gene for familial NSCL/P. Further functional studies will be required to determine its significance with regard to MYO5B structure and function.

880S

Rare Variants Within 7p Region Associated with Carotid Bifurcation Intima-Media Thickness Among Dominican Republic Families. *N.D. Dueker[1], A. Beecham[1], L. Wang[2], S. Blanton[2], C. Dong[3], S. Guo[1], D. Cabral[3], E. Sabala[2,3], T. Rundek[3], R.L. Sacco[2,3].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL; 3) Department of Neurology, Epidemiology and Public Health, Miller School of Medicine, University of Miami, Miami, FL.

Ischemic stroke (IS) and myocardial infarction (MI) are leading causes of mortality and disability in the US. Though both IS and MI are suggested to be genetically controlled, only a small proportion of the risk for both is explained. One mechanism to identify candidate loci for these vascular disorders is to identify the genetic determinants of subclinical phenotypes, such as carotid intima-media thickness (cIMT). We have previously shown cIMT measures to be heritable and found evidence for linkage and association of common variants on 7p with carotid bifurcation IMT (BIF). Therefore, we sought to further characterize the 7p region and to identify additional candidate variants, hypothesizing that rare variants in this region are associated with BIF. To test this hypothesis, we sequenced the 1 LOD unit down region on 7p in nine extended families from the Dominican Republic with strong evidence for linkage to BIF in this region (family specific LOD score > 0.1). Using this data, we performed the family-based sequence kernel association test (famSKAT) on all genes within the 7p region to identify those associated with BIF. Analyses were restricted to single nucleotide variants (SNVs) with minor allele frequency (MAF) <5%. A total of 70 genes were included in our analyses and one gene, nucleotide-binding oligomerization domain (*NOD1*), met our Bonferroni corrected threshold of p=0.0007 (famSKAT p=0.0003; # SNVs=23). NOD1 is an excellent candidate gene as it is expressed in the brain and the heart, and a previous study found the *NOD1* G796A polymorphism to be significantly more common in stroke patients seropositive for Chlamydia pneumoniae compared to controls seropositive for Chlamydia pneumoniae. In addition to *NOD1*, we also identified 10 moderately-associated genes with p<0.05, including *GHRHR* which encodes the growth hormone releasing hormone receptor, a receptor that when activated, improves injury response after MI (famSKAT p=0.02; # SNVs=24). In exploratory analyses, we restricted our analysis to exonic variants with MAF<5% and found similar results. Taken together, our study provides suggestive evidence for a role of rare variants within our previously identified 7p region in BIF. Sequencing within additional families is currently in progress.

881M

No association of PTPN22 and SUMO4 Polymorphisms with predisposition to type 1diabetes (T1D) in a cohort of south Indian subjects. *B.C. Gorijala[1], U. Ratnamala[2], S.K. Nath[3], U. Radhakrishna[4].* 1) Oncology Department, Krishna Institute of Medical Sciences, Secunderabad, Andhra Pradesh, India; 2) Department of Pharmacology, Creighton University, Omaha, NE, United States; 3) Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK, United States,; 4) Beaumont Research Institute, Royal Oak, MI, USA.

Type 1 diabetes, (T1D) (juvenile or insulin-dependent diabetes mellitus) is an autoimmune disease characterized by the lack of insulin due to an autoimmune destruction of pancreatic beta cells. The estimated incidence of T1D varies from 8-17/100,000 in Northern Europe and the U.S. The general incidence of T1D in India is 10.6 cases/year/100,000. The exact etiology and pathogenesis of T1D is still unknown, however genetic factors are believed to be a major component for the development of T1D. Additionally, modifying epigenetic factors such as diet, environmental, infections and lifestyle play an important role in disease expression. To date, several genome-wise association and candidate gene studies have identified more than 25 genetics association with high confidence, including PTPN22, SUMO4, CTLA4 and IL2RA. However no common gene mutations or pathogenic causative genomic variations have been identified. We have recruited 500 sporadic T1D patients, and an equal number of age-matched controls from southern India. We have recently screened PTPN22 and SUMO4 genetic variants in 100 T1D patients and equal controls using a SNP array. Our results did not show any association with PTPN22 (rs2476601, 1858C>T). This SNP was monomorphic (G/G) in both T1D. This is consistent with other Asian populations. The data of SUMO4 variation (rs237025, 163A > G, M55V) which has been shown to be a susceptibility variant for TD1, a higher frequency of combined AG and GG genotypes (64%) in the affecteds than in matched controls (56%). However, the “G” allele frequency between cases and controls are not statistically significant (40% in cases vs. 35% in controls, OR (C.I.) = 1.23 (0.82-1.85), chi2=1.07, p=0.30), mostlikely due to small sample sizes.

882T

Evaluation of Genetic Polymorphism of *MBL2* Gene and Pulmonary Function Test in Chronic Obstructive Pulmonary Disease. *A. Sharma[1], G. Gandhi[1], B. Malhotra[3], J. Singh[2], S. Singh[2], M. Kaur[1].* 1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, India; 2) Department of Molecular Biology and Biochemistry, Guru Nanak Dev University, Amritsar, India; 3) Department of Tuberculosis and Chest Diseases, Govt. Medical College, Amritsar, India.

Chronic Obstructive Pulmonary disease (COPD) is characterized by narrowing of airways due to abnormal inflammatory response to noxious particles leading to progressive reduction in the pulmonary function. Worldwide, COPD is the fourth leading cause of morbidity and mortality. In India, its prevalence is 4.1%. Several molecules of innate and adaptive immune system are implicated in disease etiology and pathogenesis. Mannose-binding lectin (MBL) is a pattern-recognition protein which binds to carbohydrates of infectious agents or cells resulting in activation of complement system and thus causing stimulation of inflammatory reactions. Objectives of the present study were to evaluate *MBL2* gene polymorphisms in COPD patients and its association with pulmonary function test (PFT). The study was approved by Institutional Ethical Committee. After informed consent, blood samples were collected from 80 COPD patients and 72 age, gender and ethnicity matched controls. Genomic DNA was isolated from blood samples using inorganic method. Single nucleotide polymorphisms (SNPs) of codon 54 and 57 of the *MBL2* gene were studied by PCR-RFLP using restriction enzymes BanI and MboII respectively. SNP of 5’UTR region (allele P/Q) was studied by ARMS-PCR. Pulmonary function test was performed using Medicaid system. For the diagnosis and categorization of COPD patients GOLD (2012) guidelines were used. All statistical analyses were performed using SPSS version 16.0. No significant difference was observed in genotypic and allelic distribution of codon 54, 57 and P/Q alleles in comparison with controls. Genetic combination, heterozygosity for codon 54, P/Q and homozygosity for codon 57 (GAAACT), appears to be a risk factor (OR=4.301; CI=1.06- 17.45; p=0.04) for susceptibility to develop COPD with respect to AAAACT. All the pulmonary function test parameters viz. forced vital capacity (FVC), forced expiratory volume in one second (FEV1), FEV1/FVC ratio and peak expiratory flow (PEF) were found to be significantly reduced in patients than controls (p<0.001). On the basis of disease severity, patients were stratified into mild (3.75%), moderate (23.75%), severe (40.00%) and very severe (32.50%) disease stages. Significant association was found with GGAACC for FEV1 (p<0.001) in case of severe COPD patients and GAAACC for FVC (p=0.039) in case of mild COPD patients. The results of the present study showed that *MBL2* polymorphisms may be involved in pathogenesis of COPD.

883S

Age-related hearing impairment associated with GJB2 single mutation IVS1+1G>A in the Yakut population in Eastern Siberia. *N.A. Barashkov[1,2], F.M. Teryutin[1,2], V.G. Pshennikova[1], A.V. Solovyev[2], L.A. Klarov[3], N.A. Solovyeva[1], A.A. Kozhevnikov[4], L.M. Vasilyeva[5], E.E. Fedotova[5], M.V. Pak[6], S.N. Lekhanova[7], E.V. Zakharova[8], K.E. Savvinova[2,9], N.N. Gotovtsev[2,9], G.P. Romanov[2], A.M. Rafailov[9], N.V. Luginov[3], A.N. Alexeev[10], O.L. Posukh[11,12], L.U. Dzhemileva[13], E.K. Khusnutdinova[13,14], S.A. Fedorova[1,2].*
1) Department of Molecular Genetics, Yakut Scientific Centre of Complex Medical Problems, Siberian Branch of the Russ, Yakutsk, Sakha Republic, Russian Federation; 2) Laboratory of Molecular Biology, Institute of Natural Sciences, M.K. Ammosov North-Eastern Federal University, Yakutsk, Russian Federation; 3) Department of Radiology, Republican Hospital #2 - Center of Emergency Medicine, Ministry of Public Health of the Sakha Republic, Yakutsk, Russian Federation; 4) Republican Centre of Professional Pathology, Republican Hospital #2 - Center of Emergency Medicine, Ministry of Public Health of the Sakha Republic, Yakutsk, Russian Federation; 5) Audiology-Logopaedic Center, Republican Hospital #1 - National Medical Centre, Ministry of Public Health of the Sakha Republic, Yakutsk, Russian Federation; 6) Department of Pediatric, Medical Institute, M.K. Ammosov North-Eastern Federal University, Yakutsk, Russian Federation; 7) Department of Normal and Abnormal Anatomy, Operative Surgery with Topographic Anatomy and Forensic Medicine, Medical Institute, M.K. Ammosov North-Eastern Federal University, Yakutsk, Russian Federation; 8) Institute of Foreign Philology and Regional Studies, M.K. Ammosov North-Eastern Federal University, Yakutsk, Russian Federation; 9) Institute of Natural Sciences, M.K. Ammosov North-Eastern Federal University, Yakutsk, Russian Federation; 10) Institute of Humanitarian Research and Indigenous Peoples of the North, Siberian Branch of the Russian Academy of Sciences, Yakutsk, Russian Federation; 11) Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation; 12) Novosibirsk State University, Novosibirsk, Russian Federation; 13) Department of Genomics, Institute of Biochemistry and Genetics, Ufa Scientific Centre, Russian Academy of Sciences, Ufa, Russian Federation; 14) Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russian Federation.

Age-Related Hearing Impairment (ARHI) is one of the frequent sensory disorders registered in 50% of individuals over 80 years. ARHI is a multifactorial disorder due to environmental and poor-known genetic components. In this study, we present the data on age-related hearing impairment of 48 heterozygous carriers of mutation IVS1+1G>A (GJB2 gene) and 97 subjects with GJB2 genotype wt/wt in the Republic of Sakha/Yakutia (Eastern Siberia, Russia). This subarctic territory was found as the region with the most extensive accumulation of mutation IVS1+1G>A in the world as a result of founder effect in the unique Yakut population isolate. The GJB2 gene resequencing and detailed audiological analysis in the frequency range 0.25, 0.5, 1.0, 2.0, 4.0, 8.0 kHz were performed in all examined subjects that allowed to investigate genotype-phenotype correlations between the presence of single mutation IVS1+1G>A and hearing of subjects from examined groups. We revealed the linear correlation between increase of average hearing thresholds at speech frequencies (PTA0.5,1.0,2.0,4.0 kHz) and age of individuals with GJB2 genotype IVS1+1G>A/wt (rs=0.499, p=0.006860 for males and rs=0.427, p=0.000277 for females). Moreover, the average hearing thresholds on high frequency (8.0 kHz) in individuals with genotype IVS1+1G>A/wt (both sexes) were significantly worse than in individuals with genotype wt/wt ($p<0.05$). Age of hearing loss manifestation in individuals with genotype IVS1+1G>A/wt was estimated to be ~ 40 years (rs=0.504, p=0.003). These findings demonstrate that the single IVS1+1G>A mutation (GJB2) is associated with age-related hearing impairment (ARHI) of the IVS1+1G>A carriers in the Yakuts.

884M

Enzymatic properties of the catalytic domain of mouse acidic mammalian chitinase expressed in *Escherichia coli*. *A. Kashimura, K. Okawa, M. Kimura, K. Okazaki, M. Sakaguchi, Y. Sugahara, F. Oyama.* Applied Chemistr, Faculty of Engineering, Kogakuin University, Hachioji, tokyo, Japan.

Acidic mammalian chitinase (AMCase) was discovered to perform compensatory role of chitotriosidase and was named for its acidic isoelectric point. Unlike other chitinases, which are inactive at low pH, mouse AMCase has been shown to be most active at pH 2.0 with acidic stability. AMCase has attracted considerable attention due to its increased expression under pathological conditions. Significant increases of AMCase mRNA and protein were detected in an induced asthma mouse model. Polymorphisms and haplotypes of AMCase are associated with bronchial asthma in humans. Furthermore, AMCase expression is increased by antigen-induced mouse models of allergic lung inflammation. Little is known, however, about the pathophysiological functions of AMCase in mice and humans. Mouse AMCase contains an N-terminal catalytic domain and a C-terminal chitin-binding domain (CBD). It has been generally considered that CBD recognizes chitin and catalytic domain degrade it. We expressed the catalytic domain of AMCase in Escherichia coli as a fusion protein. We found that the enzymatic properties of the recombinant catalytic domain against 4-nitrophenyl N,N'-diacetyl-β-D-chitobioside were essentially consistent with those of full-length AMCase as for the pH and temperature optima as well as pH and thermal stabilities. This recombinant catalytic domain can be used to elucidate detailed biological functions of the mouse chitinase.

885T

***Nicastrin* knockdown in keratinocytes induces expression profiles for decreased expression of cell cycle genes and increases gene expression related to type-1 interferon response.** *E.D.O. Roberson[1,2], L. Cao[1].*
1) Dept of Medicine, Washington University, St. Louis, MO; 2) Dept of Genetics, Washington University, St. Louis, MO.

Hidradenitis suppurativa (HS) is chronic, painful and disfiguring skin disease. It is characterized by the development of intensely painful, deep-seated nodules and abscesses in the axillae, inner thighs, and groin. Mild forms are nodular, but more severe disease progresses to interconnected abscesses. Nicastrin (NCSTN) is frequently mutated in familial HS, primarily via frameshifting indels. We sought to recapitulate the likely NCSTN haploinsufficiency state in cell culture models using siRNA knockdowns. We knocked down the gene expression of NCSTN in both human embryonic kidney (HEK293) and human epidermal keratinocyte (HEK001) cell lines to less than 50% of wildtype level. RNA from the resulting cells, along with luciferase siRNA control, was profiled using Illumina HT12v4 microarrays. Genes were considered differentially expressed if the false-discovery rate corrected p-value was less than 0.05 and the absolute fold-change was at least 1.50. HEK293 cells with NCSTN knockdown had 650 genes with significantly increased expression and 872 genes with significantly decreased expression. Among genes with significantly increased expression, there was enrichment for genes related to the p53 signaling pathway and caspase cleavage of cytoskeletal proteins. Among the genes with significantly decreased expression, there was enrichment of genes related to response to interferon-alpha and cholesterol biosynthesis. HEK001 cells with NCSTN knockdown demonstrated 407 genes with significantly increased and 359 genes with significantly decreased expression. Genes with significantly increased expression were enriched for type-1 interferon-mediated signaling and a response to interferon-beta. This was accompanied by an enrichment of genes with interferon regulatory motif (IRF) transcription factor motifs. Genes with significantly decreased expression were enriched for genes involved in the cell cycle, organelle fission, DNA replication and cytokinesis. The transcriptome of HS affected skin is largely unexplored. Using this cell line model, we are able to demonstrate a possible increase in caspase activity (HEK293), increased inflammatory propensity (HEK001), and decreased cell proliferation (HEK001) under NCSTN depletion. NCSTN haploinsufficiency may predispose to loss of hair follicular support cells by increased cell death in response to stress and decreased follicular cell proliferation. This study highlights the need to study affected HS tissue.

886S

Functional study of Peptidylarginine deiminase type 4 as genetic risk factor for RA. *A. Suzuki[1], Y. Kochi[1], F. Shoda[2], K. Fujio[2], E. Kanno[1], H. Matsumura[1], R. Yamada[1,3], K. Yamamoto[1,2].* 1) IMS, RIKEN, Yokohama City, Japan; 2) Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 3) Center for Genomic Medicine, Kyoto University, Kyoto, Japan.

Rheumatoid arthritis (RA) is well-known as an autoimmune disease and is a chronic inflammatory disorder characterized by the destruction of multiple joints. Many genome wide association studies were performed and multiple RA-susceptibility loci and autoimmune-susceptibility loci have been identified. These studies suggested that multiple genes and its functions were related with disease causing and development. Previouly, peptidylariginine deiminase type 4 (PADI4) was identified as a susceptibility gene for RA in a Japanese population by case-control association study (Ref 1). PADI4 is a member of the PADI gene family and converts arginine residue (peptidylarginine) to citrulline residue (peptidylcitrulline). PADI4 is highly expressed in bone marrow, macrophages, neutrophils and monocytes. Peptidylcitrulline is an interesting molecule in RA, because it is an antigen of ACPA and only PADs (translated protein from PADI genes) can provide peptidylcitrullines, via modification of protein substrates. To evaluate the importance of PADI4 gene in the progression of RA, we generated Padi4 -/- DBA1J mice. We used Padi4-/- mice to show that PAD4 is significantly affected to progress of collagen induced arthritis (CIA), well known as an RA model animal. Expression of various inflammatory cytokines and Padi genes in immune cells was detected by real-time TaqMan assay. Cytokine concentrations in sera were measured by enzyme-linked immunosorbent assay. We demonstrated that Padi4 expression was induced by CII immunization. In Padi4-/- mice, inflammatory cytokine levels were significantly decreased compared with those in wild-type mice. Interestingly, Padi2 expression was induced in immune cells of Padi4-/- mice in compensation for the defect in Padi4. 1) Suzuki, A. et al Nat. Genet.34, 395-402 (2003).

887M

Functional characterization of a TERT-CLPTM1L multi-cancer risk locus on chr5p15.33. *L. Amundadottir[1], J. Jia[1], J. Choi[1], I. Collins[1], M. Xu[1], J. Hoskins[1], T. Zhang[1], M. Zhang[1], R. Koster[1], P. PanScan Consortium[2], G. Petersen[3], J. Fang[1], K. Brown[1].* 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD; 3) Department of Health Sciences Research, Mayo Clinic, Rochester, MN.

Genome wide association studies (GWAS) have mapped multiple independent cancer risk loci (n=6) to a small region on chr5p15.33 for at least ten distinct cancers, including bladder, breast, glioma, lung, melanoma, non-melanoma skin, ovarian, pancreas, prostate, and testicular germ cell cancer. This region harbors two plausible target genes, *TERT* which encodes the catalytic subunit of telomerase reverse transcriptase which maintains chromosome ends by adding telomeres repeats, and *CLPTM1L* which encodes the cleft lip and palate transmembrane protein 1-like protein which promotes cancer cell growth, protect cells from apoptosis, and can induce abnormal cytokinesis. The most significant pancreatic cancer GWAS SNP on chr5p15.33 was rs401681 (P=3.7×10^{-7}, OR=1.19) located in the 13th intron of the *CLPTM1L* gene. Imputation and fine mapping in pancreatic cancer improved this signal by three orders of magnitude (P=1.4×10^{-10}, OR=1.30) to a set of 7 highly correlated SNPs. Electrophoretic mobility shift assays (EMSA) and Luciferase assays across these SNPs led us to a single SNP that showed allele specific effects in both assays and overlapped with prominent ENCODE marks indicating regulatory potential. Examining this SNP across pancreatic cancer, melanoma, lung cancer and testicular cancer cell lines established a consistent pattern of a specific protein binding and enhanced regulatory activity for the risk allele across 8 cancer cell lines (1-2 per cancer type). Promoter siRNA targeting analysis across the regulatory region harboring the putative functional SNP resulted in a strong inhibition of *TERT* expression (average 60%, range 45-85%) but no effect on *CLPTM1L* expression. Furthermore, allele specific siRNA targeting showed stronger inhibition of *TERT* expression from the risk as compared to the protective allele. Current work focuses on identifying the protein(s) that differentially bind this variant through proteomics, and performing eQTL analysis to correlate genotypes to expression of *TERT* and *CLPTM1L*. Our results indicate that for pancreatic cancer, the multi-cancer risk locus within CLPTM1L (tagged by rs401681) may be explained by a single SNP that confers allele specific enhancer effects and influences *TERT* gene expression. The effects were consistent with elevated levels of the *TERT* expression in carriers of the risk allele. Additionally, our data suggest that this variant may represent a risk-conferring variant at this locus for multiple cancers.

888T

Functional BDNF gene variants increase risk to moderate-severe allergic rhinitis (AR). *AK. Andiappan[1,2], JM. Quek[2], C. Schurmann[5,6], A. Teumer[5], HJ. Westra[7], T. Esko[8], L. Franke[7], JJ. Liu[3], FT. Chew[2], A. Larbi[1], DY. Wang[4], O. Rotzschke[1].* 1) Singapore Immunology Network, Singapore, Singapore; 2) Department of Biological Sciences, NUS; 3) Genome Institute of singapore, Singapore; 4) Department of Otolaryngology, NUS; 5) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; 6) Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany; 7) Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 8) Estonian Genome Center, University of Tartu, Tartu, Estonia.

Background: BDNF is a secretory protein belonging to the neurotrophin family which regulates the inflammatory cascade leading to allergic disease. Increased BDNF levels have been associated to severity of allergic phenotypes such as AR, asthma and eczema. No prior genetic study has investigated the relationship between BDNF polymorphisms and moderate-severe AR susceptibility. Objective: Aim of the study was to assess the association of genetic variants of BNDF with moderate-severe allergic rhinitis, and to determine whether this has in any functional consequences. Methods: TagSNPs spanning the BDNF gene were selected from the human HapMap CHB (Chinese) population. These BDNF tagSNPs were then tested for association with moderate-severe AR in a population of 2216 Shandong Chinese. The association was replicated in another independent Singapore Chinese population of 1239 individuals. The impact of the associated tagSNP on BDNF mRNA expression was determined using data from whole blood eQTLdata. The results were confirmed by using a luciferase reporter assay to evaluate the impact of associated SNP on BDNF expression in vitro. Results: The association analysis revealed that the tagSNP rs10767664 was significantly associated with susceptibility to moderate-severe AR in both the Shandong and Singapore Chinese populations (meta p= 0.000298 and OR= 1.242). The individuals bearing the minor AA genotype of rs10767664, associated with increased risk for moderate-severe AR. The tagSNP rs10767664 was linked to the functional coding variant rs6265. Importantly, the minor alleles of rs10767664 and rs6265 were signifcantly associated with BDNF mRNA expression on whole blood. This correlation was also validated in a large eQTL meta-analysis comprising of 5416 individuals in total. Finally we demonstrated in an in vitro model that the “GG” genotype of the functional rs6265 lead to higher BDNF expression, which could result in a higher risk to the AR phenotype. Conclusion: A common variant of the BDNF gene is associated with increased risk towards moderate-severe AR. The potentially causative "GG" genotype of rs6265 appears to augment expression of BDNF mRNA in whole blood as demonstrated in an in vitro system and thereby lead to disease through activating an inflammatory cascade.

889S

A non-coding variant near *BMP2* associated with sagittal non-syndromic craniosynostosis causes differential GFP expression in zebrafish. *C.M. Justice[1], J. Kim[2], S.D. Kim[2], G. Yagnik[2], B. Carrington[3], R. Sood[3], A.F. Wilson[2], S.A. Boyadjiev[2].* 1) Genometrics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; 2) Section of Genetics, Department of Pediatrics, University of California Davis Medical Center, Sacramento, CA; 3) Zebrafish Core, Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Craniosynostosis (CS) is a common congenital malformation in which one or more of the cranial sutures of an infant skull fuse prematurely. Approximately 80% of the CS cases are non-syndromic with unknown etiology, and approximately 50% of these involve the sagittal suture. Sagittal non-syndromic craniosynostosis (sNSC) has an estimated prevalence of about 2 per 10,000 live births. Our recent genome-wide association study of sNSC in 130 non-Hispanic White (NHW) trios identified a significant association to rs1884302, which is 346kb 3’ of *BMP2*. We did not identify any coding variant in *BMP2*, suggesting that variants in the region near *BMP2* may harbor regulatory elements responsible for the phenotype. We measured the expression of *BMP2* in primary calvarial osteoblasts derived from patients with sNSC and observed overexpression of *BMP2* in 2 out of 8 sNSC osteoblasts cell lines, as well as heightened BMP2-signaling. A 716bp fragment amplified from a proband encompassing rs1884302 was observed to enhance the promoter activity of *BMP2* in a *Renilla* luciferase assay. In order to determine if variation at this SNP causes functional changes, 716 bp fragments with the risk allele C (MAF=0.34 in NHW) and wild-type T allele at rs1884302 were cloned into a zebrafish enhancer detector (ZED) vector for *Danio rerio* transgenesis. Stable transgenic lines from 2 independent founders for each allele were observed for green fluorescent protein (GFP) expression driven by the cloned fragment. We observed strong expression of GFP in the brain region of the transgenic fish with the C (risk) allele, but not with the T (wild-type) allele of rs1884302. Our *in vitro* results suggest that the C allele at rs1884302 may, in fact, be regulating expression of *BMP2*, while our *in vivo* results indicate that the presence of the C allele at this SNP acts as an enhancer.

890M

Functional Investigation of Celiac Susceptibility Gene LPP in T Cells. *B. Molloy, M. Freeley, E. Quinn, R. McGinn, A. Long, R. McManus.* Department of Clinical Medicine, Institute of Molecular Medicine, St. James's Hospital, Dublin 8, Ireland.

Celiac Disease (CeD) is a common, complex and chronic immune-mediated disease affecting the small intestine. CD4 T cells are known to play an important role in celiac disease etiology as they initiate an immune response to gluten displayed by antigen presenting cells. A large case control study using the Immunochip identified the transcription factor /adhesion protein Lipoma-preferred partner (LPP) as the most significantly associated non-HLA risk locus with an associated p value of 10^-49 (Trynka et al, 2011). LPP has known roles in cancer and smooth muscle migration but little is known if it plays a role in T cells or how it may contribute to celiac disease pathophysiology (Grunewald et al, 2009). mRNA sequence data of CD4 T cells from our lab (data not shown) shows that LPP is expressed at higher levels in celiac samples compared to controls. We aimed to investigate the role of LPP in T cells. In this study we wanted to firstly confirm the expression of LPP in CD4 T cells and examine the effect LPP may have on cell migration through siRNA knockdown. In addition, using qPCR we tested a number of potential LPP targets that demonstrated dysregulation in our sequencing study for differences in expression in the presence or absence of LPP. We confirmed LPP expression in peripheral blood T lymphocytes. T cells knocked down for LPP showed defects in transwell migration in response to chemotactic signals. Furthermore, preliminary data shows that when stimulated with the chemokine CXCL12, knockdown of LPP is associated with alterations in the mRNA levels of the potential LPP interactors or transcriptional targets MMP25, TIMP1 and CXCR4 suggesting a possible mechanism by which LPP contributes to disease pathophysiology. Ongoing investigation including the use of flow cytometry aims to further delineate the role of LPP in T cells.

891T

Why do Genetic "Risk Factors" for Major Diseases not Always Negatively Affect Survival? *S. Ukraintseva[1], K. Arbeev[1], A. Kulminski[1], I. Akushevich[1], D. Wu[1], G. Joshi[2], I. Culminskaya[1], K. Land[1], E. Stallard[1], A. Yashin[1].* 1) Center for Population Health Aging, Duke University, Durham, NC; 2) University at Buffalo, NY.

Common complex diseases, such as cancer, CVD, diabetes and AD, are major contributors to mortality. However, genetic variants that have been associated with increased risks of such diseases are often found in genomes of long-lived people, and do not seem to compromise longevity. Here we discuss several genetic mechanisms that might plausibly explain the seemingly paradoxical situations in which genetic "risk factors" for major diseases may be neutral or even beneficial in relation to survival and longevity of their carriers. Such mechanisms include (but are not limited to): (i) trade-off-like effects of genes on risks of/mortalities from different health disorders; (ii) age-specific influence of genes on vulnerability to diseases and death; (iii) gene-gene interaction (epistasis); (iv) gene-environment interaction. We review current evidence in support of this explanation and conclude that being a genetic risk factor for major disease does not necessarily mean being the risk factor for all-cause mortality. The net effect of a disease risk allele on person's survival may be negative, neutral or positive, depending on a balance of detrimental and beneficial effects of such allele on various health and aging related traits. This balance may change with age, environmental conditions and genetic surrounding. Facilitating research on conditional effects of genes is of critical importance for better understanding the complex relationships between diseases and longevity and for advancing the area of personalized prevention.

892S

Disruption of the CTNND2 gene causes learning problems within the dyslexia spectrum. *A. Lindstrand[1,2,3], W. Hofmeister[1,2], D. Nilsson[1,2,3,4], A. Topa[5], BM. Anderlid[1,2,3], F. Darki[6], H. Matsson[7], I. Tapia Páez[7], T. Klingberg[6], L. Samuelsson[5], V. Wirta[8], F. Vezzi[9], J. Kere[7,10], M. Nordenskjöld[1,2,3], E. Syk Lundberg[1,2,3].* 1) Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 2) Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 3) Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden; 4) Science for Life Laboratory, Karolinska Institutet Science Park, Solna, Sweden; 5) Department of Clinical Genetics, Sahlgrenska University Hospital, Gothenburg, Sweden; 6) Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden; 7) Department of Biosciences and Nutrition, and Center for Innovative Medicine, Karolinska Institutet, Huddinge, Sweden; 8) SciLifeLab, School of Biotechnology, KTH Royal Institute of Technology, Stockholm, Sweden; 9) SciLifeLab, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden; 10) Molecular Neurology Research Program, University of Helsinki, and Folkhälsan Institute of Genetics, Helsinki, Finland.

Cytogenetically visible chromosomal translocations are highly informative as they can pinpoint strong effect genes even in complex genetic disorders. Here we report a mother and daughter with borderline intelligence and learning problems within the dyslexia spectrum and two apparently balanced reciprocal translocations; t(1;8)(p22;q24) and t(5;18)(p15;q11). By low coverage mate-pair whole genome sequencing we were able to pinpoint the genomic breaks to 2 kb intervals. We then located the chromosome 5p breakpoint to intron 9 of *CTNND2* by direct sequencing. An additional case with similar phenotypic presentation and a 163 kb microdeletion exclusively involving *CTNND2* was identified with genome wide array comparative genomic hybridization. This microdeletion at 5p15.2 is also present in mosaic state in the patient's mother, but absent from the healthy siblings. We investigated the effect of *CTNND2* polymorphisms on normal variability and identified a polymorphism (rs2561622) with significant effect on phonological ability and white matter volume in the left frontal lobe, close to cortical regions previously associated with phonological processing. Finally, given the potential role of CTNND2 in neuron motility, we used morpholino knockdown in zebrafish embryos to assess its effects on neuronal migration *in vivo*. Analysis of the zebrafish forebrain revealed a subpopulation of neurons misplaced between the diencephalon and telencephalon. Taken together, our human genetic and *in vivo* data suggest that defective migration of subpopulations of neuronal cells due to haploinsufficiency of CTNND2 contribute to the cognitive dysfunction in our patients.

893M

Genetic and phenotypic correlations between surrogate measures of insulin release obtained from oral glucose tolerance test data. *A.P. Gjesing[1], R. Ribel-Madsen[1], M.N. Harder[1], H. Eiberg[2], N. Grarup[1], C.T. Ekstrøm[3], O. Pedersen[1], T. Hansen[1]*. 1) The Novo Nordisk Foundation Center for Basic Meta, University of Copenhagen, Copenhagen, Denmark; 2) Department of Cellular and Molecular Medicine, Faculty of Health and Medical Sciences, University of Copenhagen; 3) Department of Biostatistics, Faculty of Health and Medical Sciences, University of Copenhagen.

Aim: Decreased beta-cell function is a key player in development of type 2 diabetes. Large hypothesis-free studies aiming to identify new gene variants that explain the heritable component of beta-cell function apply various surrogate measures of insulin release. We examined to which extent common surrogate measures of insulin release have shared genetic causes. Methods: Genetic and phenotypic correlations were calculated in a family-cohort (n=292) in which beta-cell indices were estimated based on fasting and oral glucose-stimulated plasma glucose and serum insulin levels. Furthermore, we genotyped by the Metabochip a large population-based cohort (n=6269) for common genetic variants known to associate with type 2 diabetes, fasting plasma glucose levels or fasting serum insulin levels to examine their association with various indices. Results: The phenotypic and genetic correlations differed noteworthy for the traits compared, emphasizing that the phenotypic correlation is an insufficient measure of the magnitude of shared genetic impact. Also, we found that corrected insulin response, insulinogenic index, and area under the curve for insulin after an oral glucose challenge shared the majority of their genetic background with genetic correlation between 0.81 and 0.99. The BIGTT index for acute insulin release differed genetically from the latter traits with genetic correlations between 0.51 and 0.81 due to less influence by incretin- related genes and more influence by insulin sensitivity-related genes on the BIGTT index. The homeostasis model assessment for beta-cell function was genetically closely related to fasting insulin with a genetic correlation of 0.90. By examining variants known to associate with type 2 diabetes, fasting plasma glucose levels or fasting serum insulin levels in a large population-based study sample, it was evident that traits displaying a high level of genetic correlation tended to share a larger number of associated SNPs. We also identified a few SNPs that associated exclusively with only one of the measures of insulin secretion and the biological effect of these SNPs gave clues to understand how the indexes for insulin secretion could reflect different physiological mechanisms. Conclusion: The level of shared genetic background varies between surrogate measures of insulin release, and this should be considered when designing a genetic association study to best obtain information on different mechanisms of insulin release.

894T

Potential Transcriptional Mediators for Established Type 2 Diabetes Variants in Southwestern American Indians. *R.L. Hanson[1], L.J. Baier[1], S. Kobes[1], R.G. Nelson[1], E.J. Weil[1], A. Nair[1], Y.L. Muller[1], M. Traurig[1], S. Kumar[2], H.H. Göring[2], J. Blangero[2], J.E. Curran[2]*. 1) Diabetes Epidemiology Clin Res, NIDDK, Phoenix, AZ; 2) Texas Biomedical Research Institute, San Antonio, TX.

A number of single nucleotide polymorphisms (SNPs) reproducibly associated with type 2 diabetes mellitus (T2DM) have been identified, but the molecular mechanisms by which these variants influence susceptibility to T2DM remain largely unknown. We analyzed potential mediation between T2DM variants and gene transcription in peripheral blood. Participants included 1416 American Indians (23% with T2DM), from urban Phoenix, Arizona, in whom transciptomic measurements had been made on the Illumina HumanHT-12 v4 Expression Beadchip. Genotypes were generated for 44 established T2DM-susceptibility SNPs, 42 ancestry informative markers, used to control for admixture, and 42 “random” SNPs, which were used for genomic control. In these samples 7 of the T2DM markers had nominally-significant ($p<0.05$) associations consistent with the established direction, including SNPs in *GCK*, *CDC123*, *GRB14*, *SLC16A11*, *FTO* and 2 SNPs in *KCNQ*1. We analyzed association of these 7 SNPs with all 15,854 unique transcripts that were significantly expressed and that, according to REMOAT, did not contain SNPs. This resulted in 85 transcripts that had suggestive association ($p<0.005$) with at least one of the 7 SNPs, and these were tested for potential mediation of the relationship between the SNP and T2DM using the Sobel test. Nominally significant mediation was identified for 4 transcripts, involving 3 SNPs (in *SLC16A11*, *FTO* and *CDC123*). The strongest results were for the *SLC16A11* SNP rs75493593 and the *RNASEK* transcript, located 33 kb away. This SNP was associated with T2DM (odds ratio=1.36 per copy of the risk allele, $p=0.0017$) and the risk allele was associated with lower expression of *RNASEK* (by 0.17 SD per copy, $p=1.6\times10^{-5}$). Lower expression of *RNASEK* was associated with T2DM (odds ratio=1.22 per SD decrease in expression, $p=0.0062$, controlled for, age, sex, heritage and the SNP). Mediation analysis was consistent with *RNASEK* expression as a partial mediator of the SNP effect ($p=0.012$). Additional potential mediators included rs75493593 and expression of *ADAM15* ($p=0.021$), rs8050136 (*FTO*) and expression of *GTF2H3* ($p=0.041$), and rs10906115 (*CDC123*) and expression of *PRO0641* ($p=0.048$). These analyses identify potential transcriptional mediators of the relationship between T2DM-associated variants and the risk of type 2 diabetes, but the results need to be confirmed in additional populations.

895S

A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *I. Moltke[1,2], N. Grarup[3], M.E. Jørgensen[4], P. Bjerregaard[5], J.T. Treebak[3], M. Fumagalli[6], T.S. Korneliussen[7], M.A. Andersen[3], T.S. Nielsen[3], N.T. Krarup[3], A.P. Gjesing[3], J.R. Zierath[3], A. Linneberg[8], X. Wu[9], G. Sun[9], X. Jin[9], J. Al-Aama[9], J. Wang[9], K. Borch-Johnsen[10], O. Pedersen[3], R. Nielsen[6], A. Albrechtsen[2], T. Hansen[3].* 1) Department of Human Genetics, University of Chicago, Chicago, IL, USA; 2) The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark; 3) The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 4) Steno Diabetes Center, Gentofte, Denmark; 5) National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark; 6) Department of Integrative Biology, University of California, Berkeley, CA, USA; 7) Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark; 8) Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark; 9) BGI-Shenzhen, Shenzhen, China; 10) Holbæk Hospital, Holbæk, Denmark.

The Greenlandic population is a historically small and isolated founder population, which has experienced a dramatic increase in type 2 diabetes (T2D) prevalence. Motivated by this, we performed genetic association mapping of four T2D-related quantitative traits in up to 2757 Greenlandic individuals without clinically diagnosed T2D. Using a linear mixed model to test for association while controlling for false positives due to admixture and relatedness, we analyzed Illumina MetaboChip data and exome sequencing data. Our analyses led to the discovery of a nonsense variant in the gene *TBC1D4* with an allele frequency of 17% in the Greenlandic population. Under a recessive model, homozygous carriers of this variant have markedly higher levels of plasma glucose (β=3.8 mmol/L, P=2.5×10^{-35}) and serum insulin (β=165 pmol/L, P=1.5×10^{-20}) two hours after an oral glucose load compared to individuals with other genotypes. Furthermore, they have marginally lower plasma glucose (β=-0.18 mmol/L, P=1.1×10^{-6}) and serum insulin (β=-8.3 pmol/L, P=0.0014) at fasting and markedly higher T2D risk (OR 10.3, P=1.6×10^{-24}). Heterozygous carriers of the variant have a significant, but moderate, increase in 2-hour plasma glucose (β=0.43 mmol/L, P=5.3×10^{-5}). These findings were all replicated in up to 1064 individuals from a different Greenlandic cohort. The nonsense variant is located in an exon that is exclusive to a long isoform of TBC1D4, which is mainly expressed in skeletal muscle. Analyses of muscle biopsies showed decreasing mRNA and protein abundance of this isoform with increasing number of copies of the variant. A similar decrease was observed in protein abundance of the glucose transporter GLUT4. Since TBC1D4 is known to be a mediator of insulin-stimulated glucose uptake in cells through regulation of GLUT4 mobilization, these results suggest that the variant, which leads to premature termination of the long TBC1D4 isoform, causes insulin resistance in skeletal muscle and thereby increased risk of developing T2D. The observed effect sizes are several times larger than any previous findings in large-scale genome-wide association studies of these traits and the identified variant accounts for more than 10% of all cases of T2D in Greenland. This finding provides new insights into T2D and constitutes further proof of the value of conducting genetic association studies outside the traditional setting of large homogeneous populations.

896M

The Type 1 Diabetes Susceptibility Gene CLEC16A encodes protein which restrains NK Cells function. *R. Pandey[1], M. Bakay[1], S. Yoeun[1], J. Kushner[2], H. Hakonarson[1].* 1) Centre for Applied Genomics, Children's Hospital Of Philadelphia, Philadelphia, PA; 2) Department of Pediatric Medicine, Endocrine-Metabolism, Texas Children's Hospital, Houstan, TX.

Type 1 Diabetes (T1D) is a multi-factorial childhood disease with a strong genetic component. Several GWAS had showed association of CLEC16A with T1D, which encodes a protein of unknown function. The highest levels of CLEC16A expression were identified in NK cells, which are required for development of T1D. Therefore, defining the role of CLEC16A in NK cells will provide insight into the pathogenesis of T1D. We reported previously that protective alleles of CLEC16A [A/A] are associated with higher levels of mRNA. Thus, we hypothesize that CLEC16A functions in NK cells to restrain secretory functions including cytokine release and cytotoxicity. First we investigated the expression of CLEC16A in human immune cells and non-immune tissues on mRNA level by RT-PCR and on protein level by western blot. To address the role of CLEC16A in NK cells we studied consequences of knockdown in NK cell lines and ex-vivo NK cells and over-expression of this protein in NK cell line. Using a retroviral expression system, we created NK cell lines stably over-expressing the canonical form of CLEC16A with a GFP expression reporter. Expression was validated by Western blot analysis. Cytotoxicity and INF-γ production were decreased in NK cells stably overexpressing CLEC16A. Optimized protocols using CLEC16A siRNA mediated knockdown enabled a 70% reduction in Clec16A protein levels in NK cells and 35% increase in cytotoxicity compared to cells receiving control siRNA. We performed conjugation assay to rule out decreased target cell killing. CLEC16A over-expresing NK cells formed smaller number of conjugates for all time points with no difference in CD107a expression. Subcellular localization studies revealed cytosolic localization. CLEC16A knockdown in mice resulted in increase in NK cell cytotoxicity of the YAC-1 targets in comparison to control. Taken together, our results indicate that CLEC16A serve a role in restraining two major functions of NK cells, cytotoxicity and cytokine release. Studies are in progress to validate the mechanism. Our improved understanding of this novel TD1-linked gene and the protein it codes will likely suggest new therapeutic interventions in T1D.

897T

Understanding genetic interactions underlying type I diabetes based on chromatin interactions and across-pathway interactions. *MK. Sung, H. Bang, WJ. Yang, KS. Lee, KB. Lee, K. Kim, K. Jang, JK. Choi, First and second author have equal contribution.* KAIST, Daejeon, South Korea.

Genome-wide association studies (GWAS) prove the highly polygenic architecture of complex diseases or traits; single-locus-based methods are usually unable to detect all involved loci especially when individual loci exert little effects. Moreover, the majority of associated SNPs reside in non-coding regions, making it difficult to understand their phenotypic contribution. In this work, we studied epistatic interactions associated with seven common disease phenotypes using the data from the Wellcome Trust Case Control Consortium (WTCCC). For a systematic annotation of non-coding SNPs, we employed whole-genome multiple-cell-type enhancer data by Maurano et al. and Andersson et al. Especially, Maurano et al. identified distal enhancer-to-promoter connections using DNase I profiles across 349 diverse cell types and Andersson et al. identified extensive active enhancers and their target promoters by using CAGE data based on FANTOM5 panel of ~ 400 distinct cell types. We performed functional enrichment analysis using affected gene list. Two T1D related pathways are of particular interest because they appeared in numerous epistatic interactions within as well as between pathways, which are NK cell mediated cytotoxicity (FDR 2.2e-03) and antigen processing & presentation pathway (FDR 0). T1D is considered to be primarily a T-cell mediated disease and the strongest genetic association identified so far is that with HLA class II genes, which interact with T cell. However, β-cell destruction is influenced by the crosstalk between immune system such as NK cells, B lymphocytes and APC. In addition, several studies have shown that NK cells are involved both in T1D progression and prevention. We found a number of epistatic pairs with significant interaction effect, which act as bridge to connect two pathways. For example, 28 pairs involve SNP rs115340904 (chr6: 31245821) which are known to be previously linked to T1D (p-value 8.6e-93) and is located in DHS correlated enhancer (in LD with r2 >0.8) with putative promoter regions of MICA, MICB, TNF, NCR3, HLA-B and HLA-C, participating in NK cell function. The paired SNPs of rs115340904 are located around MHC II and III regions (chr6: 31296226 - 32740411). Our results illustrate how extensive non-coding variant annotation based on the chromatin interactions of distal enhancers can be used to dissect complex genetic interactions across multiple pathways in complex diseases.

898S

DIO2 rSNPs, transcriptional factor binding sites and disease. *N. Buroker.* Pediatrics, 356320, University of Washington, 1959 Pacific Ave NE, Seattle, WA.

The TIO2 gene transcribes deiodinase type 2 that converts the thyroid prohormone, thyroxine (T4), to the biologically active triiodothyronine (T3). The thyroid hormone T3 plays an important role in the regulation of energy balance and glucose metabolism. Regulatory SNPs (rSNPs) in the promoter region novel SNP (-2035bp), 5'UTR (rs12885300), intron one (rs225010, 225011 and rs225012), exon two [rs225014 (Thr92Ala)] and 3' UTR (rs6574549 and rs225015) of the DIO2 gene are in linkage disequilibrium. These rSNP alleles alter the DNA landscape for potential transcriptional factors (TFs) to attach resulting in changes in transcription factor binding sites (TFBS). The alleles of each rSNP were found to produce unique TFBS resulting in potential changes in TF DIO2 regulation. These regulatory changes are discussed with respect to disease and sickness.

899M

Harnessing genome engineering to characterize the role of STRs in gene regulation. *D. Zielinski[1], M. Gymrek[1,2], Y. Erlich[1].* 1) Whitehead Institute for Biomedical Research, Cambridge, MA; 2) Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, MA.

Most variants associated with complex traits reside in noncoding DNA, suggesting an important role for cis-regulatory elements. Efforts to discover cis eQTLs (expression Quantitative Trait Loci) have mainly focused on the contribution of SNPs to gene expression. However, several candidate gene studies in human and model organisms suggest that Short Tandem Repeat (STR) variations can modulate expression levels and splicing of nearby transcripts. After identifying significant associations between STR variations and expression profiles across hundreds of samples from the 1000 Genomes Project, we are experimentally validating the effects of these candidate STRs on expression levels in human lymphoblastoid cell lines. Specifically, we are using the CRISPR/Cas genome engineering system to generate a distribution of STR alleles and measuring the associated gene expression levels. In addition to validating our in silico findings, this technique will allow us to fine map individual eQTL signals and determine whether the STR is the true causal signal at these candidate loci.

900T

Functional genomics of the costimulatory locus in autoimmune disease. *L. Petukhova[1, 2], T. Yamany[2], L. Bian[2], Z. Dai[2], E.N. Drill[2,5], J.D. Broadbent[4], P.L. Nagy[4], R. Clynes[2,6], A.M. Christiano[2,3].* 1) Department of Epidemiology, Columbia University, New York, NY; 2) Department of Dermatology, Columbia University, New York, NY; 3) Department of Genetics and Development, Columbia University, New York, NY; 4) Department of Pathology and Cell Biology, Columbia University, New York, NY; 5) Department of Biostatistics, Columbia University, New York, NY; 6) Department of Medicine, Columbia University, New York, NY.

The costimulatory locus contains three key immunoregulatory genes (CD28, CTLA4, and ICOS), and was one of the first genomic regions associated with autoimmune disease outside of the HLA locus. GWAS have provided robust agnostic evidence for association of this region with type 1 diabetes, rheumatoid arthritis, celiac disease, Graves disease and alopecia areata (AA). Despite the extensive evidence for conferring risk, the mechanisms by which genetic variants contribute to disease have remained elusive. Furthermore, the efficacy of therapies that target the costimulatory pathway demonstrates that it is a critical axis of autoimmunity for some patients. In order to systematically identify all disease risk variants at this locus, we performed targeted resequencing of the costimulatory locus in 122 AA patients from our GWAS cohort, targeting the entire 297Kb region. This experiment identified 1209 variants that passed rigorous QC filters, which we computationally phased, allowing us to assign alleles to chromosomes. Among the 244 sequenced chromosomes, we identified 88 chromosomes that carried GWAS-identified risk haplotypes. We next identified 208 variants that were significantly enriched on the chromosomes carrying GWAS risk haplotypes ($p<4.1e-5$). Among these enriched SNPs, there is one CTLA4 protein coding variant (rs231775; p. T17A), and eight SNPs annotated with regulatory functions in CD4 cells in public databases. Importantly, we validated regulatory effects by demonstrating that four of these regulatory variants affect the distribution of CTLA4 isoforms in activated human T-cells and show that this shift inhibits T cell cytokine production, and TCR-mediated MAPK activation in CTLA4 deficient Jurkat cells. The identification of risk variants with biological consequences perturbing the costimulatory axis provides a crucial step forward in understanding how this locus contributes to autoimmune disease.

901S

Allele specific chromatin interaction of 9p21 endometriosis risk locus regulates expression of *ANRIL*. *H. Nakaoka, A. Gurumurthy, T. Hayano, K. Hosomichi, I. Inoue.* Division of Human Genetics, National Institute of Genetics, Mishima, Shizuoka, Japan.

With the advent of genome-wide association studies (GWASs), a large number of SNPs associated with human complex diseases and phenotypes have been discovered. Most of the identified SNPs are located on intron and intergenic regions rather than coding regions, implying that the SNPs are associated with the disease risk through the regulation of expression levels of nearby genes. Although recent studies from ENCODE project demonstrated that a large proportion of the SNPs identified by GWASs were located on open chromatin regions or DNase I hypersensitive sites (DHSs), regulatory mechanisms underlying these genotype-phenotype relationships remain to be completely elucidated. In this study, we explored the regulatory mechanism of a SNP on chromosome 9p21 that was associated with the risk of endometriosis by using functional genomics approaches.

In order to fine-map the 9p21 endometriosis risk locus, we performed a target re-sequencing of 9p21 region for 48 Japanese individuals and constructed a linkage disequilibrium (LD) map of this region. Combined with DNase-seq data from ENCODE project, we identified three candidate SNPs that were in high LD with the GWAS hit SNP and located on DHSs. Since these candidate SNPs were distant from 9p21 genes, we hypothesized that these SNP were on regulatory elements with enhancer activities through long-range chromatin interactions. Therefore, we performed chromatin conformation capture (3C) analysis followed by next generation sequencing. We observed that one of the candidate SNPs showed chromatin interaction with the promoter region of *ANRIL* (or *CDKN2B-AS1*) in an allele specific manner. By using bioinformatics analysis and chromatin immunoprecipitation (ChIP) assay, we found that the SNP disrupted a DNA binding motif of a transcription factor that was involved in Wnt signaling pathway. Finally, we demonstrated that the SNP was a cis-acting expression quantitative trait nucleotide of *ANRIL* by using allele specific expression analysis of RNA-sequencing data for normal endometrial tissues and endometrial carcinoma cell lines. These results may elucidate the regulatory mechanism underlying 9p21 endometriosis risk locus.

902M

A Genomewide Association Study of Alcohol Dependence in the Irish Affected Sib Pair Study of Alcohol Dependence. *B.P. Riley[1, 2, 3, 4], A.E. Adkins[1, 2, 3], L.M. Hack[1, 2, 3], T.B. Bigdeli[1, 2, 3], M.S. Grotewiel[1, 4], A.G. Davies[1, 5], J.C. Bettinger[1,5], C.A. Prescott[6], M. Mamdani[1, 2, 3], V. Williamson[1, 2, 3], D.M. Dick[1, 2, 3, 4, 7], V.I. Vladimirov[1, 2, 3], B.T. Webb[1, 2, 3, 4], K.S. Kendler[1, 2, 3, 4].* 1) VCU Alcohol Research Center, Virginia Commonwealth Univ, Richmond, VA; 2) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth Univ, Richmond, VA; 3) Dept of Psychiatry, Virginia Commonwealth Univ, Richmond, VA; 4) Dept of Human & Molecular Genetics, Virginia Commonwealth Univ, Richmond, VA; 5) Dept of Pharmacology and Toxicology, Virginia Commonwealth Univ, Richmond, VA; 6) Dept of Psychology, University of Southern California, Los Angeles, CA; 7) Dept. of Psychology, Virginia Commonwealth Univ, Richmond, VA.

Background: We report results from a genomewide association study (GWAS) in an ethnically homogeneous Irish sample (N=706 related cases, 1755 population controls) with strong supporting evidence from VCU Alcohol Research Center (VCU ARC) model organism (MO) studies. Methods: GWAS cases from the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) were diagnosed using DSM-IV criteria. Affymetrix V6.0 arrays were genotyped at 3 separate core facilities and BeagleCall was used to call genotypes. IMPUTE2 and the 1000 Genomes reference haplotype panel (March 2012 freeze) were used to impute unmeasured genotypes. After QC filtering, imputation, and post-imputation filtering, 710 AD cases, 1755 controls and 8.2 million SNPs remained. Probabilities were converted to dosages with MACH2. Case/control association analysis was run using MQLS to correct for the non-independence of siblings. A sex weighted prevalence estimate of 8.9% was used for controls. We used a significance threshold of p<3.06E-8 based on the number of independent LD blocks in 1000 genomes data. FDR q-values were calculated with QVALUE in R. Results: SNPs in the COL6A3 gene on chromosome 2 and an intergenic region of chromosome 3 near the ECT2 gene were genomewide significant. SNPs in 8 independent genomic regions had FDR q-values <1% (notably in KLF12, POMT2/TMED8, LOC339975 and RYR3). Knockdown of a COL6A3 ortholog by RNAi in C. elegans leads to significant reductions in the development of acute functional tolerance (AFT). Knockdown of POMT2 by RNAi in D. melanogaster leads to significant reduction in ethanol sensitivity. The C. elegans klf-3 (orthologous to human KLF12) mutant does not develop acute functional tolerance and RNAi knockdown of the D. melanogaster homolog of KLF12, luna, results in enhanced sensitivity to ethanol. Finally, a loss of function allele of unc-68, the C. elegans homolog of RYR3, confers resistance to ethanol. Human replication data support LOC339975, a long non-coding RNA predicted to regulate 12 microRNAs that in turn regulate genes including numerous genes previously implicated in alcohol phenotypes such as SLC6A3, GABRG1, CNR1, ESR1, ACSL4 and multiple alcohol aldehyde dehydrogenase genes. Discussion: Our case-control GWAS detected numerous associated loci supported by functional assays in model organisms or directly linked to implicated genes in humans, a powerful combined approach elucidate genetic risk factors underlying alcohol dependence.

903T

eMERGE Phenome-Wide Association Study (PheWAS) Identifies Clinical Associations and Pleiotropy for Functional Variants. *A. Verma[1], S. Verma[1], S. Pendergrass[1], D. Crawford[2], D. Crosslin[4], H. Kuivaniemi[3], W. Bush[2], Y. Bradford[5], I. Kullo[7], S. Bielinski[7], R. Li[8], J. Denny[5], P. Peissig[6], S. Hebbring[6], E. Pugh[9], M. Andrade[7], M. Ritchie[1], G. Tromp[2].* 1) Center for Systems Genomics, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA; 2) Case Western University, Cleveland, OH; 3) Geisinger Health System, Danville, PA; 4) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 5) Vanderbilt University, Nashville, TN; 6) Marshfield Clinic, Marshfield, WI; 7) Mayo Clinic, Rochester, MN; 8) National Human Genome Research Institute, Bethesda, MD; 9) John Hopkins University, Baltimore, MD.

We performed a phenome-wide association study (PheWAS) exploring the association between stop-gained genetic variants and a comprehensive group of phenotypes to identify novel associations and potential pleiotropy. Using multiple bioinformatics tools we selected 38 functionally relevant stop-gained/null genetic variants within the genotypic data of 37,972 unrelated patients from seven study sites in the Electronic Medical Records and Genomics (eMERGE) Network. We calculated comprehensive associations between these variants and case-control status for 3,518 ICD-9 diagnosis codes (requiring ≥ 3 visits per individual to identify case status, ≥10 case subjects per ICD-9 code). Associations were adjusted for age, sex, site, platform and the first 3 principal components. A total of 418 associations passed a liberal significance threshold of p< 0.01. The most significant association was between GLG1 rs9445 and "chronic non-alcoholic liver disease" ($p=4.12\times10^{-5}$, $\beta=2.60$). We identified many potentially pleiotropic associations at p< 0.01, 35 out of 38 SNPs demonstrated associations with more than one phenotype, and 17 SNPs were each associated with > 10 different ICD-9 codes. For example, we found associations for IL34 rs4985556 with 25 diagnoses, such as "lupus erythematosus" ($p=5.94\times10^{-3}$, $\beta=0.98$) and for GBE1 rs2229519 with 33 diagnoses, such as "hypertension" ($p=1.2\times10^{-3}$, $\beta=0.067$), "hyperlipidemia" ($p=6.66\times10^{-3}$, $\beta=0.058$), and "ocular hypertension" (2.49×10^{-3}, $\beta=0.21$). We will seek replication of these results. In conclusion, our PheWAS shows stop-gained variants may have important pleiotropic effects, and that PheWAS are a powerful strategy to mine the full potential of the EMR for genome-phenome associations.

904S

Lupus associated coding polymorphism rs1143679 within ITGAM acts in both nucleotide and protein level to develop disease phenotypes. *A.K. Maiti[1], K. Bhattarai[1], P. Motghare[1], X. Kim-Howard[1], J-P. Anaya[2], S.K. Nath[1].* 1) Gen Epidemiology Unit, A & CI, OMRF, Oklahoma City, OK; 2) Universidad del Rosario-Corporación para Investigaciones Biológicas, Bogotá, Colombia.

Integrin alpha M (ITGAM; CD11b) is a component of the macrophage-1 antigen complex, which mediates leukocyte adhesion, migration and phagocytosis as part of the immune system. We previously identified a missense polymorphism, rs1143679 (R77H) at the exon 3, that is strongly associated with systemic lupus erythematosus (SLE). We explored the molecular mechanism by which the rs1143679 risk allele in ITGAM contribute to the development of SLE disease manifestations. We show that rs1143679 carrying DNA sequences act as strong transcriptional enhancer and the risk allele affects the expression of ITGAM gene in vivo. This SNP binds with NFKB1, EBF1 and Ku70/80 protein in vitro and in vivo in the monocytes of SLE patients and healthy individuals but with reduced efficiency in the presence of risk allele. The observed reduced ITGAM expression in risk allele carrying individual's monocytes is attributed to reduced interaction of these proteins with RNA pol II at the ITGAM promoter implicating reduced transcription of this gene resulting in less mRNA and protein production. KU70/80 is a lupus auto-antigen and apart from its DNA repair function, also acts as transcription factor and binds with several enhancer SNPs that are associated with lupus. Using NextGen ChIP-seq, we mapped Ku70/80 binding genomic regions that could indicate lupus associated genomic regions. Our genome-wide analysis reveal that genes that are regulated by Ku70/80 interacts with numerous previously identified lupus associated genes, implicating the disruption of these pathways in presence of Ku70/80 autoantibody could play critical role in SLE pathogenesis. At the protein level, we also show that the mutant ITGAM (CD11b) protein for this SNP inefficiently interacts with one of its ligand protein, vitronectin,that plays critical role in glomerulonephritis. Thus, risk allele carrying sequences affects in both nucleotide and protein levels due to reduced transcription of ITGAM gene resulting in reduced protein production as well as mutant protein production that could explain the strong association of this SNP with lupus.

905M

Host genetic variation and Kaposi's Sarcoma-associated herpesvirus infection. *N. Sallah, C. Franklin, A. Palser, P. Kellam, I. Barroso.* Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom.

Kaposi's sarcoma-associated herpesvirus (KSHV) is an oncogenic gammaherpesvirus associated with a variety of lymphoproliferative diseases particularly Kaposi's sarcoma (KS), Primary effusion lymphoma (PEL) and Multicentric Castlemans disease (MCD). The virus establishes a lifelong infection with a lytic phase in which active virus replication facilitates spread and pathogenesis; and an immunologically silent latent phase that promotes persistence in B-cells. Despite high seroprevalence, only a small proportion of infected people develop tumours, in addition, striking geographic distribution and familial clustering of disease is suggestive of a possible genetic predisposition to disease. Host and virus genetic variation and their influence on KSHV infection and epidemiology remain largely uncharacterized. Here we establish a pipeline to investigate how the genetics of host-virus interactions influence KSHV pathogenesis. Studies have identified a B-cell transcription factor, X-box binding protein-1 (*XBP-1*) as a key regulator of the lytic switch that links terminal differentiation, B-cell receptor activation and virus production. Thus far, we have identified 40 genetic variants of *XBP-1* in whole genome and exome datasets and are currently using PEL cell lines to investigate the biological effects of polymorphisms found in known functional domains. The pipeline used here from identification and annotation of variants to assaying for phenotypic differences can be employed for other host genes important in viral pathogenesis to elucidate function.

906T

Studying the effects of pubertal timing-associated gene *LIN28B* on early vertebrate development using zebrafish as a model. *J.T. Leinonen[1], Y.C. Chen[2], H. Koivula[2], P. Panula[2], E. Widén[1].* 1) Institute for Molecular Medicine Finland, FIMM, University of Helsinki, Finland; 2) Neuroscience Center and Institute of Biomedicine, Anatomy, University of Helsinki, Finland.

Sequence variation near *LIN28B* has repeatedly been associated with pubertal timing in several genome-wide association studies, and animal models suggest that the evolutionarily conserved *LIN28* genes encoding for RNA-binding proteins are crucial regulators of development and growth, expressed already during fetal life. Normal puberty requires appropriate fetal development. The mechanism linking *LIN28B* to puberty is unknown, but we hypothesized it may influence early developmental processes, in particular the formation of the hypothalamic pituitary gonadal -axis. Our specific aim is to elucidate the function of the puberty-associated gene *LIN28B* by assessing 1) its embryonic temporal and spatial expression patterns, 2) the impact of its downregulation and, 3) the impact of its upregulation in developing zebrafish, a model system allowing efficient evaluation of early development and growth. Knockdown- and overexpression-experiments showed that the correct function of *LIN28B* is essential for proper embryonic development. Both *LIN28B* knockdown, as induced by morpholino oligonucleotide injection, and overexpression, induced by mRNA injections, caused a dose-dependent variety of phenotypes, with high doses being lethal and lower doses causing severe to mild malformations. The brain morphology associating with *LIN28B* downregulation was severely compromised. Co-injection of morpholino and mRNA resulted in phenotypic rescue, suggesting that the knockdown phenotypes are specific to *LIN28B* downregulation. The spatiotemporal expression patterns of *LIN28B* in wild-type zebrafish, assessed by RNA in situ hybridization on whole mount embryos and tissue sections, indicated that while the expression is not spatially restricted during the first 24 hours of development, it subsequently appears most prominent in developing organ primordia, including regions in the telencephalon. Interestingly, the timing of the formation and migration of gonadotrophin releasing hormone (GnRH) neurons to the hypothalamus, which is essential for normal sexual development in vertebrates, coincide with *LIN28B* expression. We therefore intend to investigate whether *LIN28B* and *GnRH3*-expression may be co-localized. Taken together, the experiments indicate that changes in *LIN28B* expression have far-reaching consequences for the development of vertebrate body plan, and we further plan to study the impact of *LIN28B* on the development of the GnRH neuronal network.

907S

Microglia deficiency in TREM2 R47H carriers with familial late onset Alzheimer disease. *E. Korvatska[1], S. Jayadev[2], P. McMillan[1,3], J.B. Leverenz[4], G.A. Garden[2], T.D. Bird[2,3,5], W.H. Raskind[1,5,6].* 1) Psychiatry, University of Washington, Seattle, WA; 2) Neurology, University of Washington, Seattle, WA; 3) GRECC, VA Puget Sound Health Care System, Seattle, WA; 4) Pathology Anatomy and Cell Biology7, Thomas Jefferson University, Philadelphia, Pen; 5) Medicine, University of Washington, Seattle, WA; 6) MIRECC, VA Puget Sound Health Care System, Seattle, WA.

A rare missense variant R47H in the TREM2 gene is a recently identified strong risk factor for Alzheimer disease (AD). TREM2 is expressed by microglia and plays a role in immune responses in the CNS. We report a large multigeneration family affected with late-onset AD in which multiple affected members (11/15) carry the variant. Neuropathologic examination of 11 available brains from family members found features typical of AD that included neuritic amyloid plaques and neurofibrillary tangles (Braak stages 5-6) in all affected patients. Vascular pathology features were common in the R47H-positive patients (4/7). Results of ongoing assessments of microglia status in TREM2-R47H carriers will be reported.

908M

Functional study of a novel unexpected interferon-responsive gene, GRAMD1B, identified in Multiplex MS families. *F. Martinelli Boneschi[1], A.M. Osiceanu[1], F. Esposito[1], A. Zauli[1], M. Sorosina[1], B. Bettegazzi[4], D. Cittaro[2], E. Mascia[1], S. Santoro[1], A. Calabria[3], D. Lazarevic[2], D. Zacchetti[4], G. Comi[1], E. Stupka[2].* 1) DEPT NEUROLOGY AND INSPE, SCIENTIFIC INSTITUTE SAN RAFFAELE, MILAN, MI, Italy; 2) CENTER OF TRANSLATIONAL GENOMICS AND BIOINFORMATICS, SCIENTIFIC INSTITUTE SAN RAFFAELE, MILAN, MI, Italy; 3) SAN RAFFAELE TELETHON INSTITUTE FOR GENE THERAPY (HSR-TIGET), SCIENTIFIC INSTITUTE SAN RAFFAELE, MILAN, MI, Italy; 4) CELLULAR NEUROPHYSIOLOGY UNIT, SCIENTIFIC INSTITUTE SAN RAFFAELE, MILAN, MI, Italy.

BACKGROUND: While the role of common genetic variants is clearly established in multiple sclerosis (MS[MIM126200]), the heritability of the disease is still poorly explained, suggesting the existence of rare variants implicated in the susceptibility to the disease. OBJECTIVE: To identify low-frequency and rare genetic variants contributing to MS susceptibility in an Italian multiplex family. DESIGN/METHODS: SNP microarray genotyping and whole-genome sequencing (mean coverage: 20x) in 4 MS patients and 4 unaffected individuals belonging to an Italian multiplex family originating from a first cousin consanguinueous marriage have been performed. The Merlin software was used for the linkage analyses and SNPeff and GATK software were applied to prioritize rare variants. RESULTS: Filtering criteria, including a linkage analysis that identified a unique signal showing a LOD score <2.0 at chr11q23.3, narrowed down the list of variants up to a rare functional variant which determines an amino acid change, S601P, at an unexplored gene, *GRAMD1B*. The mutation is in a context that is highly conserved across species, and it segregates along the family consistent with an autosomal recessive transmission (p=0.02). By performing WG expression, we found that the gene was downregulated in affected relatives (p=0.01) with the exception of the only case who was IFNβ treated. We performed a IFNβ stimulation of PBMCs isolated from 20 healthy controls, showing an increase in *GRAMD1B* expression (p<0.001). We also observed a significantly higher expression of *GRAMD1B* in the brain tissue, and in immune cells, compared to other cells and tissues (p<0.05). Moreover, we observed a significant increase of expression in activated rat microglial cells compared to unstimulated ones (p<0.05), suggesting a putative role of the protein in the modulation of the immune system. Ongoing experiments are aimed to: 1) explore the role of GRAMD1B in the IFNβ pathway; 2) identify the localization of the protein; 3) study the effect of the mutation on the expression, localization and/or function of the GRAMD1B protein in MS etiology. . CONCLUSIONS: The use of next-generation sequencing approach in an Italian MS multiplex family has been successful in identifying a novel rare variant in an unknown gene recently found to be associated with IgE levels. Further investigations are ongoing to explore the role of the variant.

909T

The causal basis of Hirschsprung disease risk: functional consequences of polymorphisms in two *RET* shadow enhancers. *S. Chatterjee, A. Kapoor, A. Chakravarti.* Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

A significant proportion of the genetic risk of Hirschsprung disease (HSCR) in European ancestry subjects arises from two common variants at the gene encoding the receptor tyrosine kinase RET. We have previously shown that the first variant, rs2435357 (risk/non-risk alleles T/C, T control allele frequency 27%, odds ratio ~4), lies within RET intron 1 and disrupts binding of the transcription factor SOx10. We now demonstrate that the second variant, rs2506030 (risk/non-risk alleles G/A, G control allele frequency 40%, odds ratio ~2), maps 125kb upstream of RET and affects the binding of the transcription factor, Retinoic acid receptor beta (RARβ). In mouse Neuro2a cells, siRNA mediated knock down of either transcription factor reduces *Ret* expression, by 8 fold for *Sox10* ($P= 1.8\times10^{-4}$) and 5.5 fold for *Rarβ* ($P=4\times10^{-4}$). Surprisingly, when *Ret* is significantly ($P=1.1\times10^{-4}$) depleted by siRNA, expression of *Sox10* is decreased by 5 fold ($P=2\times10^{-5}$) but *Rarβ* remained unaffected (P=0.66), suggesting that Sox10 and Ret regulate each other. These results are supported by *in vivo* gene expression data in *Ret* homozygous null mice gut at 2 stages in development (E11.5 and E12.5), as compared to wild type mice at the same stages, where early gut neurogenesis occurs: *Sox10* expression is reduced 4 fold (E11.5, $P=1.2\times10^{-7}$) and 8 fold (E12.5, $P=1.2\times10^{-4}$) but *Rarβ* expression is, as expected, unaffected (E11.5, P=0.64; E12.5, P=0.19). In concert, *Gdnf*, the Ret ligand, expression is increased 1.5 fold (E11.5, P=0.03) and 3.4 fold (E12.5, $P=6\times10^{-4}$) as is the expression of the co-receptor *Gfra1* (3.7 fold at E11.5, P=0.02; 3 fold at E12.5, P=0.03). These in vitro and in vivo data show the presence of compensatory mechanisms within the *Ret* gene regulatory network. Specifically, *Ret*, the major gene for HSCR, is a positive regulator of its own transcription through Sox10. These results explain how common polymorphisms can lead to large genetic effects in HSCR since transcription attenuation of *Ret* from enhancer mutations are amplified through its auto-regulation. These results implicate *RET* as a key rate limiting step in early enteric nervous system (ENS) development and explains why >95% of HSCR cases have at least one *RET* loss-of-function allele. As a more general lesson, the phenotypic impact of transcriptional changes by common regulatory variants can only be understood in the context of its gene regulatory network.

910S

Functional regulatory assessment of the *APOL1* kidney disease risk variants. *P. An[1], J. Kopp[2], C.A. Winkler[1].* 1) Basic Research Laboratory, Leidos Biomedical Research, Inc., Frederick National Lab for Cancer Research, Frederick, MD; 2) Kidney Disease Section, National Institute of Diabetes and Digestive and Kidney Disease, Bethesda, MD.

Background: Apolipoprotein L1(encoded by *APOL1*) is a serum apolipoprotein bound to high-density-lipoprotein (HDL) particles. Common trypanolytic *APOL1* coding variants G1 (S342G:I384M) and G2 (2 aa deletion) are associated with markedly increased risk to kidney diseases including focal segmental glomerulosclerosis (FSGS) (OR 2-29) in Africa Americans and explain much of the excess risk for chronic and endstage kidney diseases. However, the functional basis of the APOL1 and the *APOL1* G1/G2 variants in relation with renal risk are largely unknown. We aimed to assess the potential regulatory functions of the *APOL1* kidney disease risk variants. **Methods:** We performed electrophoretic mobility shift assay (EMSA) to assess the binding of *APOL1* G1/G2 with nuclear transcription factors in the HEK 293 cell line, real-time PCR to quantify gene expression in the lymphoblastoid cell lines (LCLs) derived from FSGS patients, and gene reporter assay to test the regulatory potential of the variants. **Results:** We observed that G1 and G2 variants conferred differential transcription factor binding patterns, G2 allele carriers had increased mRNA gene expression, and the 3' terminus appeared to confer regulation activity. Our preliminary experimental data suggest that the *APOL1* G1/G2 variants may influence *APOL1* gene expression via differential transcription factor binding and gene regulation. **Summary:** *APOL1* gene expression may be differentially regulated by G1 and G2 variants; these results warrant further investigation in podocytes and other renal cell types. (Funded by the National Cancer Institute Contract HHSN261200800001E).

911M

Four regulatory variants alter protein binding and contribute to transcriptional activity at the *ANGPTL8* HDL-C GWAS locus. *M.E. Cannon, K.L. Mohlke.* Department of Genetics, University of North Carolina-Chapel Hill, Chapel Hill, NC.

Human genome-wide association studies (GWAS) have identified at least 157 loci that are associated with high-density lipoprotein cholesterol (HDL-C) and other blood lipid traits. A recent GWAS identified four DNA variants strongly associated with HDL-C ($r^2>0.8$ with rs737337, $P=4.6\times10^{-17}$) located upstream of the *ANGPTL8* gene, but the molecular mechanism of how the variants alter gene expression is unknown. All four variants are located in candidate regulatory elements (CREs) defined by epigenetic enhancer marks, transcription factor binding, and open chromatin marks that are correlated with *ANGPTL8* expression across tissues. *ANGPTL8* expression is limited to liver and adipose tissue, despite being entirely contained within one intron of ubiquitously expressed *DOCK6*. We evaluated allelic effects on transcriptional activity and protein binding across liver, adipose, embryonic kidney, and pancreatic islet cell lines. The promoter of *ANGPTL8* and the CREs containing the HDL-associated variants were cloned upstream of a luciferase gene in a promoterless or minimal promoter-containing vector and tested for allelic differences in enhancer activity using transcriptional reporter assays. The *ANGPTL8* promoter demonstrated transcriptional activity in the human HepG2 hepatocellular carcinoma cell line but not in human SW872 liposarcoma, mouse MIN6 insulinoma, or human 293T embryonic kidney cells. The CREs containing rs737337 and rs200788077 increased transcriptional activity 20-fold and 10-fold in HepG2, respectively. The remaining two CREs decreased transcriptional activity in HepG2, suggesting they have repressive activity. We also tested the variants for allelic differences in protein binding in vitro using electrophoretic mobility shift assays and observed allele-specific binding for all four variants. The HDL-decreasing allele rs737337-C demonstrated allele-specific protein binding in HepG2, but not in SW872 cells or MIN6 cells, suggesting that a liver-specific protein complex is binding to rs737337-C. Together, these results indicate that the *ANGPTL8* promoter is an important driver of transcriptional activity, which may be assisted by interactions with the variant-containing CREs. These data suggest a unique and complex molecular mechanism involving promoter-enhancer and/or promoter-repressor interactions that alter transcriptional activity at the *ANGPTL8* HDL-C GWAS locus.

912T

The expression of adipokines in the peripheral blood leukocytes of young patients with myocardial infarction. *R. Richterova[1], P. Cibulkova[1], H. Janyskova[1], P. Drahosova[1], J. Vaclavik[3], M. Svestak[2], D. Stejskal[2].* 1) Laboratore AGEL a.s., Novy Jicin, Novy Jicin, Czech Republic; 2) AGEL Research and Training Institute - Prostějov Branch, Central Moravian Hospital Group, Branch - Prostějov Hospital, Department of Laboratory Medicine, Czech Republic; 3) 3Department of Internal Medicine I - Cardiology, University Hospital Olomouc and Palacký University Faculty of Medicine and Dentistry, Olomouc, Czech Republic.

Adipokines are components of adipose tissue, which have important roles in energy homeostasis, sugar and fat metabolism, control of body temperature, reproduction, immunity and also have an effect on cardiovascular system (CS). Based on the effects on the CS they can be divided into cardioprotective adipokines and adipokines with negatively influence on the CS. Nowadays, little is known about the effect of adipokines in the pathogenesis of the acute coronary syndromes (ACS) in young patients. The aims of our study were: to evaluate possibility to detect mRNA expressions of selected adipokines in the peripheral blood leukocytes of young patients with ACS. We want to verify correlation between mRNAs expression and their products in blood and to evaluate these results with clinical status of tested probands. We prospectively examined 67 patients who had an acute myocardial infarction at a young age (men 18-45 years of age and women 18-55 years of age) and control groups of 79 patients with metabolic syndrome and 49 healthy individuals. RNA was isolated from their peripheral blood leukocytes and transcribed into cDNA. We examined expression below adipokines in peripheral blood leukocytes: aFABP4, Adiponectin, Adiponectin receptor 1, Adiponectin receptor 2, ANGPTL-4 Chemerin, FGF-21, GLP-1, Insulin, Insulin receptor, IL-6, Leptin, Leptin Receptor, Lipocalin 4 MCP-1, Omentin-1, Resistin, Retinol Binding Protein 4, Vaspin, Visfatin; as a housekeeping gene was used Beta-2-microglobulin. To confirm that it establishes mRNA from peripheral blood leukocytes was performed the following experiment: for 3 persons we separated liquid portion (plasma) cell fraction, which we have repeatedly purify with PBS. Then we performed expression isolation and quantification of adipokine's mRNA in whole blood, fluid and cellular portion. Adipokine's mRNA was found only in whole blood and cell content. We confirmed that we are really determined adipokines expression in peripheral leukocytes. The expression of adiponectin receptor 2, leptin and aFABP4 was detected in the group of patients after ACS. A correlation could not be found between aFABP4 gene expression and the concentration of its product in the serum of the group with ACS. We have shown that the mRNA expression of selected adipokines can be quantified in the peripheral blood leukocytes. Whether low expressions of protective adipokines contribute to early manifestations of ACS needs to be explored in further studies.

913S

An atopy-associated variant in the 11q13.5 locus regulates promoter activity. *J. Manz[1,2], A. Kretschmer[1,2,3], G. Möller[4], A. Peters[1,2], J. Adamski[4], M. Waldenberger[1,2], E. Reischl[1,2], S. Weidinger[3].* 1) Research Unit of Molecular Epidemiology, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health Neuherberg, Germany; 2) Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 3) Department of Dermatology, Venereology and Allergy, University Hospital Schleswig-Holstein, Kiel, Germany; 4) Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, Neuherberg, Germany.

Atopic dermatitis (AD) is a common chronic inflammatory skin disorder with a strong genetic component. Genome-wide association studies (GWAS) robustly identified common risk variants at 11q13.5 located in the intergenic region between *leucine rich repeat containing 32* (*LRRC32*) and *chromosome 11 open reading frame 30* (*C11orf30*). The same locus has also been associated with asthma, hayfever and Crohn's disease (CD). One GWA proxy SNP (rs2155219) is associated with the expression of *LRRC32* and *C11orf30*, but the functional relevance of this SNP has not been investigated yet. Therefore, the aim of this study was to characterize the functional impact of rs2155219 on cis-regulatory transcriptional activity and differential transcription factor binding. Reporter vector constructs carrying the major or minor allele of rs2155219, including two additional variants naturally occurring in this sequence (rs34455012indel; rs11236797C>A), were tested for regulating transcription activity in HaCaT (human keratinocyte), A549 (human alveolar basal epithelium), HeLa (human cervical cancer) and Jurkat (human T lymphocyte) cells using luciferase assays. In addition, different variant combinations within this haplotype (differing from the natural occurring allele composition) were investigated. Differential transcription factor binding was analyzed by electrophoretic mobility shift assays (EMSAs). Reporter vector constructs containing rs2155219 together with the naturally occurring variants acted as enhancer on promoters in HaCaT, A549, HeLa and Jurkat cells. The AD- and CD-risk associated T allele of rs2155219 in combination with rs34455012del and rs12236797A showed significantly higher promoter activity compared to the non-risk allele (in concert with rs34455012ins and rs12236797C allele). Analysis of further variant combinations (not naturally occurring) revealed cell-type specificity of single variants: an even stronger enhancer effect was observed dependent on rs34455012 deletion compared to rs34455012 insertion in A549 and HaCaT cells. Differential binding of putative protein complexes at selected variants was identified. Our study adds functional knowledge at the AD associated 11q13.5 locus by demonstrating that the GWA proxy SNP rs2155219 together with rs34455012indel and rs11236797C>A controls regulatory activity and impacts transcription factor binding.

914M

Obesity and metabolic disorders in children as part of the Reward Deficiency Syndrome: association with the SNP TaqIA C32806T of DRD2 gene. *R.M. Pinto[1,2], D.M. Silva[3], T.C Vieira[1,2], F.J. Queiroz[1,2], F.R. Godoy[1,2], I. Lacerda[1,2], L.S. Teodoro[1,2], M.W. Gonçalves[1,2], A.D da Cruz[1,2].* 1) Pontifícia Universidade Católica de Goiás - PUC/Goiás- Brasil; 2) Núcleo de Pesquisas Replicon- PUC Goiás - Brasil; 3) Universidade Federal de Goiás - Brasil.

The Reward Deficiency Syndrome (RDS) is a hypo-dopaminergic state that predisposes to obsessive-compulsive and impulsive behaviors. Obesity is part of RDS since the individual eats compulsively to compensate the defect in dopamine levels. The allele A1 (T) of the polymorphism TaqIA C32806T in the Dopamine D2 receptor gene (DRD2) is associated with reduction up to 40% of DRD2 levels, and is associated with higher BMI in adults. DRD2 are expressed in β cells and modulate insulin secretion. This is a case control study designed to investigate the relation between the SNP TaqIA C32806T of the DRD2 gene and obesity and metabolic alterations in children. We accessed the nutritional state of 105 children (55 obese/50 eutrophic) based on the definition of the World Health Organization (Z score of BMI). Peripheral blood samples were taken to determine the lipid profile, glucose, insulin and for analysis of DRD2 polymorphism through RFLP-PCR. The Homeostatic Model Assessment (HOMA) was calculated. We found A1 and A2 alleles, resulting in A1A1 (12,4%), A1A2 (33,3%) and A2A2 (54,3%) genotypes. The frequency of the A1 allele in obese and controls was 34,5%; and 23% respectively (p=0,05; RR=1,3). This allele was also associated with lower triglycerides (TG) levels and increased parents' BMI. The children were divided into groups according to the reference values of the metabolic variables studied. Significant difference in allelic distribution was observed in children with total cholesterol (TC) <170mg/dl or TC ≥ 170mg/dl. The HOMA β was abnormal in 52,7%of obese and 10% of eutrophic patients. Allele A1 was present in 38,2% of children with altered HOMA β and 24,6 % of children with normal HOMA β (p=0,037; RR=1,5). Up to the present, only 5 studies have studied the allele A1 frequency in obese children, ranging from 17% to 51%. Our results show for the first time that the A1 allele is associated with TC ≥ 170 mg/dl and lower TG levels. Another unpublished result in the literature is the association of A1 allele with HOMA β ≥175, relative risk conferred of 1,5. The A2 allele was associated with the normality of HOMA β in both obese and eutrophic groups, implicating this allele as a protective factor for pancreatic secretion. The recognition of predisposed individuals through determinations of risks polymorphisms can lead to new paths for treatment and prevention of metabolic abnormalities. We believe that in the future children will be treated based on their genomes.

915T

The International Genomics & Translational Research in Transplantation Network (iGeneTrain). *J. van Setten[1], Y.R. Li[2,3], M.V. Holmes[2], S. Bala[4], K. Birdwell[5], P.I.W. de Bakker[1], P. Jacobson[6], K.J. Karczewski[7], D. MacArthur[7], J.H. Moore[8], W. Oetting[6], M. Snyder[9], J.D. Christie[2], G. Lord[10], E. Schadt[11], A.K. Israni[12], A. Shaked[2], F.W. Asselbergs[1], B.J. Keating[2,3] on behalf of the iGeneTrain consortium.* 1) University Medical Center Utrecht, Utrecht, Netherlands; 2) University of Pennsylvania, Philadelphia, USA; 3) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, USA; 4) Yale Medical School, New Haven, USA; 5) School of Medicine, Vanderbilt University, Nashville, USA; 6) College of Pharmacy, University of Minnesota, Minneapolis, USA; 7) Harvard Medical School and Broad Institute, Massachusetts General Hospital, Boston, USA; 8) School of Medicine, Dartmouth, Hanover, USA; 9) Stanford Center for Genomics and Personalized Medicine, Stanford University, Stanford, USA; 10) Guy's and St. Thomas' Hospital and King's College London, London, UK; 11) Institute for Genomics and Multiscale Biology, Icahn School of Medicine, Mount Sinai, New York, USA; 12) Hennepin County Medical Center, Minneapolis, USA.

Over the last two decades, over 300,000 solid organ transplantations have been performed in the USA alone. While there has been significant progress with immuno-suppression therapies, the risk of graft organ rejection is still substantial, even when appropriate HLA matching has been performed. Transplant rejections from male-donor to male-recipient and females to females are much lower than male to female. Cross-ethnic rejection rates are also higher, suggesting that additional genetic factors are at play in the rejection process. Possible sources of genetic variation underpinning rejection are homozygous deletion copy number variants (CNVs) spanning whole gene or exon regions and Loss of Function (LoF) variants ablating two copies of a given gene, resulting in incompatibility across the proteomes of donor and recipient. It has been shown through large-scale whole genome sequencing studies that the average human has 20 genes disrupted (i.e. with LoF) in two copies. Here, we present iGeneTrain, a large-scale international consortium, which ultimately aims to translate genetic data into clinical applications such as more optimal genomic compatibility matching of donor-recipient (D/R) pairs and immune suppression therapy dosing. We designed a low-cost genome-wide SNP array specifically tailored for the transplantation community, including exonic and LoF content, pharmacogenomic markers, CNVs, expression QTL markers, and HLA alleles. To date, we have included over 21,000 individuals. We will perform association testing within groups and cross-organ, maximizing statistical power to identify genetic variation associated with vasculopathy, rejection, and death.

916S

Repair activity of primate-specific alternative single-stranded DNA binding protein aRPA may explain brain-region repeat instability in CAG/CTG trinucleotide repeat diseases. *J. Luo[1], A. Mason[2], M.S. Wold[3], Y.H. Wang[4], A. LaSpada[2], C.E. Pearson[1].* 1) The Hospital for Sick Children, Toronto, Ontario, Canada; 2) University of California, San Diego; 3) University of Iowa; 4) University of Virginia.

Expansion of gene-specific trinucleotide repeat sequences, CAG/CTG causes at least 14 neuromuscular and neurodegenerative diseases. Continued somatic CAG expansions can contribute to disease onset and progression, thereby highlighting a need to understand the expansion process or processes that inhibit this. The processing of mutagenic slipped-DNA structures formed within the repeats is thought to play a role in repeat expansion. Replication protein A (RPA) is a single-strand binding protein that is essential in many DNA metabolic processes, including DNA replication, repair, and recombination. We previously revealed higher levels of RPA in the cerebellum compared to the striatum correlated with the lower levels of CAG expansions in the cerebellum versus the larger expansions in the striatum - which is sensitive to degeneration. A related primate-specific protein called alternative RPA (aRPA) has been shown to be expressed in human tissues at a very low levels relative to RPA; however, its function and relationship to RPA is not known. Here we show that aRPA transcript levels are greatly amplified in the brains of Huntington's disease (HD) patients compared to control brains, while expression of RPA is unaltered. We assessed the ability of RPA and aRPA to repair and process mutagenic slipped-DNA structures at the CAG/CTG repeats and found that RPA enhances the correct repair and processing of slipped-DNAs. However, high levels of aRPA (at aRPA:RPA ratio approximating those in HD brains) can inhibit RPA and eliminate repair. These findings suggest that RPA may protect against trinucleotide repeat mutations and aRPA may modulate this, a phenomena that may be particularly exacerbated in Huntintons patient brains. This furthers our understanding of the both disease-causing repeat mutations, the functions of single-strand binding proteins RPA and aRPA, and disease-specific changes.

917M

Novel Indoleamine 2, 3-Dioxygenase (*IDO*) gene mutation in the Pathogenesis of age-related cataract. *P. Gunda[1], M. Mamata[1], R.K.R Kondreddy[2], N. Thurlapati[3], P. Tirunilai[1].* 1) Dept of Genetics, Osmania University, Hyderabad, India; 2) Sarojini Devi Eye Hospital, Mehdipatnam, Hyderabad, India; 3) Department of Zoology, Osmania University, Hyderabad, India.

Purpose: Exposure to UV light is considered as the major risk factor for the development of age-related cataract. UV filters that are produced during tryptophan catabolism maintain the transparency of the lens and also protect retina from photodamage. Indoleamine 2, 3 dioxygenase (IDO) is the first rate limiting enzyme in the tryptophan catabolism which is encoded by *IDO* gene localized on chromosome 8p12-p11 region. Mutations in *IDO* gene can affect synthesis of UV filters qualitatively or quantitatively. Hence the present study was planned to screen for mutations in *IDO* gene and to evaluate their role in the causation of age-related cataract. **Methods:** Genomic DNA from 331 age-related cataract cases [110-Nuclear cataract, 110-Cortical cataract, 111-Posterior Subscapular types and 210 normal controls were PCR amplified for all regions covering the entire coding sequence and splice junctions of *IDO* gene. Amplified PCR products were analyzed using SSCP followed by sequencing of variant samples in both directions with an automated genetic analysis system (3100; ABI). Structure prediction and energetic analysis of wild-type IDO compared with mutants were predicted using various Bioinformatic tools. **Results:** Mutational screening in exon-7 of *IDO* gene showed the presence of a novel variation (c.596_597 delins TT; rs267606590) in heterozygous pattern in 2 (1 NC & 1 CC) of the cases studied and none among controls, leading to substitution of alanine (A) at position 199 by valine (V) in mutants. SIFT and polyphen tools predicted possible damaging effect of mutation on protein. Protein modeling by Triton package indicated wide variation between the wild type and mutant protein structure. Km values of wild type (68.66±0.26 μM) and mutant IDO proteins (74.92±0.40 μM) showed a significant difference indicating reduced affinity between the enzyme and substrate for mutant protein as compared to wild type. **Conclusion:** The novel variation p.A199V in exon 7 of *IDO* gene among patients with NC and CC cataracts resulting in the low affinity between mutant IDO enzyme and substrate tryptophan is likely to reduce the rate of UV filter synthesis in lens. The reduced levels of UV filters may cause exposure of lens to UV light resulting in aggregation of lens proteins and development of cataract.

918T

Genetic Variants Concomitantly Influence Nonalcoholic Fatty Liver Disease and Correlated Metabolic Traits. *M.F. Feitosa[1], A.T. Kraja[1], E.K. Speliotes[2], M.K. Wojczynski[1], N.D. Palmer[3], L.M. Yerges-Armstrong[4], T.B. Harris[5], C.S. Fox[6], I.B. Borecki[1], GOLD Consortium, GLGC Consortium, CHARGE Adiposity Consortium, GIANT Consortium, MAGIC Consortium.* 1) Department of Genetics, Washington University School of Medicine, St. Louis, MO; 2) Department of Internal Medicine, Division of Gastroenterology and Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 3) Forest School of Medicine, Winston-Salem, NC; 4) Department of Medicine, University of Maryland, Baltimore, MD; 5) National Institute on Aging, National Institutes of Health, Bethesda, MD; 6) Framingham Heart Study, National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health, Framingham, MA.

Nonalcoholic fatty liver disease (NAFLD) is associated with metabolic disorders, including overall and central obesity, dyslipidemia (high levels of triglycerides and low levels of high-density lipoprotein cholesterol), and impaired glucose tolerance (high fasting levels of glucose and insulin). Several variants have been discovered by genomewide association (GWA) studies for NAFLD and for these correlated metabolic disorders; however, few efforts have focused on determining whether the discovered genes/loci concomitantly influence these traits. To identify common variants contributing to pleiotropic effects on NAFLD and correlated metabolic traits, we employed a correlated meta-analysis (CMA) assessing the large published GWA meta-analysis Consortium data of CT measured fatty liver (FL, GOLD): with (i) triglycerides and (ii) high density cholesterol (GLGC), (iii) fasting levels of glucose and (iv) insulin (MAGIC), (v) body mass index and (vi) waist circumference (GIANT), and (vii) abdominal adipose tissues (CHARGE). Our methodological approach yielded evidence of genetic variants (e.g., ZNF14-rs12608822, SUGP2-rs3810444, NOTCH4- rs3130320, ATRAID-rs7437, NAT2-rs1495743, SMPD3-rs2863973, and GCKR-rs780093) with pleiotropic effects on CT measured FL and these metabolic traits. In addition, the CMA approach offered the opportunity to discover novel variants (APBA1-rs1330326, TYW3-rs11161851, LOC400940-rs2693827 and HLA-DKA2-rs2858310) that did not reach the GWA significance level in the primary discovery efforts. The associated variants are enriched for regulatory genetic functions, predict signaling pathway networks, and are associated with cell-specific-types and tissues involved in central nervous system, energy balance, glucose homeostasis, lipid metabolism, and fat liver accumulation. Our findings provide insight into the genetic basis of the correlated architecture of NAFLD metabolic pathway.

919S

Epigenetic effects of environmental enrichment and EGCG treatment on a mouse model of Down Syndrome. *C.N. Hor[1,2], S. Jhanwar[1,2], S. Espeso-Gil[1,2], M. Pons-Espinal[1,2], M.M. de Lagran[1,2], M. Friedländer[1,2], M. Dierssen[1,2], X. Estivill[1,2], S. Ossowski[1,2].* 1) Centre for Genomic Regulation, Barcelona, Barcelona, Spain; 2) Universitat Pompeu Fabra (UPF), Barcelona, Spain.

Epigenetic marks, defined as modifications to the DNA molecule and higher order structure not affecting the nucleotide sequence, have been hypothesized to be the long sought after interface between genes and the environment. The epigenetic state of a locus, i.e. the local combination of epigenetic marks, is thought to be determined partly by the DNA sequence and influenced by environmental cues. To directly explore the effects of genotype and environment under controlled conditions, we subjected transgenic Dyrk1A overexpressing mice (a model of Down syndrome) and their wild-type littermates to an enriched environment and/or treatment with the green tea polyphenol epigallocatechin gallate (EGCG, a Dyrk1A kinase inhibitor), both of which were previously shown to improve the cognitive performance of the transgenic mice. We have generated genome-wide profiles of cytosine methylation, a selection of histone marks and open chromatin in the cerebral cortices of these mice, as well as transcriptome and proteome. The epigenomic data are combined by computational analysis to yield a map of chromatin states, which can then be correlated with transcript and protein levels in each genotype and treatment group. Our analysis can (1) guide the discovery of regulatory elements active in the mouse brain, (2) uncover environment-specific signatures in the epigenome when comparing the profiles of different exposures across one genotype, and (3) elucidate the differential response attributable to the transgene by comparing these changes between the two genotypes. Preliminary analyses suggest that DNA methylation patterns remain stable upon treatments, while H3K36me3, a histone modification present along active gene bodies, is differentially detected in a set of genes enriched in annotation terms related to long term potentiation, which is important for memory and learning.

920M

The Molecular Convergence of non-HLA Ankylosing Spondylitis Risk Genes with Autoimmune Diseases. *D. O'Rielly[1], M. Uddin[2], M. Husan[1], D. Codner[1], P. Rahman[1].* 1) Faculty of Medicine, Memorial University, St. John's, Newfoundland, Canada; 2) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario.

Genome-wide association studies (GWAS) and subsequent follow-up studies validated a significant number of non-HLA ankylosing spondylitis (AS) risk genes. Although these genes are independently recognized as small contributory risk factors to the pathogenesis of AS, the combined effect is yet to be quantified. In this study, we hypothesize that AS risk genes converge at the molecular level with related autoimmune disease genes. To identify such convergence a systematic approach is required to further investigate and understand AS pathobiology. A protein-protein interaction network-based method was utilized to infer connectivity between AS genes identified in GWAS studies with other autoimmune risk gene sets. We investigate 418 risk genes from ten (10) autoimmune diseases and their complex interactions with AS. The constructed PPI network demonstrated significantly ($p<1.8 \times 10^{-4}$ after 50,000 permutations) dense modular connectivity for AS genes with six other autoimmune disease genes including Crohn's disease (CD), psoriasis (Ps), ulcerative colitis (UC), celiac disease (CeD), multiple sclerosis (MS), and primary biliary cirrhosis (PBC). The most significant connectivity observed (after correcting for gene number) was with CeD and UC, which is consistent with the known pathobiology of AS, where CD and CeD often manifest as comorbid conditions. To better understand the shared pathway and to identify etiological genes, each non-AS risk gene was ranked according to their contributing connectivity to AS associated genes. We have identified STAT3, JAK2, IL12RB2, LSP1, and NFKBIA genes which are highly clustered with AS associated genes. We have developed a systematic approach to infer causal genes that converge at the molecular level to a common pathway involved in disease pathogenesis. The highly connected non-AS genes identified in this study will be prioritized for further investigation to identify their contribution to AS pathogenesis and to better understand the complex relationship of autoimmune diseases.

921T

Characterizing the Nphp10 (Sdccag8Tn(sb-Tyr)2161B.CA1Cove) mouse model. *K. Weihbrecht[1,2], V. Sheffield[1,3], S. Seo[2].* 1) Dept. of Pediatrics, Univ. of Iowa, Iowa City, IA; 2) Dept. of Ophthalmology, Univ. of Iowa, Iowa City, IA; 3) Howard Hughes Medical Institute, Univ. of Iowa, Iowa City, IA.

A subset of nephronophthisis genes show normal localization to the ciliary-centrosomal complex, creating a linked group of disorders described as nephronophthisis-related ciliopathies (NPHP-RC). Aside from early end-stage renal failure and kidney cysts, seen in patients with just NPHP, the two other main phenotypes of NPHP-RCs are retinal and cerebellar degeneration. One such disease with NPHP-RC phenotypes is Bardet-Biedl Syndrome (BBS), with cardinal phenotypes of retinal degeneration, renal abnormalities, obesity, polydactyl, and learning disabilities. One NPHP gene associated with BBS is serologically defined colon cancer antigen 8 (SDCCAG8), later defined as NPHP10. However, the association with this disease was only recently made and little is known about the molecular function and how it leads to the observed phenotypes. BBS patients with NPHP10 mutations exhibit retinal and renal abnormalities, obesity, and learning disabilities. The purpose of this study is to characterize mouse model Sdccag8Tn(sb-Tyr)2161B.CA1COve, generated by a Sleeping Beauty Transposon (SBT) insertion between exons 12 and 13 of Nphp10. The goal in characterizing this mouse model is to determine its efficacy as a BBS or NPHP disease model, as well as to determine its use for functional analysis of NPHP10. NPHP10 has not been shown to interact with known NPHP complexes, nor has it been shown as an interactor with the other current BBS genes. Thus, the pathway through which NPHP10 leads to BBS phenotypes is novel and this mouse can be used to study the severity of typical BBS phenotypes in a BBS gene found in only a small subset of patients. While Jackson Laboratory reports an SBT insertion between exons 12 and 13 of Nphp10, our work has determined that a deletion occurred after the SBT insertion site, resulting in a loss of Nphp10's exons 13-18 as well as exons 2-13 of neighboring gene Akt3. Using these animals, we have also characterized a subset of phenotypes, including fully penetrant hind limb preaxial polydactyl and patterning defects of the secondary palate via Alcian Blue and Alizarin Red staining and developmental defects of the brain and abnormalities of the lung via H&E staining. Polydactyl, patterning defects, and lung abnormalities are consistent with phenotypes seen in other ciliopathy gene deficient mice and are likely due to loss of Nphp10, whereas the developmental defects of the brain are consistent with an Akt3 knockout animal.

922S

Transcriptome analysis of differentially expressed isoforms in hypertrophied cardiomyocytes derived from human iPSCs using RNA-Seq. *W. Li, P. Aggarwal, A. Turner, A. Matter, U. Broeckel.* Pediatrics Genomics, Medical College of Wisconsin, Milwaukee, WI.

Purpose. Left ventricular hypertrophy (LVH) is a potent risk factor for cardiovascular disease. The development of LVH is accompanied by significant changes in gene expression patterns in cardiomyocytes (CMs). However, traditional methods of transcriptome analysis do not analyze the abundance of individual splice forms (SFs), leaving the potential role of differentially expressed isoforms in complex diseases, including LVH, unexplained. Taking advantage of RNA sequencing, we identify a set of genes that are differentially expressed in hypertrophied CMs and explore the contribution of individual splice-forms to the LVH-transcriptome profile. Methods. We developed a model of human CM hypertrophy using induced, pluripotent stem cell derived CMs (iPSC-CMs). Stimulating the iPSC-CMs with isoproterenol induced characteristic changes associated with the development of LVH, evidenced by an increase in cell surface area and two well-established LVH markers, c-fos and BNP. RNA-Seq was performed for three replicates. Transcripts were assembled using TopHat2 and Bowtie2, and analyzed for relative levels of expression using Cufflinks2. Results. Upon hypertrophic induction, 920 genes exhibited significantly differential expression. 433 of these genes have one SF and 487 have two or more (55 genes had ≥ 5 SFs). Of those that are alternatively spliced, 367 (75.4%) showed significant changes at the global level only, 15 (3.1%) at the splice-form level only, and 105 (27.1%) at both. To explore potential functional implications of differentially expressed SFs, we compared the exons that are unique to only some SFs of each gene to those that are common (common exon) to all SFs. On average, unique exons are approximately twice as long as common exons. Further, unique exons contain numerous well-characterized functional domains. Functional characterization of differentially expressed isoforms will likely provide novel insights into the etiology of LVH. Conclusions. Our study demonstrates that differential splice form expression is prevalent during the development of LVH and likely contributes to the physiology of hypertrophied CMs. The same is likely true for many other complex diseases. Reliance on analysis of global expression patterns alone obscures changes associated with individual splice forms, resulting in a substantial loss of information and potential insights into the development and physiology of complex diseases.

923M

Maternal effects influence the heritability of adult obesity traits but not obesogenic growth trajectories in a model system. *C.A. Schmitt[1,2], S. Service[1], R.M. Cantor[3], A.J. Jasinska[1], M.J. Jorgensen[4], J.R. Kaplan[4], N.B. Freimer[1], The International Vervet Research Consortium.* 1) Center for Neurobehavioral Genetics, University of California - Los Angeles, Los Angeles, CA; 2) Human Evolution Research Center, Museum of Vertebrate Zoology, University of California - Berkeley, Berkeley, CA; 3) Department of Human Genetics, University of California - Los Angeles, Los Angeles, CA; 4) Department of Pathology, Section on Comparative Medicine, Wake Forest University School of Medicine, Winston-Salem, NC.

Obesity arises from a complex interaction of genetic predisposition and environment that can accumulate throughout life. Previous work has shown that there are maternal effects on obesity outcomes in humans and nonhuman primates, but few have investigated the impact of maternal effects on obesogenic growth itself throughout the lifespan. This research investigates the impact of maternal effects on the genetic underpinnings of obesogenic growth from birth to adulthood in a genetically well-characterized model system under a controlled diet and environment: the African green monkey (*Chlorocebus aethiops sabaeus*) in the Vervet Research Colony at Wake Forest School of Medicine. We used growth curve analysis on measures taken thrice yearly on body size and composition - such as body weight (BW) and BMI - in a population of 641 monkeys measured from 2000 through 2013. Of these, 33 individuals, 6 males and 27 females, presented with signs of chronic abdominal obesity - defined as having an adult waist circumference above 40.5 cm for at least three successive measurements. Individuals measured < 6 times were excluded from analysis (max measures = 30, mean = 12.6 ± 6.12). Growth was modeled using three-parameter logistic growth curves in nonlinear mixed models, with parameters modeled as fixed effects and subject and sex/obesity status modeled as random effects. We assessed heritability of individual growth parameters using maximum likelihood variance components analysis in SOLAR, with the variance attributed to maternal ID (c^2) partitioned from environmental variance to determine maternal effects. We found significant heritability and maternal effects on all static measures of adult body condition (e.g., BW, $h^2 = 0.86$, $p = 6.52\times10^{-10}$ and $c^2 = 0.11$, $p = 0.04$; BMI, $h^2 = 0.77$, $p = 1.39\times10^{-09}$ and $c^2 = 0.12$, $p = 0.05$), and high heritability but no evidence for significant maternal effects on parameters of growth (e.g., BW: θ_1, asymptote of growth, $h^2 = 0.77$, $p = 4.93\times10^{-21}$ and $c^2 = 0.05$, $p = 0.11$). This study suggests that although adult obesity is a developmental process driven in part by heritable obesogenic trajectories resulting in faster and longer growth to larger adult size, those trajectories do not appear to be influenced by maternal effects. A better understanding of how growth can be decoupled from maternal effects on adult obesity outcomes will be necessary to assess early obesity risks and inform potential biomedical interventions.

924T

The Genetic Landscape of Hematopoietic Stem Cells. *H. Allayee[1], A.L. Crow[1], J. Hartiala[1], T.J. Spindler[2], A. Ghazalpour[3], L.W. Barsky[2], B.J. Bennett[4], B.W. Parks[3], E. Eskin[3], R. Jain[5], J.A. Epstein[5], A.J. Lusis[3], G.B. Adams[3].* 1) Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90033; 2) Broad Center for Regenerative Medicine and Stem Cell Research at USC, University of Southern California, Los Angeles, CA 90033; 3) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095; 4) Department of Genetics and Nutrition Research Institute, University of North Carolina, Chapel Hill, Kannapolis, NC 28081; 5) Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104.

The ability to carry out association studies on a genome-wide basis has revolutionized the field of human genetics. Such GWAS have identified hundreds of novel genes for atherosclerosis, diabetes, cancer, and related quantitative phenotypes. However, human GWAS can be limited since other biomedically relevant phenotypes, such as hematopoietic stem cell (HSC) number/function, are logistically difficult to obtain in large numbers of subjects. To overcome these limitations, we used a GWAS approach with a recently developed panel of inbred and recombinant mouse strains, termed the Hybrid Mouse Diversity Panel (HMDP), to characterize the genetic landscape of HSCs. We used flow cytometry to quantitate the frequency of long-term (LT) and short-term (ST) HSCs, as well as hematopoietic progenitor cells (HPCs), in bone marrow (BM) from 111 HMDP strains. These cell populations varied 25 to 80-fold, thus providing strong evidence for a genetic basis to HSC variation. Consistent with this notion, a GWAS analysis identified 8 loci significantly associated with HSCs (p=6.9E-6 - 7.0E-14), including a locus for ST-HSCs on chromosome 5 (p=6.2E-7). Of the genes in this region, highly significant cis eQTLs was identified for homeodomain only protein (Hopx) in liver (p=3.4E-18) and heart (p= 6.2E-19), suggesting that genetic variation at this locus is functional with respect to Hopx expression. Hopx expression in BM from 25 HMDP strains was significantly positively correlated (r=0.44; p<0.05) with ST-HSCs. To functionally validate Hopx, we characterized BM from Hopx knockout (Hopx -/-) mice by flow cytometry and observed a significant 4-fold reduction in ST-HSC frequency compared to wildtype littermates (p<0.01). Additional analyses revealed that Hopx-/- mice had significantly reduced numbers of quiescent ST-HSCs (60% vs 85%; p<0.05) and increased numbers of cells in the G1 phase (35% vs 10%; p<0.05). Hopx deficiency specifically affected ST-HSCs since there were no differences in the frequency or cell cycle status of HPCs or LT-HSCs. These findings are entirely consistent with Hopx mRNA levels being positively correlated with ST-HSCs and the Hopx locus only being associated with this HSC subset in the GWAS. Taken together, these results provide compelling evidence that Hopx is a novel positive regulator of HSC physiology in mice and demonstrate the power of the HMDP for elucidating the genetic architecture of complex traits that would otherwise not be feasible in humans.

925S

Genotyping-by-sequencing in outbred CFW mice yields a powerful approach for genome-wide mapping of complex trait loci. *P. Carbonetto[1], S. Gopalakrishnan[1], C.C. Parker[2], N.M. Gonzales[1], A. Lionikas[3], C.L. Ackert-Bicknell[4], E.H. Leung[1], E. Aryee[1], J. Park[1], J. Davis[5], A.A. Palmer[1].* 1) Dept. of Human Genetics, University of Chicago, Chicago, IL; 2) Dept. of Psychology and Program in Neuroscience, Middlebury College, Middlebury, VT; 3) School of Medical Sciences, College of Life Sciences and Medicine, University of Aberdeen, Foresterhill, Aberdeen, UK; 4) The Jackson Laboratory, Bar Harbor, Maine; 5) Dept. of Genetics, Stanford School of Medicine, Stanford, CA.

Despite the successes of genome-wide association studies for augmenting our understanding of many common human diseases and disease-related traits, mice still remain an important complementary resource for studying genetics of complex disease. Advantages include being able to control the environment, and measure gene expression in tissues that are ordinarily inaccessible in humans. (We present gene expression from multiple brain regions in a separate poster.) One important disadvantage of working with traditional mouse lab strains compared to humans is that the mice exhibit long-range correlations on their chromosomes, due to a lack of accumulated recombination, often making it very difficult to assign any functional interpretation to a complex trait locus. We describe our efforts to develop a new resource for mapping complex trait loci from CFW mice, a commercial stock of outbred mice, that permits better resolution to map complex traits. We show that CFW mice have desirable properties for QTL mapping, including rapid decay of linkage disequilibrium and a high proportion of common alleles. To ascertain whole-genome variation at low cost, we adapt "genotyping-by-sequencing" (GBS) protocols originally developed for plants, and contribute additional technical improvements to this protocol for outbred mice. To address the problem of false positives caused by "cryptic" or "hidden" relatedness, we test for association using the linear mixed model approach, implemented in the software GEMMA. In a cohort of ~1000 CFW mice, we obtain strong support for genetic associations in a variety of complex traits, including bone-mineral density, testes weight, musculoskeletal traits, prepulse inhibition and methamphetamine sensitivity. In some QTL regions, the association signal isolates only a few genes. We also identify loci that overlap with established complex trait genes, such as testes weight gene Inhba and osteoporosis gene Col1a1. Many of these associations have been replicated in an independent panel of CFW mice developed by our collaborators at University of Oxford. Our findings show that the CFW mouse stock is a practical and powerful resource for identifying loci underlying a variety of complex traits.

926M

An APP, BACE expressing C. elegans model of Alzheimer's disease. *K.N. Ly[1], S.J. Redi[1], K. Henty[1], W.N. Grant[2], R.G. Russell[1].* 1) School of Biolocial Sciences, The University of Auckland, Auckland, New Zealand; 2) Department of Genetics, La Trobe University, Melbourne, Victoria, Australia.

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and is the most common form of dementia, affecting over 44 million people worldwide. Pathologically, AD is characterized by two important hallmarks: extracellular amyloid plaques and intracellular neurofibrillary tangles, made up predominantly by proteins known as β-amyloid (Aβ) and hyperphosphorylated Tau, respectively. Mutations in the Aβ precursor protein (APP) and two proteins involved in Aβ production (presenilin 1 and 2; components of the gamma secretase complex) cause heritable early onset AD, typically prior to 65 years old. However, the majority of AD cases occur in people over the age of 65 with unknown mutations. Despite over two decades of intensive research, the physiological roles and mechanistic contribution of APP and Aβ to late onset AD still remain elusive. *Caenorhabditis elegans* is emerging as a powerful tool in the research of neurodegenerative disorders, complementing the traditional mammalian systems with ease of genetics, simplicity of their neuronal network, and short generation time. Previous studies have shown that transgenic *C. elegans* over-expressing human Aβ in muscle cells develop intracellular Aβ aggregation and a progressive paralysis, while those expressing Aβ in neurons reveal behavioral deficits and rapid paralysis. To further understand Aβ and the processing of its precursor APP, transgenic *C. elegans* strains which express full-length human APP and human β-secretase (an enzyme responsible for the first step of Aβ production) under three promoters: neuronal specific (*snb-1*), *apl-1* (*C. elegans* APP homologue), *sel-1* (*C. elegans* presenilin homologue) have been developed. The approach taken has been to integrate single copies of the human genes using the MOSSCI technique, with and without fluorescent tags, in order to achieve stable expression levels. Promoter appropriate expression has been observed and results from developmental and behavioral studies will be presented.

927T

Characterization of neuronal development in autism using induced pluripotent stem cells reveals disease-specific changes in neurite outgrowth and expression of synaptic function genes. *B.A. DeRosa[1,2], K.C. Belle[1,2], J.M. Lee[1], M.L. Cuccaro[1,2], J.M. Vance[1,2], M.A. Pericak-Vance[1,2], D.M. Dykxhoorn[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Dr. John T. MacDonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL.

Autism spectrum disorders (ASD) comprise a genetically and phenotypically heterogeneous group of neurodevelopmental conditions that are characterized by manifestations of several broadly defined phenotypes, including impairments in communication, social interaction, and behavior. The rapid expansion in the number of newly identified ASD risk loci, coupled to the absence of therapeutics and pharmacological treatments, has shed light on the need for a better understanding of the mechanisms by which specific genetic variants lead to pathophysiological changes associated with the disease. Until recently, functional studies of idiopathic autism were hindered by the lack of genetically-relevant human disease models. With the advent of human induced pluripotent stem cell (iPSC) technology, and recent advancements in neural differentiation methods, it is now possible to functionally analyze specific populations of developing neurons derived from ASD-affected individuals. Since in vitro neuronal differentiation mimics in vivo neurogenesis and neuronal maturation, it is possible to analyze ASD-specific phenotypes throughout neuronal development. In order to characterize the impact of non-syndromic forms of ASD on neurodevelopmental processes, including neurite outgrowth dynamics and network formation, we generated a panel of patient-specific iPSC lines from peripheral blood mononuclear cells attained from clinically and genetically well characterized ASD patients. Neurite outgrowth dynamics and ASD-risk gene expression were analyzed in iPS-derived ASD and control forebrain neurons over a broad time course spanning in vitro neuronal differentiation and maturation. Our results show ASD-associated changes in gene expression profiles and patterns of neurite outgrowth. These analyses provide a solid framework for understanding the molecular and cellular mechanisms that underlie ASD pathophysiology.

928S

Association of Age-related Macular Degeneration (AMD) Susceptible Genes with Second Eye Involvement of AMD. *M. Miyake[1,2], K. Yamashiro[1], K. Kumagai[1], M. Saito[3], M. Kuroda-Sugahara[1], M. Yoshikawa[1], H. Tamura[1], Y. Akagi-Kurashige[1,2], M. Oishi[1], I. Nakata[1,2], H. Nakanishi[1,2], A. Oishi[1], N. Gotoh[1,2], A. Tsujikawa[4], T. Sekiryu[3], Y. Kurimoto[5], C.C. Khor[6], C.Y. Cheng[7], T.Y. Wong[7], R. Yamada[2], F. Matsuda[2], N. Yoshimura[1].* 1) Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Center for Genomic Medicine/Inserm U.852, Kyoto University Graduate School of Medicine, Kyoto, Japan; 3) Fukushima Medical School. Fukushima, Japan; 4) Kagawa University, Kagawa, Japan; 5) Kobe City General Hospital, Kobe, Japan; 6) Genome Institute of Singapore, Singapore, Singapore; 7) Singapore Eye Research Institute, Singapore.

BACKGROUND: Age-related macular degeneration (AMD) is a major cause of progressive, irreversible visual impairment among elderly population in developed countries. Especially, bilateral involvement of AMD severely impairs quality of life, so that predicting fellow-eye-involvement of AMD in unilateral cases is of great importance in clinical setting. In 2013, the largest genome-wide association study on age-related macular degeneration (AMD) identified 19 AMD susceptible loci, including 7 novel loci (The AMD Gene Consortium, Nature Genetics). Of these, ARMS2 was reported to be associated with the second eye involvement of AMD, while others have not been evaluated. In this study, we investigated whether these loci were also associated with second eye involvement using Japanese AMD cohorts recruited from 3 institutes. **METHODS:** A total of 1,576 AMD cases were genotyped using illumina OmniExpress, HumanOmni2.5-8, and/or HumanExome, which was imputed to 1,000 genomes cosmopolitan data after a standard quality control. We retrospectively reviewed medical records of 500 unilateral AMD patients from Kyoto University Hospital, 264 patients from Fukushima Medical University, and 130 patients from Kobe City General Hospital. Written informed consent for genetic research had been obtained from all patients. Cox proportional hazard regression analysis was used to examine the association between the genotypes of AMD susceptible single nucleotide polymorphisms (SNPs) and the duration until the development of AMD in the second eye. **RESULTS:** ARMS2 rs10490924 recessive model (Hazard Ratio [HR]$_{meta}$ = 2.04; $P_{meta} = 3.4 \times 10^{-3}$) and CFH rs800292 additive model (HR_{meta} = 1.77; P_{meta} = 0.013) revealed significant association with second eye involvement. Dominant model of TNFRSF10A rs13278062, VEGFA rs943080, and CFI rs4698775 showed consistent effect across 3 datasets (I^2 = 0%; HR_{meta} = 1.46, 1.30, 1.51, respectively). Next, we calculated genetic risk score (GRS) using these 5 SNPs. After 10 years from the first visit, the patients with top-10-percentile of GRS showed 51% of hazard rate, in contrast to 2.3% among patients with lower-10-percentile of GRS. **CONCLUSIONS:** GRS using ARMS2, CFH, TNFRSF10A, VEGFA, and CFI is a good predictor of second eye involvement, which is especially of great importance in the clinical setting.

929M

Nasopharyngeal microbiome composition is associated with lung function in adult Hutterites. *C. Igartua[1], E.R. Davenport[1], J.B. Becker[2], S.R White[2], Y. Gilad[1], C. Ober[1], J.M. Pinto[3].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Section of Pulmonary and Critical Care Medicine, Department of Medicine, University of Chicago, Chicago, IL; 3) Section of Otolaryngology-Head and Neck Surgery, Department of Surgery, University of Chicago, Chicago, IL.

The human microbiome contains trillions of microbes and has been associated with disorders of the airway, such as sinusitis and asthma. We predict that host immunity is related to the composition of bacteria in the airways and that ecologic characteristics of these bacterial communities may be predictive of respiratory disease state. We explore this possibility by considering the relationship between the nasopharyngeal microbiome (an accessible, representative measure of airway bacteria) in 142 adult Hutterites, including 24 asthmatics, and FEV_1/FVC ratio as a measure of lung obstruction, which is significantly associated with asthma in this population. The Hutterites are a founder population of European decent who live communally, thereby ensuring that environmental exposures are relatively similar among individuals. We sequenced the V4 region of the 16S rRNA gene from bacterial cells collected from the nasopharynx and calculated the relative abundance of each bacterium after bioinformatic classification of QC filtered reads using Mothur. After subsampling a maximum of 550,000 reads from each individual (range 273,064-550,000; median 550,000), standardized bacterial reads present in at least 75% of individuals were fit to a standard normal distribution. At the taxonomic level of order, after regressing out age, sex and date of collection, the first principle component, which accounted for 44.4% of the variance between individuals, was associated with the FEV_1/FVC ratio ($p=0.008$). After removing correlated bacteria (Spearman correlation >0.8), we performed association tests for the relative abundance of the 60 bacteria in the taxonomic level of order using a linear mixed model as implemented in GEMMA, which incorporates information on relatedness into the model. Sex, age, and date of collection were included as covariates. The abundance of Enterobacteriales bacteria (gram-negative facultative anaerobes in the class Gammaproteobacteria of the phylum Proteobacteria) was positively correlated with FEV_1/FVC ratio after Bonferroni correction for 60 tests (uncorrected $p=5.3\times10^{-4}$). Our study demonstrates that the nasopharynx is an accessible airway site in which to study microbiome-lung disease associations and suggests that microbes in this anatomic location may reflect environmental exposures and/or host mucosal immunity status related to important disease-associated phenotypes.

930T

Hematopoeitic stem cells target neovascular tissue in a novel preclinical model of proliferative diabetic retinopathy. *K. Wert[1], V. Mahajan[2,3], Y. Yan[4], Y. Li[5,6], J. Tosi[5,6], T. Nagasaki[6], M. Grant[7], S. Tsang[5,6,8].* 1) Jaenisch Laboratory, Whitehead Institute for Biomedical Research, Boston, MA; 2) Department of Ophthalmology & Visual Sciences, University of Iowa, Iowa City, IA; 3) Omics Laboratory, University of Iowa, Iowa City, IA; 4) Department of Pharmacology and Therapeutics, University of Florida, Gainesville, FL; 5) Bernard & Shirlee Brown Glaucoma Laboratory, Columbia University, New York, NY; 6) Edward S. Harkness Eye Institute, Columbia University, New York, NY; 7) Eugene and Marilyn Glick Eye Institute, Department of Ophthalmology, Indiana University School of Medicine, Indianapolis, IN; 8) Department of Pathology & Cellular Biology, Columbia University, New York, NY.

Currently, more than 382 million people have been diagnosed with diabetes, and this number is expected to rise to 592 million people by 2035. Approximately one third of patients diagnosed with diabetes display signs of diabetic retinopathy (DR). Approximately 60% of non-proliferative DR patients progress to the neovascular stage, proliferative DR (PDR). PDR patients will develop vitreous hemorrhage, tractional retinal detachment, neovascular glaucoma and blindness. Most currently available mouse retinopathy models are inadequate for the study of PDR. They only display mild, non-proliferative disease within the lifespan of the animals. An animal model for PDR is critical to study signaling pathways involved in this disease stage. In this study, we examined protein levels from vitreous samples of human patients with PDR. We discovered an increase of hypoxia inducible factor 1 alpha (HIF1a) protein in untreated human PDR patients compared to control groups. During hypoxic conditions, von Hippel Lindau tumor suppressor protein (VHL) releases its normal degradative action on HIF1a. We tested whether the human PDR condition could be modeled by using a knockout of VHL, which would constitutively activate HIF1a and release its inhibition of HIF1a transcriptional activity. Therefore, we created a novel, tissue-specific preclinical model in which *Vhl* was knocked out within the neural retina using a *Ch×10 cre* promoter. This promoter allowed for a complete embryonic retinal loss of *Vhl* gene expression. We found that the mice developed severe proliferative retinopathy that correlated with human PDR. Since hematopoietic stem cells (HSCs) are found to localize to sites of angiogenesis, we hypothesized that HSCs delivered into the intravitreal space of the eye might localize to the sites of neovascularization, where angiogenesis is occurring and abnormal and leaking blood vessels are formed. So, the PDR mouse model was treated using transplantation of HSCs into the intravitreal space of the eye. We found that HSCs localized to the sites of damaged retinal vasculature after intravitreal injection into our novel preclinical PDR model. In summary, we developed a new mouse model for the study of human PDR and other ischemic retinal diseases by constitutive down regulation of HIF1 regulator, VHL. Rapid progression of retinopathy in these mutants will expedite the evaluation of therapeutic agents for this group of blinding disorders, including treatment using HSCs.

931S

The Role of Copy Number Variants in Latino Children with Asthma. *M.L. Spear[1, 2, 3], M. Pino-Yanes[2], C.R. Gignoux[3], C. Eng[2], B.P. Coe[4], C.D. Campbell[4], E.E. Eichler[4, 5], D.G. Torgerson[2], E.G. Burchard[2, 3] on behalf of the GALA I Investigators.* 1) Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA; 2) Department of Medicine, University of California, San Francisco, San Francisco, CA; 3) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA; 4) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA; 5) Howard Hughes Medical Institute.

Asthma is a complex respiratory disease influenced by social, environmental, and genetic factors. Twin studies have found the heritability to be between 75% and 92%, suggesting an important genetic contribution to asthma susceptibility. Although variation in over 150 genes has been associated with asthma, they only explain a small fraction of the disease heritability. Copy number variants (CNVs) are an important source of variation in the human genome due to their large size. CNVs have been previously associated with immune-related diseases, however knowledge of their role in asthma is limited. In this study we performed a genome-wide assessment of common CNVs in Latino children from the Genetics of Asthma in Latino Americans (GALA I) study, to identify novel CNVs associated with asthma. CNVs were called from genome-wide SNP genotypes obtained from the Affymetrix 6.0 GeneChip Array (Affymetrix, Santa Clara, Calif) using the Affymetrix Genotyping Console 4.1 software. A total of 733 Puerto Ricans and Mexicans, including 445 cases and 288 controls were included in the study. We limited our analysis to common CNVs present in more than 5% of our case/control population, and greater than 30 kb in length. We performed logistic regression adjusting by genetic ancestry and self-reported ethnicity, to test for an association between asthma and copy number of common CNVs. We identified a significant association between asthma and CNVs on chromosome 3q22.1 (OR = 0.85 [0.79-0.91]; p-value = 4.36×10^{-6}) following multiple testing correction by performing 10^6 permutations. The association was still significant after adjusting by genetic ancestry and ethnicity (OR = 0.85 [0.79-0.92]; p-value = 1.84×10^{-5}). The associated CNV is located within a genomic region previously associated with atopic dermatitis through linkage analysis, and is 300 kb upstream of *COL6A5*. Overall, our findings suggest an important role of a common CNV at 3q22.1 in asthma susceptibility in Latino children.

932M

Family-based Associations and Parent of Origin analyses reveal novel associations with inflammatory bowel disease (IBD). *X. YAN[1], M. DUBINSKY[2], T. HARITUNIANS[1], J. ROTTER[3], K. TAYLOR[3], S. RICH[4], S. Onengut-Gumuscu[4], W. CHEN[4], X. GUO[3], S. TARGEN[1], D. LI[1], D. MCGOVERN[1].* 1) F. Widjaja Foundation Inflammatory Bowel & Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 2) Pediatric IBD Center, Gastroenterology, Cedars-Sinai Medical Center, Los Angeles, CA; 3) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center; 4) Center for Public Health Genomics, Dept Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA.

Background: Only a minority of the heritability for IBD has been explained despite the identification of >160 susceptibility loci. We aimed to identify additional IBD associations by Transmission Disequilibrium Test (TDT) and to assess for any Parent-of-Origin effect to explain, in part, the missing heritability. **Methods:** 965 IBD trios with an affected offspring and both parents genotyped by ImmunoChip were included. We tested for association using TDT and Parent- of-Origin test (PLINK). Novel signals were confirmed by checking for LD and conditional analyses on the known 193 IBD-associated SNPs. An independent confirmation cohort of 2904 cases and 4791 controls was used to validate the TDT associations. 10e4~10e6 permutation and meta-analysis were performed.

Results: SNPs with **TDT** association after conditioning on known IBD SNPs (p<0.01 for SNPs within known region or p<0.0001 for SNPs outside known region, and p<5.0e-4 for conditional analyses) and confirmed in the case/control replication cohort are listed as below: *IL1RL1*(SNP=rs1420101, p_discovery=7.7e-3,p_confirmation=4.5e-5,p_meta=1.92e-5); *SLC2A13/LRRK2*(SNP=rs76904798,p_discovery=1.0e-3,p_confirmation=2.5e-7,p_meta=1.21e-7); *SBNO2*(SNP=rs8099951,p_discovery=6.4e-3,p_confirmation=3.4e-4,p_meta=1.4e-4); *LOC441108*(SNP=rs2158101,p_discovery=4.3e-4,p_confirmation=4.8e-5,p_meta=1.2e-5). **Parent-of-Origin** analyses identified differential transmission at the following genes, which mapped outside known IBD loci: *CTSH*(rs78764205,ptu=134.5:80.5,p_pat=2.3e-4,mtu=110.5:131.5,p_mat=0.18)(known MS, T1D locus); *MYCN*(rs4669018, ptu=151:202,p_pat=6.6e-3,mtu=208:178,p_mat=0.13); *TMEM16C*(rs1353142,ptu=80:113,p_pat=0.02,mtu=107:73,p_mat=0.01); *DLG2* (rs11233971, ptu=13:33,p_pat=0.003, mtu=45:25, p_mat=0.02); *TRPS1*(rs2293888,ptu=136:176,p_pat=0.02,mtu=186:143,p_mat=0.02); *RFX3/GLIS3*(rs13286706, ptu=179:137,p_pat=0.02,mtu=161:197,p_mat=0.06)(known T1D locus); *SAMD3/TMEM200A*(rs9492584, ptu=115.5:139.5,p_pat=0.13,mtu=167.5:119.5,p_mat=0.005); *NID2/PTGDR*(rs802971,ptu=145:103,p_pat=0.008, mtu=98:128,p_mat=0.05 (all p<1.0e-4, p/mtu - paternal/maternal transmitted : untransmitted counts, p_p/mat - paternal/maternal transmission p-value). **Discussion:** TDT analyses have identified additional IBD associations. The Parent-of-Origin analyses identified new IBD signals and if validated may explain some of the missing heritability in IBD.

933T

Parent-of-origin effects of the *APOB* gene on adiposity in young adults. *H. Hochner[1], C. Allard[2], E. Granot-Hershkovitz[1], J. Chen[3], C.M. Sitlani[4,5], S. Sazdovska[1], T. Lumley[6], B. McKnight[7], K. Rice[7], D.A. Enquobahrie[8], J.B. Meigs[9], P. Kwok[10,11,12], M.F. Hivert[13], I. Borecki[14], F. Gomez[14], T. Wang[14], C. van Duijn[15], N. Amin[15], J.I. Rotter[16], J. Stamatoyannopoulos[4,17], V. Meiner[18], O. Manor[1], J. Dupuis[19], Y. Friedlander[1], D.S. Siscovick[4,5,8].* 1) Braun School of Public Health, Hebrew University-Hadassah, Jerusalem, Israel; 2) Département de Mathématiques, Université de Sherbrooke and Centre de Recherche du Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, QC, Canada; 3) Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA; 4) Department of Medicine, University of Washington, Seattle, WA, USA; 5) Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA; 6) Department of Statistics, University of Auckland, Auckland, New Zealand; 7) Department of Biostatistics, University of Washington, Seattle, WA, USA; 8) Department of Epidemiology, University of Washington, Seattle, WA, USA; 9) Harvard Medical School and General Medicine Division, Massachusetts General Hospital, Boston, MA, USA; 10) Institute of Human Genetics, University of California, San Francisco, CA, USA; 11) Cardiovascular Research Institute, University of California, San Francisco, CA, USA; 12) Department of Dermatology, University of California, San Francisco, CA, USA; 13) Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical School, Boston, MA, USA; 14) Department of Genetics Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO, USA; 15) Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, University Medical Center, Rotterdam, the Netherlands; 16) Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; 17) Department of Genome Sciences, University of Washington, Seattle, WA, USA; 18) Department of Human Genetics, Hebrew University-Hadassah, Jerusalem, Israel; 19) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA.

Background: Genome-wide association studies (GWAS) of unrelated individuals have identified multiple loci associated with cardiometabolic phenotypes. However, the magnitude of these associations is small to moderate and the loci account for only a small proportion of the traits' heritability. To date, most association studies have not considered parent-of-origin effects (POEs). **Objectives:** We sought POEs on adiposity and glycemic traits in young adults. **Methods:** The Jerusalem Perinatal Study (JPS) sample comprises 1250 young adults (mean age 32) born in Jerusalem, and their mothers, with recently-collected DNA samples and cardiometabolic measurements. We focused on a set of 18 genes (182 tag SNPs) identified by previous GWAS as associated with selected cardiometabolic traits. Using linear regression, we examined the associations of maternally- and paternally-derived offspring minor alleles with body mass index (BMI), waist circumference (WC), fasting glucose and insulin. To replicate and meta-analyze findings, we performed similar analyses in individuals of European ancestry aged≤50 belonging to extended pedigrees from three additional studies: Framingham Heart Study, Family Heart Study and Erasmus Rucphen Family study (Total maximum N=4800). We considered $p<2.7\times10^{-4}$ statistically significant to account for multiple testing. Replicated findings were followed-up by also assessing POEs on lipids and blood pressure (BP). **Results:** A common coding variant in *APOB* (rs1367117) demonstrated a significant maternally-derived effect on BMI (β=0.8; 95%CI:0.4,1.1; $p=3.1\times10^{-5}$) and WC (β=2.7; 95%CI:1.7,3.7; $p=2.1\times10^{-7}$). The corresponding paternally-derived effects were non-significant (p>0.6). Suggestive maternally-derived associations of this *APOB* SNP were observed with fasting glucose (β=0.9; 95%CI:0.3,1.5; $p=4.0\times10^{-3}$) and insulin (ln-transformed, β=0.06; 95%CI:0.03,0.1; $p=7.4\times10^{-4}$). There was little evidence to suggest POE on lipids, yet maternal POE was observed for systolic BP (β=1.6; 95%CI:0.8,2.5; $p=1.6\times10^{-4}$). Maternal-specific associations with glucose, insulin and systolic BP were attenuated after further adjustment for BMI. **Conclusions:** Our findings demonstrate that a common genetic variation in *APOB* is associated with adiposity through maternal POE, an association that was not previously detected in GWAS of adiposity. These results provide support for a role of POEs in adiposity and related cardiometabolic traits, and motivate further research in this area.

934S

Heritability explained by common SNPs for dietary intake: genome-wide analysis in three US cohorts. *Q. Qi[1], A. Chu[2], J. Huang[3], J. Kang[2], G. Curhan[2,3], L. Pasquale[2], I. De Vivo[2,3], A. Chan[2,3], H. Choi[2,4], R. Tamimi[2,3], P. Ridker[2], D. Hunter[2,3], W. Willett[2,3], E. Rimm[2,3], D. Chaseman[2], F. Hu[2,3], L. Qi[2,3].* 1) Albert Einstein College of Medicine, Bronx, NY; 2) Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 3) Harvard School of Public Health, Boston, MA; 4) Boston University School of Medicine, Boston, MA.

Habitual dietary intake is a complex behavior that has been suggested to have a genetic component, with estimated heritability ranging from 8% to 70% in family and twin studies. However, genetic determinants of dietary intake are poorly understood and it is unknown how much heritability for dietary intake can be captured by common SNPs in genome. Using a linear mixed model to fit all SNPs simultaneously by GCTA software, we estimated narrow sense heritability for dietary intake explained by common SNPs (~300,000 to 700,000 SNPs in the genome) among 37,421 unrelated individuals of European ancestry from three US cohorts: the Nurses' Health Study, the Health Professionals Follow-up Study, and the Women's Genome Health Study. Dietary intakes of total energy, macronutrients (percentage of energy intake), and individual foods and beverages were assessed by almost identical validated semiquantitative food frequency questionnaires across three studies, and two major dietary patterns (prudent pattern and Western pattern) were derived using factor analysis. In the combined results from three studies by meta-analysis, small but significant proportions of variance for intakes of total energy (5.8 [95% CI 2.9, 8.7]%; P=0.0001), fat (5.9 [3.0, 8.8]%; P=0.0001), protein (4.8 [1.9, 7.7]%; P=0.001), and carbohydrate (3.4 [0.5, 6.3]%; P=0.02) were explained by all common SNPs. We also found that a number of food and beverage groups (egg, butter, margarine, dairy products, vegetables, whole grains, refined grains, condiments, coffee, alcohol, fruit juice, sugar-sweetened beverages, diet beverages, and water) had genetic influences, with significant heritability estimates ranging from 3.1 [0.2, 6.0] to 9.3 [6.4, 12.2]% (P=0.04 to 5.3×10-10). In addition, a Western dietary pattern (characterized by higher intakes of red and processed meats, sweets, desserts, French fries, and refined grains) showed the highest heritability estimate (10.3 [7.2, 13.4]%; P =4.4×10-10) among dietary intake variables. Our data provide evidence for moderate genetic influence on dietary intake. Genome-wide association studies are warranted to identify genetic variants associated with dietary intake.

935M

The role of STRs in shaping complex traits. *T. Willems[1,2], A. Gordon[1], M. Gymrek[1], Y. Erlich[1].* 1) The Whitehead Institute, Cambridge, MA; 2) Computational and Systems Biology, MIT, Cambridge, MA.

Despite the advent of whole-genome sequencing data and increasingly large sample sizes, genome-wide association studies have failed to fully uncover the genetic basis for the heritability of most complex traits. While various studies have proposed that this "missing heritability" may stem from rare SNPs or epistatic interactions, other variant classes may also be responsible. Recently, a handful of single-gene and genome-wide studies have identified short tandem repeats (STRs), sequences of DNA consisting of repeating patterns of 2-5 base pairs, as modulators of gene expression. To assess whether this class of variant may also influence complex traits, we genotyped ~700K STRs in over 1500 individuals using data from the UK10K project. In conjunction with a wide range of phenotypes, these genotypes were utilized to perform a genome-wide STR GWAS. The resulting analysis identified a handful of STRs strongly associated with a variety of phenotypes, even after controlling for tagging SNPs. In addition, we utilized linear mixed models to jointly assess the phenotypic variance explained by SNPs and STRs. In aggregate, our results shed new light on the issue of "missing heritability" and suggest that STR variations substantially contribute to complex traits.

936T

Regulatory variants explain much more heritability than coding variants across 11 common diseases. *A. Gusev[1], S.H. Lee[2], B.M. Neale[3,4], G. Trynka[5,6,7,8], B.J. Vilhjalmsson[1], H. Finucane[9], H. Xu[10], C. Zang[10], S. Ripke[3,4], E. Stahl[11], . SCZ-PGC[16], . SWE-SCZ[17], A.K. Kahler[12], C.M. Hultman[12], S.M. Purcell[3,4,11], S.A. McCarroll[4], M. Daly[3,4], B. Pasaniuc[13], P.F. Sullivan[14], N.R. Wray[2], S. Raychaudhuri[5,6,7,8,15], A.L. Price[1,8].* 1) Epidemiology, Harvard School of Public Health, Boston, MA; 2) The University of Queensland, Queensland Brain Institute, Brisbane, Queensland, Australia; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 4) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 5) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; 6) Division of Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; 7) Partners Center for Personalized Genetic Medicine, Boston, Massachusetts, USA; 8) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 9) Department of Mathematics, Massachusetts Institute of Technology, Massachusetts, USA; 10) Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts, USA; 11) The Department of Psychiatry at Mount Sinai School of Medicine, New York, New York, USA; 12) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 13) Department of Pathology and Laboratory Medicine, Geffen School of Medicine at UCLA, Los Angeles, California, USA; 14) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; 15) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK; 16) Schizophrenia Working Group of the Psychiatric Genomics Consortium; 17) Swedish Schizophrenia Consortium.

Common variants implicated by genome-wide association studies (GWAS) of complex diseases are known to be enriched for coding and regulatory variants. We applied methods to partition the heritability explained by genotyped SNPs (h^2_g) across functional categories (while accounting for shared variance due to linkage disequilibrium) to imputed genotype data for 11 common diseases in >60,000 samples. DNaseI Hypersensitivity Sites (DHS), identified in 218 cell-types, spanned 16% of the genome and explained an average of 79% of h^2_g (5.1× enrichment; $P<10^{-20}$); further significant enrichment was observed at enhancer DHS elements ($P=5\times10^{-4}$) and cell-type specific DHS regions with relevance to phenotype. The enrichments were larger and more statistically significant than in analyses that did not use imputed genotypes or were restricted to GWAS-associated loci. In contrast, coding variants, which span 1% of the genome, explained only 8% of h^2_g (13.8× enrichment; $P=5\times10^{-4}$). We showed by extensive simulations that these estimates are robust to diverse genetic architectures, and substantially more accurate than estimates from GWAS that do not account for correlation between SNPs and functional classes. To investigate the potential role of rare coding variation, we extended these analyses to an independent set of 2,500 schizophrenia cases and 3,900 controls genotyped on both GWAS and exome chips, replicating the DHS enrichment but observing no significant contribution from rare coding SNPs. Using our estimates of functional heritability to inform fine-mapping analysis substantially reduced the number of SNPs needed for follow-up. Overall, our results robustly quantify the contribution of functional features to common diseases, and highlight the value of analyzing components of heritability to unravel the functional architecture of common disease.

937S

Haplotypes explain additional heritability of complex traits. *G. Bhatia[1,2], A. Gusev[1,2], S.H. Lee[3,4], S. Ripke[5,6], E.A. Stahl[7], B.M. Neale[5,8], M.J. Daly[5,8], S.M. Purcell[6,7,9,10], N.R. Wray[3], N.A. Zaitlen[11], B. Pasaniuc[12], A.L. Price[1,2], Schizophrenia Working Group of the Psychiatric Genomics Consortium.* 1) Epidemiology, Harvard School of Public Health, Boston, MA; 2) Biostatistics, Harvard School of Public Health, Boston, MA; 3) Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia; 4) Queensland Institute of Medical Research, Brisbane, Queensland, Australia; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 6) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard University, Cambridge, Massachusetts, USA; 7) Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 8) Broad Institute of Harvard and MIT, Cambridge, MA 02142;; 9) Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA; 10) Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA; 11) Department of Medicine, Lung Biology Center, University of California, San Francisco, San Francisco, California, USA; 12) Department of Pathology and Laboratory Medicine, Geffen School of Medicine at UCLA, Los Angeles, California, USA.

While genome-wide significant associations generally explain only a small proportion of the narrow-sense heritability of complex traits (h^2), recent work has shown that the heritability explained by all genotyped SNPs (h^2_g) explains more of h^2 (Yang et al. 2010). However, much of the heritability is still missing ($h^2_g < h^2$). For example, for schizophrenia, h^2 is estimated at 0.7-0.8 but h^2_g is estimated at 0.2-0.3 (Lee et al. 2012; all values on liability scale). Efforts at increasing coverage through imputation have yielded only small increases in the heritability explained, possibly due to the relatively poor imputation of low-frequency causal variants. We propose to estimate the heritability explained by a set of haplotypes constructed directly from the study sample (h^2_{haplo}). Our method constructs a set of haplotypes by extending shared segments subject to a 4-gamete test using data phased with the HAPI-UR software package (Williams et al. 2012). The haplotype values (0, 1 or 2 copies per individual) are then used to estimate genetic relationships between individuals. These relationships are used as the covariance structure in a linear mixed model, which is used to estimate the heritability explained by this set of haplotypes. Our simulations show that this approach can yield substantial gains in heritability explained relative to genotyped SNPs, and produces robust estimates. We applied our method to WTCCC2 data consisting of 9K multiple sclerosis cases and 5K controls, genotyped at 452K SNPs. We built 37M haploSNPs with MAF > 1%. The heritability explained by this set of haploSNPs was $h^2_{haplo} = 0.39$ (s.e. 0.02), substantially larger than $h^2_g = 0.26$ in the same data. We also applied our method to PGC2 data consisting of 10K schizophrenia cases and 10K controls of European ancestry, all of which were genotyped at the same set of 348K SNPs. We built 28M haploSNPs with MAF > 1%. The heritability explained by this set of haploSNPs was $h^2_{haplo} = 0.38$ (s.e. 0.02), significantly larger than $h^2_g = 0.33$ in the same data. Overall, our results suggest that haplotypes can explain substantially more heritability than genotyped or imputed SNPs. Additionally, haplotype based approaches can shed light on the genetic architecture of complex traits and informing strategies for disease mapping.

938M

Heritability Estimates and Genetic Association for 60+ Complex Traits in a Young Healthy Sibling Cohort. *Q. Ma[1], A.B. Ozel[1], D. Siemieniak[2], K.C. Desch[3], D. Ginsburg[1,2,3,4], J.Z. Li[1].* 1) Departments of Human Genetics, University of Michigan, Ann Arbor, MI 48109; 2) Howard Hughes Medical Institute, Ann Arbor, MI 48109; 3) Pediatrics and Communicable Disease, University of Michigan, Ann Arbor, MI 48109; 4) Internal Medicine, University of Michigan, Ann Arbor, MI 48109.

As genotyping becomes more efficient, sample recruitment and phenotyping remain a major limiting factor. In a GWAS of bleeding and blood clotting traits we sought to increase the utility of the cohort by collecting > 60 self-reported complex traits through web-based questionnaires. The cohort of 1,191 healthy young subjects consists of 509 sibships, 80% Europeans, and age of 14-35 yrs. The traits include 16 quantitative traits (e.g., weight, height, age of menarche, hematological measures RBC, HCT, MCV, MCH, MCHC, RDW, WBC, HGB, PLT, MPV), 21 ordinal traits (e.g., Smoking, BleedingTendency, SkinTags, Acne, TanningTendency, SkinColor, Freckles, DentalCaries, VisionCorrection, EatingSweets, EatingSaltyfood, Athleticability, Aphthousulcers), and 27 nominal traits (e.g., Immunization, ToothExtraction, EyeColor, HairColor, Hairline, EarLobeCreased, EarLobeAttachment, Dimples, Dyslexia, Migraines, Stuttering, Allergies, Flatfeet, Handedness, PhoticSneeze, BrainFreeze, InterlockingFingers, etc.). We used the known relatedness to estimate heritability using Merlin-regress and found that >1/2 of the traits have heritability > 40%. Since the samples have been genotyped over ~800K SNPs in the original GWAS we used SNP data to calculate the actual genetic relatedness, and estimated the variance explained by all the genotyped SNPs using GCTA. With all subjects, pedigree-based estimates were similar to SNP-based estimates; but the latter were often reduced when we select one subject from each sibship to analyze the unrelated subsets. For many traits we identified common variants of significant association. This study demonstrates the feasibility of simultaneous analysis of dozens of traits via web-based profiling.

939T

Additive and epistatic effects of enhancer variants at GWAS risk loci. *O. Corradin[1], P.C. Scacheri[1,2].* 1) Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH; 2) Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH.

SNPs associated with common disease through Genome Wide Association Studies (GWAS) often lie in gene enhancer elements located distal to protein-coding genes. The enhancer-SNPs likely confer disease risk by impacting gene expression, but quantifying their effects on target gene expression has proven challenging. Here, through integrative analyses of haplotype block maps, epigenomic landscapes, and GWAS data from six different autoimmune disorders, we show that 94% of instances in which a GWAS SNP lies in an enhancer predicted to regulate a given gene, there is at least one additional SNP that lies in a different enhancer predicted to target the same gene and is located distal to the LD block from which the GWAS-association arose. Using RNA-seq data matched with genotype data from 400 individuals, we show that the distal enhancer variants act additively and epistatically with enhancer-SNPs within the GWAS-associated LD block to affect expression of the target gene. Thus for these loci, the transcriptional effect is dictated by SNPs located within and outside of the GWAS-association signal. Additionally, many of the outside variants disrupt transcription factor binding sites and affect chromatin accessibility. These results have important implications for GWAS and eQTL studies as we demonstrate that for some loci, the transcriptional effect is only apparent when the additive and epistatic effects of enhancer variants are considered. Given that outside variants can boost the transcriptional effect but are often not considered, we propose that the effects of GWAS SNPs are often underestimated, which could explain some of the missing heritability for these disorders. We propose utilizing the chromatin structure to identify epistatic interactions between GWAS SNPs, to interpret the effect of common disease alleles on gene expression, and to help refine the relative risk of enhancer variants on disease susceptibility.

940S

The genetics of exceptional human longevity: new clues from big data on disease. *K. Fortney[1], P. Sebastiani[2], T.T. Perls[3], S.K. Kim[1].* 1) Developmental Biology, Stanford University, Stanford, CA; 2) Biostatistics, Boston University School of Public Health, Boston, MA; 3) Medicine, Boston University School of Medicine, Boston, MA.

In order to begin to decipher the genetic basis of human longevity, we are taking a big data approach to scour large genome-wide association (GWA) datasets on age-related disease to narrow the search for new SNPs associated with longevity. Rather than search for longevity loci directly in centenarian GWAS, we can leverage data from many age-related disease GWAS (often with > 10,000 cases each) by assuming that individuals with extreme longevity are likely to be depleted of polymorphisms associated with disease, such as Alzheimer disease, type 2 diabetes or cardiovascular disease.

So far, we have acquired data from eight large disease GWAS meta-analyses: Cardiovascular disease, diabetes, late-onset Alzheimer disease, age-related macular degeneration, chronic kidney disease, lung cancer, pancreatic cancer, and rheumatoid arthritis. In total, this preliminary dataset contains > 250,000 SNPs for > 150,000 cases of 8 diseases. We combined these data with the results of a GWAS on exceptional longevity that included 801 subjects enrolled in the New England Centenarian Study (Sebastiani et al. 2012). We applied a new statistical method to identify variants that are enriched in people with disease but depleted in centenarians. In this approach, a strong association signal in disease can boost a weak association for longevity to statistical significance. Our approach is potentially a powerful new method to identify longevity loci. We will validate all the candidate SNPs that we identify in independent cohorts of centenarians, to test whether they are robustly associated with human longevity.

941M

A meta-analysis of genome-wide association scans for nevus count reveals PPARGC1B as affecting both moliness and melanoma risk. *N.G. Martin[1], D.L. Duffy[1], G. Zhu[1], G.W. Montgomery[1], N.K. Hayward[1], M. Falchi[2], P. Hysi[2], V. Bataille[2], T.D. Spector[2], G.C. Smith[3], D.M. Evans[3], H. Nan[4], J. Han[4], S. Chanock[4], D. Hunter[4], L. Jacobs[5], T.E. Nijsten[5], F. Liu[5], M.H. Kayser[5].* 1) Gen Epidemiology, Queensland Inst Med Res, Brisbane, QLD, Queensland, Australia; 2) Dept of Twin Research & Genetic Epidemiology, St Thomas Hospital Campus, Kings College, London, SE1 7EH, UK; 3) Department of Social Medicine, University of Bristol, Bristol BS82PS, UK; 4) Channing Laboratory, Harvard Medical School, Boston, MA 02115, USA; 5) Erasmus MC, University Medical Centre, Rotterdam, The Netherlands.

The main clinical interest in total number of acquired melanocytic nevus on the skin is its strong correlation with melanoma risk. We have previously described genome-wide association scans (GWAS) in Australian and British samples that identified loci influencing both nevus count and melanoma. Here we report a meta-analysis of five nevus GWAS from Australia, the Netherlands, United Kingdom (two studies), and the United States, comprising a total of 23,371 phenotyped individuals. We confirm known loci including MTAP, PLA2G6, IRF4 and ASIP and refine the location of the causal variants. Several new regions exhibit suggestive evidence ($P \sim 10^{-6}$-10^{-7}). One notable SNP, rs251464, is in the PPARGC1B gene, previously shown to modify tanning ability via a regulatory effect on MITF, the master melanocyte regulatory gene [Shoag et al Mol Cell. 2013, 49:145-57]. Combining the results for this SNP ($P=8\times10^{-6}$) with results from a melanoma case-control analysis ($P=4.6\times10^{-4}$) reaches a genome-wide significant level and direction of effect was consistent in both phenotypes, with the C allele decreasing both melanoma risk and mole count. Another SNP in PPARGC1B, rs32579, previously flagged as affecting tanning, was shown to not affect nevus count (P=0.19). We now plan to meta-analyse these results with those from a parallel GWAS meta-analysis of melanoma to achieve even greater resolving power for variants influencing early stage oncogenesis.

942T

Identification of 4 novel susceptibility *loci* for intracranial aneurysms in Portuguese using a pooling-based GWAS. *P.CS. Abrantes[1, 2], M.M. Santos[3], J.M. Xavier[1, 2], I. Sousa[1, 2], V. Francisco[1, 2], T. Krug[1, 2], J. Sobral[1, 2], M. Matos[1, 2], A. Jacinto[1, 2, 4], D. Coiteiro[3], S.A. Oliveira[1, 2].* 1) Instituto Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal; 2) Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal; 3) Department of Neurosurgery, Hospital de Santa Maria, Lisboa, Portugal; 4) Centro de Estudos de Doenças Crónicas (CEDOC), Faculdade de Ciências Médicas, Universidade Nova de Lisboa, 1169-056 Lisboa, Portugal.

Subarachnoid hemorrhage is a life-threatening event that most frequently leads to severe disability and death. Its most frequent cause is the rupture of an intracranial aneurysm (IA). Several studies have consistently demonstrated a genetic component to the risk for IA but until now, its exact etiology remains uncertain.

To identify new susceptibility *loci* for IA, we performed a genome-wide association study (GWAS) using DNA pooling approach on a Portuguese dataset. Pools of 100 IA cases and 92 age- and gender-matched controls were allelotyped in triplicate using the Affymetrix Human SNP Array 6.0. Top SNPs with |RAS diff| ≥13.0% were selected for individual genotyping in the same GWAS dataset. Ninety nine of the 101 SNPs successfully genotyped were technically validated (P<0.05). Replication of validated SNPs was conducted in an independent Portuguese dataset of 100 IA cases and 407 controls.

We identified 4 variants (rs4667622, rs6599001, rs3932338 and rs10943471) associated with IA in both the discovery and replication datasets (individually and in combination). SNPs rs4667622 (*P*combined=4.00E-05, OR_G[95% CI]=1.75[1.33-2.33]) and rs6599001 (*P*combined=2.20E-04, OR_C[95% CI]=2.00[1.39-2.88]) were the most significant. rs4667622 is located on chromosome 2q31.1 within the regulatory region of the myosin IIIB gene (*MYO3B*) and rs6599001 maps to chromosome 3p22.2 upstream of the WD repeat domain 48 gene (*WDR48*). The third most significant genetic association was found with rs10943471 (*P*combined=3.20E-04, OR_G[95% CI]=1.81[1.31-2.51]) which is located on chromosome 6q14.1 in an intergenic region upstream of *HTR1B* (5-hydroxytryptamine (serotonin) receptor 1B). The rs3932338 SNP (*P*combined=1.29E-03, OR_A[95% CI]= 1.59[1.19-2.08]) is located on chromosome 5p14.2 in a gene desert. Additionally, we replicated the previously described association with IA of rs1333040 at the 9p21.3 genomic region (*P*combined=1.93E-02, OR_T[95% CI]=1.41[1.05-1.89]), thus validating our dataset.

Our novel findings in the Portuguese population must be replicated in further datasets to establish their pathogenic role in IA formation.

943S

A GWAS of Risk Genes for Birth of a Child With Down Syndrome. *E. Feingold[1], Z. Zeng[1], E.G. Allen[2], D. Ramachandran[2], M.E. Zwick[2], S.L. Sherman[2].* 1) Dept Human Gen, Univ Pittsburgh, Pittsburgh, PA; 2) Dept Human Gen, Emory University, Atlanta, GA.

We have conducted the first GWAS for risk of having a child with Down syndrome (DS) in approximately 700 mothers of children with maternally-derived free trisomy 21. Mothers, fathers, and children were genotyped at the Center for Inherited Disease Research (CIDR) on the Illumina OmniExpress. Genotypes for chromosome 21 in children were called using methods we previously developed. Genotypes for parents and child were then used to establish parent of origin, stage (meiosis I or meiosis II) and meiotic recombination patterns for each child. We used two different designs for the GWAS. The first uses the fathers as controls. In the second, we take advantage of the fact that the etiologies of meiosis I trisomy 21 and meiosis II trisomy 21 appear to be quite different, and perform a GWAS of mothers of meiosis I cases vs. mothers of meiosis II cases. This approach cannot find variants that are common risk factors for both meiosis I and meiosis II errors, but is a well-controlled design for finding variants that are unique to one or the other. Results include genes that are known to affect meiotic recombination, which is not unexpected since the link between altered recombination and nondisjunction is well established both in humans and experimental organisms.

944M

Genome-wide Association Study, Meta-Analysis and Linkage Study of Gamma-Prime Fibrinogen Plasma Levels in a Healthy Young Cohort. *A. Ozel[1], K. Desch[2], D. Siemieniak[3], D. Ginsburg[1,3], J.Z. Li[1].* 1) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 2) Department of Pediatrics, University of Michigan, Ann Arbor, MI; 3) HHMI, University of Michigan, Ann Arbor, MI.

Gamma-prime fibrinogen (GPF) is one of the isoforms of fibrinogen and a component of the blood-clotting complex, making up ~10% of the total fibrinogen levels in blood. Its role in cardiovascular risk prediction has been recently proposed. Plasma levels of fibrinogen are highly variable among healthy people, with ~50% of their variance attributable to inherited factors. We performed genome-wide association studies in a healthy sibling cohort of 1,152 subjects, focusing on the European subset (n = 940), and a second healthy cohort of 2,304 individuals (Desch et al., 2012). Plasma GPF levels were determined using a monoclonal antibody specific to this isoform of fibrinogen. Heritability (h^2) was estimated as 61.6% (using *Merlin-Regress*) and was consistent with results from *GCTA* (63.3%) and intra-class correlation (65.8%). Common variants at or near the fibrinogen gene cluster on chromosome 4 showed significant association with GPF plasma levels (P < 5.0E-8), explaining 22.6% of GPF variance. The top SNP was rs7654093(T) (MAF = 0.23, P = 4.9E-72) at ~19 kb upstream of the *FGG* (Fibrinogen Gamma Chain). Suggestive evidence of association was found at 8p23.2 (lead SNP: rs70051128(T) with P = 7.9E-7 near the *MCPH1* gene). Meta-analysis of the two cohorts confirmed the signals on chromosome 4, with best p-value of 4.9E-199 and the direction of the effects of the top SNPs were in strong agreement. Linkage analysis using the sibling subset of the two cohorts identified significant signals at a ~40 cM interval on 4q28.3-q34.2 (LOD=5.25, permutation P < 0.05) including the fibrinogen gene cluster, and on a novel locus: a ~37 cM interval on 8p23.3-p12 (LOD=2.00, permutation P = 0.03). The top peak in the linkage analysis explained 33.9% of the variation in the GPF levels. To our knowledge, this is the first genome-wide study investigating gamma prime fibrinogen level variation in a healthy population. These results provide new insight into the regulation of hemostasis, may identify novel genetic modifiers of bleeding and thrombosis risk, and suggest that linkage analysis of blood clotting traits may reveal novel loci supported by association studies.

945T

SORBS1 gene, a new candidate for diabetic kidney disease: results from a multi-stage genome wide association study. *M. Germain[1,2,3], M. Pelozzeli[4], N. Sandholm[5,6,7], AJ. MacKnight[8], K. Susztak[9], M. Lajer[10], C. Forsblom[5,6], M. Marre[11,12,13,14], HH. Parving[10,15], P. Rossing[10,15], I. Toppila[5,6], J. Skupien[4,16], R. Roussel[11,12,13,14], YA. Ko[9], N. Ledo[9], L. Folkersen[17], M. Civelek[18], AP. Maxwell[8,19], DA. Trégouët[1,2,3], PH. Groop[5,6,20], L. Tarnow[10,15,21], S. Hadjadj[22,23,24].* 1) UMR_S 1166, INSERM, Paris, France; 2) Sorbonne Université, UPMC Univ Paris 06, Paris, France; 3) ICAN Institute for Cardiometabolism and Nutrition, F-75013, Paris, France; 4) Research Division, Joslin Diabetes Center & Department of Medicine, Harvard Medical School, Boston,Massachusetts; 5) Folkhälsan Institute of Genetics, Folkhälsan Research Center, Biomedicum Helsinki, Helsinki, Finland; 6) Division of Nephrology, Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland; 7) Department of Biomedical Engineering and Computational Science, Aalto University, Espoo, Finland; 8) Nephrology Research, Centre for Public Health, Queen's University of Belfast, Belfast, United Kingdom; 9) University of Pennsylvania; Renal Electrolyte and Hypertension Division; Philadelphia, PA; 10) Steno Diabetes Center, Gentofte, Denmark; 11) Université Paris-Diderot, Sorbonne Paris-Cité, France; 12) Diabetologie, AP-HP, Hôpital Bichat, Paris, France; 13) Departement Hospitalo-Universitaire FIRE, Paris, France; 14) INSERM U-872, Paris, France; 15) Faculty of Health Sciences, University of Aarhus, Aarhus, Denmark; 16) Department of Metabolic Diseases, Jagiellonian University, Krakow, Poland; 17) Department of PharmacoGenetics, Novo Nordisk Park 9.1.21, Copenhagen, Denmark; 18) Department of medicine, University of California Los Angeles, Los Angeles; 19) Regional Nephrology Unit, Belfast City Hospital, Belfast, United Kingdom; 20) Baker IDI Heart and Diabetes Institute, Melbourne, Australia; 21) Nordsjaellands HospitalAarhus, Denmark; Faculty of Health Sciences, University of Aarhus, Aarhus, Denmark; 22) Université de Poitiers, UFR MédecinePharmacie, Centre d'Investigationclinique, Poitiers, France; 23) CHU de Poitiers, Service d'Endocrinologie-Diabetologie and Centre d'Investigationclinique,Poitiers, France; 24) INSERM, CIC1402and U1082, Poitiers, France.

In order to identify novel susceptibility genes for diabetic nephropathy (DN), we performed a genome-wide association study using 1,000 Genomes-based imputation in individuals with type 1 diabetes. The discovery cohort comprised 683 proteinuric patients with or without renal failure (cases) and 779 patients with diabetes for more than 15 years duration and no evidence of renal disease (controls). None of the single nucleotide polymorphisms (SNPs) tested reached genome-wide statistical significance. The top forty-six SNPs with p-value < 10^{-5} were brought for initial replication in 820 cases and 885 controls part of the US GokinD study. Two SNPs in strong linkage disequilibrium with each other, located in the SORBS1 gene, were consistently and significantly associated with DN ($p = 7.87\ 10^{-6}$ in the discovery, $p = 1.32\ 10^{-4}$ in the initial replication). In the combined samples, the minor allele of the candidate SNP was less frequent in cases than in controls (37% vs 45%) and associated with a decreased risk for DN of OR = 0.74 [0.67 - 0.82]. However, this association was not observed in a second stage DN cohorts (FINNDIANE and UK-ROI) composed of 2,142 cases and 2,494 controls (p = 0.295). We ultimately interrogated transcriptomic analysis of human kidney tissue from type 2 diabetes kidney disease patients and non-diabetic controls. In micro-dissected glomerular (n=17) and tubule (n= 39) samples, SORBS1 showed significant over-expression in tubule samples in diabetic kidney disease versus non-diabetic controls. In addition, SORBS1 expression was significantly and inversely correlated with estimated glomerular filtration rate (eGFR) reflecting kidney function. Altogether, these data suggest that SORBS1 might be a new gene involved in DN.

946S

Polymorphism upstream of cryopyrin gene (*NLRP3*) is associated with severe retinopathy in type 1 diabetes. *S. Hosseini[1], K. Howard[2], L. Sun[3], A.P. Boright[4], D.A. Tregouet[5], N. Sandholm[6], K. Hietala[6], I. Toppila[6], M.S. Lajer[7], M. Marre[8], P. Rossing[7], P.H. Groop[6], A.J. Canty[9], S. Hadjadj[10], B.E. Klein[2], S.B. Bull[3,11], R. Klein[2], A.D. Paterson[1,3], the DCCT/EDIC Research Group.* 1) Genetics & Genomic Biology Program, Hospital for Sick Children Research Institute, Toronto, ON, Canada; 2) Dept of Ophthalmology & Visual Sciences, Univ Wisconsin, Madison, WI, USA; 3) Dalla Lana School of Public Health, Univ Toronto, Toronto, Canada; 4) Dept of Medicine, Univ Toronto and LMC Diabetes & Endocrinology, Toronto, Canada; 5) Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1166 and INSERM, UMR_S 1166 and ICAN Institute for Cardiometabolism and Nutrition, F-75013, Paris, France; 6) Folkhälsan Inst of Genetics, Univ of Helsinki, Helsinki, Finland; 7) Steno Diabetes Center, Gentofte, Denmark; 8) INSERM U695 & Dept of Endocrinology, Bichat-Claude Bernard Univ Hosp, Paris, France; 9) Dept of Mathematics & Statistics, McMaster Univ, Hamilton, Canada; 10) Dept of Endocrinology & Diabetology, Univ of Poitiers, Poitiers, France; 11) Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Canada.

Background: Diabetic retinopathy (DR) is a leading cause of blindness worldwide. Despite strong effects of diabetes duration and glycemic control, several lines of evidence suggest a genetic contribution to the risk, especially for severe DR (SDR defined as severe non-proliferative DR or worse). However, no genetic variant has shown convincing association with DR in genome-wide association studies (GWAS). **Purpose:** To identify common polymorphisms associated with SDR **Methods:** White subjects with type 1 diabetes participating in the Epidemiology of Diabetes Interventions and Complications (EDIC, n=1304) and Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR, n=603) were genotyped by Illumina BeadChip assays. Genotypes of ~2.5M autosomal SNPs from phase 2 and 3 of HapMap were imputed in both studies. SDR status was defined based on fundus photographs at the most recent follow-up visit [level 53/<53 or worse on ETDRS scale or panretinal laser photocoagulation treatment]. Multivariable association analysis of SDR with additive genotype by logistic regression accounted for covariates, including glycemia (measured by A1C) and diabetes duration. GWAS were first performed separately in EDIC divided into 3 subgroups (primary cohort, secondary cohort by treatment group - conventional or intensive) and WESDR; results were combined in a fixed effects meta-analysis. Top association signals (39 loci) were evaluated for replication in three independent white T1D cohorts: Genesis-GeneDiab (n=502), Steno (n=936), FinnDiane (n=2194). **Results:** Among the top hits short of genome-wide significance, an intergenic SNP 27 kb upstream of NLRP3 associated with SDR (P=8E-6) was replicated (P<0.001 accounting for multiple testing). The index SNP at this locus showed similar direction of effect in all the examined populations. In the combined meta-analysis of the discovery and replication studies, each copy of the risk allele increased the odds of SDR by 25% (OR=1.26, 95%CI: 1.15-1.38, P=6E-7). NLRP3 is a member of the inflammasome complex which serves as a platform for the activation of caspase-1 in immune response. SNPs within NLRP3 (not in strong LD with our index) have been associated with Crohn's disease in candidate gene association studies and with C-reactive protein and fibrinogen, two inflammatory markers, in GWAS. **Conclusion:** NLRP3 is the first replicated locus with consistently strong evidence for association with SDR in a meta-GWAS.

947M

Genome-wide meta-analysis identifies novel variants associated with fasting plasma glucose in East Asians. *J. Hwang[1], Y. Cho[2], B. Han[1], B. Kim[1].* 1) KNIH, Osong, South Korea; 2) Department of Biomedical Science, Hallym University, Chuncheon, Korea.

Fasting plasma glucose (FPG) has been recognized as an important indicator for the overall glycemic state preceding the onset of metabolic diseases. Most genome-wide association loci for FPG so far been identified were derived from populations with European ancestry with a few exceptions. To extend a thorough catalog for FPG loci, we conducted meta-analyses of 13 genome-wide association studies in up to 24,740 non-diabetic subjects with East Asian ancestry. Follow-up replication analyses in up to additional 21,345 participants identified three new FPG loci reaching genome-wide significance in or near PDK1-RAPGEF4, KANK1 and IGF1R. Our results could provide additional insight into the genetic variation implicated in fasting glucose regulation.

948T

A Genome-wide Association Study of Apnea-Hypopnea Index in Children with Obstructive Breathing. *r. pellegrino[1,2], e. byrne[1], l. brooks[3], a. pack[2], h. hakonarson[1].* 1) Genetics, CHOP, Philadelphia, PA; 2) Center for Sleep and Circadian Neurobiology, Upenn, Philadelphia, PA; 3) Pulmonary, CHOP, Philadelphia, PA.

Obstructive sleep apnea syndrome (OSAS) is a complex sleep disorder that imposes a large burden on our society in terms of morbidity, quality of life, and healthcare costs. Childhood OSAS is characterized by habitual snoring, disturbed sleep and problems with daytime neurobehavioral functioning. OSAS appears to result from diverse gene-gene interactions and association with environment changes. There have been no published GWAS in children for OSAS. The identification of genetic variants associated with increased risk for OSAS could potentially translate into earlier recognition and treatment with reduced morbidity, and may also serve to identify potential targets for novel therapies. Here, we present a genome-wide association study of the Apnea-Hypopnea Index measured in a cohort of children referred to the sleep clinic at the Children's Hospital of Philadelphia for suspected obstructive breathing. Our aim was to identify common genetic variants that increase OSAS severity in children with sleep difficulties. A total of 2,473 children participated in the sleep study. The primary reason for referral was for suspected OSA and they had a Polysomnography (PSG) exam performed. 1,782 children in the sample were given a potential diagnosis of OSA prior to the sleep study. Also, a total of 1201 children were listed as having at least one mental disorder. All participants were genotyped using either the Illumina HumanHap550 or 610 Quad arrays. The association between the natural log of the Apnea-Hypopnea Index (AHI) and SNP was assessed using a linear model in PLINK. Principal components for each individual were calculated using GCTA, and the first 5 PCs were fitted as covariates in the analysis along with age, gender, BMI and total sleep recording time. No SNPs in the study passed the threshold for genome-wide significance ($p < 5 \times 10\text{-}8$), either within the ethnic groups separately, or in the meta-analysis of both groups. However, many of the marginally significant SNPs are located in or near genes that have shown evidence of association with related traits such as waist circumference and C-reactive protein, suggesting they may become useful biomarkers as we grow our sample size.

949S

The Association between Genetic Markers for Type 2 Diabetes and Carnitines: A Replication of Adult Findings in a Neonatal Population. *CJ. Smith[1], SL. Berberich[2], JC. Murray[3], KK. Ryckman[1].* 1) Department of Epidemiology, University of Iowa, Iowa City, IA; 2) State Hygienic Laboratory, University of Iowa, Iowa City, IA; 3) Department of Pediatrics, University of Iowa, Iowa City, IA.

Background: Utilizing metabolomic technologies, such as tandem mass spectrometry, is of growing interest for understanding the etiology of complex diseases, such as Type 2 Diabetes. Recent genome-wide association studies (GWAS) of the adult human metabolome have identified genetic variants associated with biomarkers that are important in Type 2 Diabetes (T2DM) etiology; however, no studies to our knowledge have examined the relationship between these disease-associated genetic markers at birth and the implications these associations would have for earlier disease prediction. Objective: We examined single nucleotide polymorphisms (SNPs) identified by GWAS to be associated with Type 2 Diabetes, in addition to genes associated with fatty acid metabolism, and their associations with neonatal metabolite levels. **Methods:** Our analysis included genetic data and analyte measurements for 832 singleton Iowa neonates. We examined analytes measured by tandem mass spectrometry as part of routine newborn screening. Sixty-eight SNPs in forty-six genes were genotyped using the Fluidigm® SNP Genotyping platform. We analyzed each SNP-analyte level for association using Kruskal-Wallis tests. **Results:** Seven SNP-analyte associations were significant by Bonferroni correction (P value <6.5×10-5). The CC genotype of rs2014355 in ACADS was significantly associated with higher levels of C4, C4/C2, and C4/C3, with p-values of 4.63E-34, 2.99E-34, and 7.96E-30, respectively. The CC genotype of rs211718 in ACADM was significantly associated with higher levels of C6, C8 and C10 and lower levels of C5DC/C8, with p-values of 2.26E-19, 8.38E-14, 2.42E-08, and 2.86E-15, respectively. **Conclusions:** The C allele of rs2014355 has previously been shown to be associated with decreased insulin sensitivity, and higher C4 levels have been observed in individuals with T2DM compared to those without the disease. The C allele of rs211718 has previously been shown to be associated with increased C10 in adults, and C6 and C8 have been found to be significantly elevated in subjects with T2DM. We have replicated adult SNP-analyte associations in a neonatal population. Capitalizing on the potential of neonatal genetic and metabolic profiles to predict risk for later life chronic conditions will be of paramount importance for personalized medicine, and may provide insight into the pathogenesis of complex metabolic disorders.

950M

A genome-wide association analysis of scarring trachoma in rural Gambia. *C.S. Franklin[1], C.h Roberts[2], S.E. Burr[3], F. Payne[1], W. Bottomley[1], S. Molina-Gonzalez[2], A. Natividad[2], H. Joof[3], P. Makalo[3], N. Faal[3], I. Sarr[3], M.J. Burton[2], T. Clark[2], K. Rockett[4], D. Kwiatkowski[1], A. Sillah[5], R.L. Bailey[2], D.C.W. Mabey[2], I. Barroso[1], M.J. Holland[2].* 1) Wellcome Trust Sanger Institute, Cambridge, Cambridgeshire, United Kingdom; 2) London School of Hygiene and Tropical Medicine, Keppel Street, Bloomsbury, London, United Kingdom; 3) Medical Research Council Unit, Atlantic Blvd, Serrekunda, Gambia; 4) Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, United Kingdom; 5) NEHP, Ministry of Health, The Gambia.

Trachoma is the most common infectious cause of blindness worldwide and is a neglected tropical disease that is endemic across much of sub-Saharan Africa. The active form of disease is a severe conjunctivitis that is caused by infection with Chlamydia trachomatis (Ct). The host response to repeated cycles of infection can lead to the formation of trachomatous scarring (TS) on the upper tarsal surface of the eyelid, causing it to deform. This can progress to trachomatous trichiasis (TT), where the eyelashes turn inwards, scratch the cornea resulting in opacity, visual impairment and blindness. Within endemic populations a variable fraction of persons develop TS or TT. Pathology in trachoma appears to be largely due to the host immune response, which suggests an important role for host diversity. Heritability of the immune response to Ct infection has been estimated at 0.39 through twin studies. Environmental risk factors for (TS) are significant and well characterized but the familial clustering of disease also suggests that there is a significant genetic component. We have conducted the first genome-wide association study of TS, based on a Gambian sample of 1090 TS cases and 1531 controls. Genotyping was performed using the Illumina 2.5M chip with subsequent imputation against the 1000 genomes global reference panel. Mixed model association was performed using EMMAX and GEMMA to account for population stratification and relatedness of samples. Although none of the variants reached the genome-wide significance threshold of $p<2.5\times10\text{-}8$, we found several suggestive association signals located in interesting candidate genes, which mapped to pathways involved in the cell cycle, microtubule organisation, cell-cell interactions and the immune response. Our strongest individual association is located within the PREX2 gene, which may have a role in mediating chlamydial cell entry to the host cell. Replication of these putative signals will be attempted in an on-going collection of 2,000 cases and 2,000 matched controls from a Trachoma endemic region in Tanzania.

951T

Genetic variation predicts serum lycopene concentrations in multi-ethnic population of postmenopausal women. *N. Zubair[1], U. Peters[1], C. Kooperberg[1], D. Chongzhi[1], L. Jingmin[1], C. Hutter[2], M. Neuhouser[1], Women's Health Initiative (WHI).* 1) Fred Hutchinson Cancer Research Center, Seattle, WA; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Background: While the human function of lycopene remains undetermined, this nutrient has been both positively and inversely associated with the risk of several chronic diseases. The inconsistencies in lycopene-disease association studies may stem from a lack of knowledge about the genetic variation in the synthesis, metabolism, and deposition of transport and binding proteins, which potentially influence serum lycopene concentrations. Objective: Here we examined the association between variation across the genome and serum concentrations of lycopene in a multi-ethnic population. Design: Participants included African (n= 914), Hispanic (n= 464), and European American (n=1,203) postmenopausal women from the Women's Health Initiative (WHI). We analyzed ~7 million single nucleotide polymorphisms (SNPs). Linear regression models were used to assess associations between each SNP and serum concentrations (log-transformed, continuous) of lycopene; we adjusted for age, BMI, and population substructure. Models were run separately by ethnicity and then combined in a trans-ethnic fixed effects meta-analysis. Results: In the meta-analysis, the SCARB1 gene, which encodes for a plasma membrane receptor for high-density lipoprotein cholesterol, significantly associated with lycopene concentrations ($rs1672879$, P-meta $< 2.68 \times 10\text{-}9$). Here each additional G allele resulted in a 12% decrease in lycopene levels for African Americans, 20% decrease for Hispanic Americans, and a 9% decrease for European Americans. In addition, two regions significantly associated with serum lycopene levels in African Americans: the SLIT3 gene, which serves as a molecular guidance cue in cellular migration, and the DHRS2 gene, which codes for an oxidoreductase that mitigates the breakdown of steroids. Conclusions: We found three novel loci associated with serum lycopene concentrations, two of which were specific to African Americans. Future functional studies looking at these specific genes may help provide insight into the metabolism and underlying function of lycopene in humans, which may help further elucidate lycopene's influence on disease risk and health.

952S

Systematic genome-wide microbiome association analysis in the Northern-German population identifies genetic variation that impacts microbial diversity in the gut. *A. Franke[1], J. Skiecieviciene[1], J. Wang[2], P. Rausch[2], U. Noethlings[3], W. Lieb[4], T.H. Karlsen[5], J. Baines[2], The PopGen Microbiome Study Group.* 1) Inst Clinical Molec Biol, University of Kiel, Kiel, Germany; 2) Max-Planck Institute of Evolutionary Biology, Ploen, Germany; 3) Institut für Ernährungs- und Lebensmittelwissenschaften (IEL), Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany; 4) Institute of Epidemiology, The PopGen Biobank, University of Kiel, Kiel, Germany; 5) NoPSC Study Group, Rikshospitalet, University of Oslo, Oslo, Norway.

The bacteria of the gut are a community that has co-evolved with the host and confers beneficial effects on human physiology and nutrition, and is crucial for human health. Increasing evidence suggests that the host's genetic variation might influence bacterial composition in addition to other direct (secretions of bile and defensins, control of gut motility etc.) or indirect (food and lifestyle) factors. An effect of host genotype on bacterial composition has been demonstrated by our group and others for single genes (MYD88, NOD2, TLRs, immunoglobin A (IgA)), and the human leukocyte antigen (HLA) region. In order to systematically evaluate the relationship between variation in the human genome to variation in the human microbiome, large-scale GWmAS (genome-wide microbial association studies) studies of population-representative cohorts are needed. Here, we performed a GWmAS of a cross-sectional, population-representative study cohort of Northern Germany (n=757). Ultra-dense SNP genotype information, available from different genotyping platforms, detailed phenotypic and dietary data, 16S rDNA microbiome data (V1-V2 region, 454 Roche platform) were generated for this Northern German cohort. Systematic association studies were performed for alpha diversity, beta diversity and core measurable microbiome measures. Likely due to the lack of statistical power and high inter-individual variability in the microbiome composition none of the detected associations bypassed traditional genome-wide significance (P-value smaller than 5×10E-8). In order to validated the best findings of the screening study, an independent study cohort was recruited from the same geographic region within Northern Germany (n=866). Again, dense SNP genotype data, phenotypic, dietary, and 16S rDNA microbiome data (again V1-V2 region, MiSeq platform; different DNA extraction method), analogous to the discovery cohort, were collected for the replication cohort. Our benchmarking study revealed that only around 6% variability was introduced by the sequencing method. We detected that genetic variation of the host contributes more than 4% to the bacterial community structure. Moreover, we managed to replicate several genetic loci that are implicated in determining microbial diversity.

953M

Identification of Susceptibility Loci for Crohn's Disease in Koreans through Immunochip. *M. Hong[1], H. Choi[1], W.J. Yun[1], J. Baek[1], T. Haritunians[2], S.H. Park[3], B.D. Ye[3], J. Liu[4], D.P.B. McGovern[2], S. Yang[3], K. Song[1].* 1) Dept. Biochemistry and Molecular Biology, Ulsan university College of Medicine, Seoul, Seoul, South Korea; 2) The F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute; Cedars-Sinai Medical Center, Los Angeles, USA; 3) Dept Internal Medicine, Univ Ulsan Colleage of Medicine, Seoul, Korea; 4) Genome Institute of Singapore, Singapore.

We have performed an immunochip study in 1,670 Korean patients with Crohn's disease and 1,438 controls. We confirmed six previously reported loci in Caucasians: GPR35 at 2q37(rs3749172; P = 5.30 x 10-11, OR = 1.45), ZNF365 at 10q21(rs224143; P = 2.20 x 10-9, OR = 1.38), ZMIZ1 at 10q22 (rs1250569; P = 3.05 x 10-7, OR = 1.30), NKX2-3 at 10q24 (rs4409764; P = 7.93 x 10-8, OR = 1.32), PTPN2 at 18p11(rs514000; P = 9.00 x 10-8, OR = 1.33), and USP25 at 21q11(rs2823256; P = 2.49 x 10-7, OR = 1.35), bringing the number of known CD loci (including HLA) in Koreans to 15. The additional six loci increased the total genetic variance for CD risk from 5.31% to 7.27 % in Koreans. Our study provides new biological insight to CD and supports the complementary value of genetic studies in different populations.

954T

Pathway Based Genome-Wide Association Studies Reveal the Association between Growth Factor Activity and Inflammatory Bowel Disease. *C. Kim, J. Li, Z. Wei, C. Cardinale, R. Baldassano, H. Hakonarson, International IBD Genetics Consortium (IIBDGC).* Children's Hospital of Philadelphia The Center for Applied Genomics 3615 Civic Center Blvd Abramson Research Center 1215A Philadelphia PA 19104.

The inflammatory bowel diseases (IBD) known as Crohn's disease (CD) and ulcerative colitis (UC) are related autoimmune conditions with a complex etiology comprised of genetic and environmental factors. Genetic studies have revealed 163 susceptibility genes for IBD, but still do not fully account for the genetic heritability of the disease. We exploited the largest CD data set (20,000 cases + 28,000 controls) and UC data set (17,000 cases + 33,500 controls) to date. We employed pathway-based approaches to identify genes that cooperatively make contributions to the genetic etiology of CD in a meta-analysis of 5 CD cohorts of European ancestry. In addition to the multiple immune-related pathways that have been implicated in the genetic etiology of IBD before, we found significant associations involving genes in growth factor signaling for CD. This result was replicated in an independent cohort genotyped on the ImmunoChip and in another pediatric cohort of European ancestry. This association with growth factor activity is not unique to CD. We found a similar significant association with UC in a meta-analysis of different cohorts. Our findings suggest that genes involved in growth factor signaling pathways may make joint contributions to the etiology of CD and UC, providing novel insight into the genetic mechanisms of these diseases.

955S

Whole-genome imputation identified 3 suggestive loci for inflammatory bowel disease in a Japanese population. *K. Yamazaki[1], Y. Fuyuno[1,2], A. Takahashi[3], T. Kawaguchi[4], M. Takazoe[4], M. Esaki[2], S. Nakamura[2], T. Matsui[5], T. Tanaka[6], M. Motoya[6], Y. Suzuki[7], Y. Kiyohara[8], T. Kitazono[2], Y. Nakamura[9], M. Kubo[1].* 1) Laboratory for Genotyping Development, Center for Integrative Medical Science, RIKEN, Yokohama, Japan; 2) Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; 3) Laboratory for Statistical Analysis, Center for Integrative Medical Science, RIKEN, Yokohama, Japan; 4) Department of Medicine, Division of Gastroenterology, Social Insurance Chuo General Hospital, Tokyo, Japan; 5) Department of Gastroenterology, Fukuoka University Chikushi Hospital, Fukuoka, Japan; 6) Department of Gastroenterology, Sapporo Kosei Hospital, Sapporo, Japan; 7) Department of Internal Medicine, Faculty of Medicine, Toho University, Chiba, Japan; 8) Department of Environmental Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; 9) Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

Crohn's disease (CD) and ulcerative colitis (UC) are the two major forms of inflammatory bowel disease (IBD), induced by multiple factors. Meta-analysis of genome-wide association studies (GWAS) identified 163 susceptibility loci for IBD from European descents. However, there was fewer reports of IBD GWAS from other ethnic groups. Recently, the human genetic resource has been growing over and it can make to impute missing genotypes. Therefore, we performed three GWAS (CD, UC and IBD) using imputation in Japanese population. We imputed genotypes of GWAS subjects which was a total of 376 CD cases, 376 UC cases and 3397 controls as previously reported. After applying stringent quality control for samples and SNPs, we performed the association analyses of 4,062,308 SNPs in CD GWAS, 4,059,167 SNPs in UC GWAS and 4,077,852 SNPs in IBD GWAS with accuracy (Rsq) > 0.9 and MAF > 0.05 of both cases and controls. However, we could not find significant association in each GWAS except for 2 reported loci (*TNFSF15* and MHC region). To identify new susceptibility loci, we performed the replication studies with independent cases and controls. We selected 1,605 SNPs (596 CD specific, 606 UC specific and 403 IBD shared) with P values < 5×10^{-5} after excluding SNPs located in previously reported loci. Among them, we picked up 137 tagging SNPs (53 SNPs as CD specific, 47 SNPs as UC specific and 37 SNPs as IBD shared) and genotyped using 948 CD patients, 361 UC patients and 4,161 controls. By combined analysis, we identified 3 suggestive loci with P value less than 1×10^{-6}. The two candidate loci as CD specific were in *RUNX3* on chromosome 1p36 (rs876109; $P = 4.01 \times 10\text{-}7$, odds ratio (OR) = 0.81) and on chromosome 11q21 (rs4435033; $P = 4.55 \times 10^{-7}$, OR = 1.37). The third locus was in *PARD6G* on 18q23 (rs4798947; $P = 5.39 \times 10^{-7}$, OR = 1.51) as UC specific. Though there was no suggestive locus shared IBD, the strongest association was shown in *TINK* on chromosome 3q26.2 (rs952209; $P = 7.29 \times 10^{-6}$, OR = 1.18). We suggested 2 candidate loci for CD and 1 candidate locus for UC in Japanese population. Further studies will be required what these loci play role in the pathogenesis of IBD.

956M

Genome-wide and exome chip study of subcutaneous and visceral adipose tissue reveals novel gender-specific adiposity loci in Hispanic Americans: The Insulin Resistance Atherosclerosis Family Study (IRASFS). *C. Gao[1,2,3], J. Ziegler[3,4], K.D. Taylor[5], J.M. Norris[6], Y.D.I. Chen[5], J.I. Rotter[5], L.E. Wagenknecht[7], C.D. Langefeld[3,4], N.D. Palmer[2,3,8,9].* 1) Program of Molecular Genetics and Genomics; Winston Salem, NC; 2) Center for Genomics and Personalized Medicine Research; Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Public Health Genomics; Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Biostatistical Sciences; Wake Forest School of Medicine, Winston-Salem, NC; 5) Institute for Translational Genomics and Population Sciences and Department of Pediatrics; Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA; 6) Department of Epidemiology, Colorado School of Public Health; University of Colorado, Aurora, CO; 7) Division of Public Health Sciences; Wake Forest School of Medicine, Winston-Salem, NC; 8) Center for Diabetes Research; Wake Forest School of Medicine, Winston-Salem, NC; 9) Department of Biochemistry; Wake Forest School of Medicine, Winston-Salem, NC.

Obesity is a major public health concern with strong correlations to numerous metabolic diseases. Moreover, regional fat deposition has been shown to increase disease risk above and beyond total adiposity. Here we report a genome-wide association study (GWAS) in 994 Hispanic Americans (N_{male}= 408, N_{female}=586) from IRASFS for ~8 million SNPs (genotyped and imputed). In addition, evaluation of low-frequency/rare variants was facilitated using the exome chip (81,560 variants). Association analyses were performed using computed tomography (CT) measures: subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), VAT adjusted by Body Mass Index (VAT_BMI), and visceral-subcutaneous adipose tissue ratio (VSR). As fat deposition appears to be gender specific, gender-stratified analyses were performed. Our results identified a total of 12 SNPs from five loci reaching genome-wide significance ($P<5.0\times10^{-8}$). SNP rs12657394 (MAF=19%), an intronic variant of *SRFBP1* on chromosome 5, showed strong association with VAT_BMI in males ($P_{male}=3.32\times10^{-8}$; P_{female}= 0.0025). *SRFBP1*, also named as *p49/STRAP*, encodes a Serum Response Factor Binding Protein. It has been shown to be involved in biosynthesis and/or processing of *GLUT4* (Glucose Transporter Type 4) in adipocytes, suggesting a role in regulating glucose homeostasis. SNP rs13247968 (MAF=42%), downstream of *SNX13* (Sorting Nexin 13), was also associated in males with VAT_BMI ($P_{male}=2.30\times10^{-9}$; $P_{female}=0.28$); rs117206355 (MAF=1%), located intergenically on 19p13.13, was strongly associated with VAT_BMI in females ($P_{male}=1.34\times10^{-4}$; $P_{female}=1.86\times10^{-9}$). In addition, two highly correlated intronic SNPs (rs2142795 and rs2425494; MAF=9%) in *PTPRT* showed strong association with VSR in males ($P_{male}=2.30\times10^{-9}$; P_{female}=0.30). rs12123452 (MAF=6%), downstream of *RIMS3*, showed a strong signal for VAT in males ($P_{male}=4.35\times10^{-8}$; $P_{female}=0.12$). To evaluate gender heterogeneity in fat deposition, interaction analyses were performed and significant gene-gender interactions were detected for all five significant loci with the most significant interaction at rs117206355 ($P=8.20\times10^{-5}$). These results revealed five novel adiposity loci displaying strong gender specificity. These observations provide genetic evidence for a differential mechanistic basis of fat deposition between genders and warrant further replication in larger cohorts.

957T

Meta-analysis of macronutrient intake in over 64,000 individuals using 1000 Genomes imputed genotypes confirms the association of *FGF21* with composition of dietary intake and suggests potential tissue-specific effects in liver and skeletal muscle. *A.Y. Chu[1], M. Graff[2], K.E. North[2], R.N. Lemaitre[3], J. Zhao[4], J. Luan[4], R.A. Scott[4], N. Tsernikova[5], M. Perola[6], J.S. Ngwa[7], L.A. Cupples[8], M.A. Nalls[9], D.K. Houston[10], J. Huang[11], Q. Qi[11,12], T.S. Ahuwalia[13], T.I.A Sorensen[13], C. Schulz[14], M. Orho-Melander[14], A.C. Frazier-Wood[15], T. Chen[15], P.S. de Vries[16], F.J.A. van Rooij[16], J.C. Kiefte[16], S. Kanoni[17], G. Dedoussis[17], T. Lehtimäki[18], O. Raitakari[19], D.I. Chasman[1], T. Tanaka[20], CHARGE Nutrition Working Group; DietGen Consortium.* 1) Brigham & Women's Hospital, Boston, MA, USA; 2) University of North Carolina, School of Public Health, Chapel Hill, NC, USA; 3) University of Washington School of Medicine, Seattle, WA, USA; 4) MRC Epidemiology Unit, University of Cambridge, Cambridge, UK; 5) Institute of Molecular and Cell Biology, University of Tartu, Estonia; 6) National Institute for Health and Welfare (THL), Helsinki, Finland; 7) Howard University, Washington DC, USA; 8) Boston University School of Public Health, Boston MA, USA; 9) National Insitute on Aging, NIH, Bethesda MD, USA; 10) Wake Forest School of Medicine, Winston Salem NC, USA; 11) Havard School of Public Health, Boston MA USA; 12) Albert Einstein College of Medicine, Bronx NY, USA; 13) University of Copenhagen, Copenhagen, Denmark; 14) University of Lund, Lund, Sweden; 15) USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston TX, USA; 16) Erasmus Medical Center, Rotterdam, the Netherlands; 17) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 18) Department of Clinical Chemistry, Fimlab Laboratories and School of Medicine, University of Tampere, Tampere, Finland; 19) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; 20) National Insitute on Aging, NIH, Baltimore MD, USA.

While dietary proportions of macronutrients, namely carbohydrate, protein, and fat, are relatively constant across populations, inter-individual variation of these quantities has been associated with chronic conditions such as obesity and diabetes. Prior genome-wide investigations based on common HapMap2 imputed genotypes from the CHARGE Nutrition Working Group and DietGen Consortium identified associations at the *FGF21* and *FTO* loci with macronutrient intake, estimated from food frequency questionnaires as the proportion of caloric intake from carbohydrate, protein, or fat. To further investigate the genetics of dietary intake, including associations from regions of low imputation quality in HapMap2 and lower frequency variants (MAF<5%), we performed a combined genome-wide inverse-variance weighted meta-analysis including >64K participants from 18 cohorts of European ancestry (ARIC, CHS, EPIC-Norfolk, FamHS, Fenland, FINRISK, FHS, GOYA MALE, H2000, HealthABC, InCHIANTI, Rotterdam Study I,II & III, WGHS, and YFS) using genotypes imputed to the 1000 Genomes phase 1 release 3 reference panel (March 2012). A total of 11.2 million variants remained after filtering on imputation quality (<0.4) and MAF (<0.01%), of which 43% had MAF<5%. In combined analysis, we confirmed the association of the minor allele for synonymous variant rs838133 (MAF=42%) at the *FGF21* locus with higher carbohydrate (β=0.3%) and lower fat (β=-0.2%) intake at genome-wide significance (p<5e-8) and at suggestive significance with lower protein intake (β=-0.1%, p=1.2e-6). We identified two missense SNPs in *FGF21* (MAF=36%) and *RASIP1* (MAF=49%) in moderate LD (r^2>0.6) with rs838133; only the *RASIP1* SNP is predicted to be damaging based on 3 of 4 protein prediction algorithms. Conditional analyses of rs831833 attenuated associations for both SNPs (p>0.05). No other variants attained genome-wide significance. However, among variants associated with intake of at least one macronutrient (p<1e-5) we found a greater proportion than expected that overlap with chromatin signatures for active gene regulation (H3K4me3) in liver, skeletal muscle and immune cells (p<0.05). An ongoing effort to include 6 additional cohorts (~46K participants) will increase the sample size to >100K making it the largest genome-wide investigation of dietary intake. The larger sample has potential to identify novel loci and therefore pathways for dietary habits and diet-disease relationships.

958S

Genome-Wide Association Study Identifies Novel Genetic Determinants of Emphysema Distribution Patterns. *A. El Boueiz[1,2,3], S.M. Lutz[4], R.P. Bowler[5], M.H. Cho[1,2,3], M.L. McDonald[1,3], N.M. Laird[6], T.H. Beaty[7], J.D. Crapo[5], E.K. Silverman[1,2,3], P.J. Castaldi[1,3], D.L. DeMeo[1,2,3].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 2) Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 3) Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 4) Department of Biostatistics, Colorado School of Public Health, University of Colorado, Aurora, CO, USA; 5) Division of Pulmonary Medicine, Department of Medicine, National Jewish Health, Denver, CO, USA; 6) Harvard School of Public Health, Boston, MA, USA; 7) Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.

Background: Emphysema is a chronic obstructive lung disease (COPD) with an irreversible destruction of lung parenchyma. The distribution of this destruction throughout the lung varies considerably. Upper lobe emphysema has emerged as an important predictor of response to lung volume reduction surgery. Yet, the determinants of apical versus basal emphysema distributions remain largely enigmatic. Alpha 1-antitrypsin deficiency is often associated with basilar emphysema whereas polymorphisms in xenobiotic enzymes (*GSTP1* and *EPHX1*) in non-alpha 1-antitrypsin deficient smokers have been related to apical predominance. These findings suggest the presence of genetic influences on emphysema distribution patterns. To investigate this hypothesis and identify additional genetic markers of emphysema distributional phenotypes, we performed a GWAS in the COPDGene study. Methods: From the full cohort of smokers in COPDGene, 6,094 non-Hispanic white (NHW) and 2,989 African American (AA) subjects with complete genotype and CT densitometry data were included in this analysis. Genotyping was performed on the Illumina Omni Express platform with additional markers imputed using 1,000 Genomes reference data. Standard subject and genotype level quality filters were applied. Under an additive genetic model adjusting for age, gender, pack-years of smoking, and genetic ancestry, each of the four CT scan emphysema distribution phenotypes (apical percent emphysema, basal percent emphysema, difference, and Log ratio between the two) was tested for genetic associations. Separate analyses in NHW and AA subjects were followed by a meta-analysis. Results: In the meta-analysis, markers from 7 distinct loci were associated with one or more of the emphysema distribution phenotypes at genome-wide significance. 3 loci (4q31 near *HHIP*; 15q25 near *CHRNA3/CHRNA5*; 1q41 near *TGFB2*) have been previously associated with COPD susceptibility, and 4 loci are newly identified (*TLE1*; *RANBP17*; *TOR1*; *NR2F1-AS1*). Of these 7 loci, 4 were nominally significant in both NHW and AA subjects (*CHRNA3/CHRNA5*; *RANBP17*; *TOR1B*; *NR2F1-AS1*) and 3 in NHW subjects only (*HHIP*; *TGFB2*; *TLE1*). Conclusion: This GWAS in NHW and AA individuals identified new loci that may represent risk loci for upper lobe predominant versus diffuse emphysema. Our results support the notion that genetic factors may impact emphysematous destruction. Grant support: The COPDGene study is supported by NIH R01 HL089897 and R01 HL089856.

959M

Heritability and locus susceptibility in age-related macular degeneration varies by clinical phenotypes. *L. Shen[1], T. Hoffmann[2,3], R. Melles[4], L. Sakoda[1], M. Kvale[2], Y. Banda[2], N. Risch[2], C. Schaefer[1], E. Jorgenson[1].* 1) Division of Research, Kaiser Permanente Northern California, Oakland, CA 94612; 2) Institute for Human Genetics, University of California San Francisco, San Francisco, CA 94143; 3) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA 94143; 4) Department of Ophthalmology, Redwood City Medical Center, Kaiser Permanente Northern California, CA.

Age related macular degeneration (AMD) is a complex, late onset vision disorder that progresses in stages. Majority prior studies focused on advanced AMD, and the most recent genome-wide association study (GWAS) meta-analysis identified 19 independent loci. Other AMD related clinical phenotypes have been less extensively examined. Here, we conducted a GWAS of overall and subgroups of AMD based on imputation to the 1000 Genomes Project reference panel. The analysis included a total of 5,762 overall AMD cases (2,504 with advanced AMD) and 45,929 controls of non-Hispanic white ancestry from the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort as part of Kaiser Permanente Research Program on Genes, Environment and Health. We observed large differences in genetic susceptibility across AMD stages and subtypes. The estimated variance explained by all autosomal SNPs was lower for overall AMD (32%) than advanced AMD (40%), lower for geographic atrophy (GA) (36%) than choroidal neovascularization (NV) (47%), and highest for cases with both NV and GA (52%). For the 19 previously reported risk loci, the proportion of variance explained, area under the ROC curve, and odds ratios also vary by AMD phenotypes. GWAS of the five AMD categories identified a novel association near CSN1S1 with the GA subtype ($p=3.02\times10\text{-}8$). After conditioning on the top SNPs at each risk locus, we identified a non-synonymous SNP in HLA-DBQ1 to be associated with the risk of overall AMD ($p=1.64\times10\text{-}12$). These findings suggest that shared genetic factors underlie the risk of AMD of all stages and forms. The strength of the effects vary by disease severity and subtype.

960T

Genome-Wide Association Study of Serum Sodium Concentration in Han Chinese Population residing in Taiwan. *I. Song[1], J. Yang[1], C. Chen[1], C. Sung[2], S. Lin[2], Y. Chen[1], J. Wu[1].* 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, Taiwan; 2) Division of Nephrology, Department of Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan.

Water is the most important human body component that counts for 60-70% of body weight. Water homeostasis is the essential vital balance which maintains human survival. Serum sodium concentration is the index reflecting the systemic water balancing condition. The normal serum sodium concentration is between 135 to 145 mM. Hypernatremia (>145 mM) was reported to associate with increase motility and hyponatremia (<135 mM) was suggested to be a risk factor for osteoporosis, myocardial infarction, and attention deficits. Recently, high heritability of serum sodium concentration was demonstrated in non-Hispanic Caucasian, African American, and American Indian populations. However, the genetic influences on serum sodium concentration in Han Chinese Taiwan have not yet been studied. Therefore, in current study, we aimed to identify the loci affecting serum sodium concentration. We performed GWAS on 375 individuals with relative high sodium concentration (138-153 mM) and 394 individuals with relative low sodium concentration (85-133 mM). All subjects carry no severe medical illness. We found that chromosomes 14q31.1 ($P<10\text{-}7$) and 17p13.3 ($P<10\text{-}6$) showed strong association. The replication with another independent cohort is in processed. We hope the findings from this study can help in early identifying subject that predisposing to spontaneous or iatrogenic hyper/hyponatremia.

961S

Y Chromosome degradation and male longevity in the Long Life Family Study. *M. Bailey[1], W. Daw[1], J. Lee[2], J. Zmuda[3], B. Thyagarajan[4], P. Sebastiani[5], R. Lin[1], T. Perls[6], K. Christensen[7], M. Province[1].* 1) Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO; 2) College of Physicians and Surgeons, Columbia University, New York, NY; 3) Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 4) Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN; 5) Department of Biostatistics, Boston University / Boston Medical Center, Boston, MA; 6) Geriatrics, Boston University / Boston Medical Center, Boston, MA; 7) Danish Aging Research Center, University of Southern Denmark, Odense, DENMARK.

Y chromosomal degradation and loss is a known phenomenon in male aging and has been the focus of recent studies. We explored the hypothesis that Y chromosome variants contribute to Y chromosome stability and also to successful aging in the Long Life Family Study (LLFS). We used data on 2182 men, of whom 2135 have 450 Y chromosome genetic markers and intensity data from the Illlumina HumanOmni2.5 GWA chip. LLFS was designed to examine the genetics of healthy aging and families were recruited consisting of siblings who showed exceptional longevity, their children, and spouses where available. The mean age of male subjects on first examination, when DNA was collected, was 71. The mean age of men with Y chr intensity < 0.9 (indicating substantial Y chromosome degradation/loss) was 91, consistent with prior studies. Using a minor allele persistence algorithm, we have identified several SNPs with significant differences in age-specific allele frequencies, possibly associated with Y chromosome loss (p-values < 1.1×10-4, significant for 450 Y chromosome SNPs—Bonferroni alpha correction). Of these, rs17316547 is a variant within a predicted snRNA and is within 30kb of PRKY. Additionally, of the 231 individuals who carry this allele in our data, the youngest recorded death was 83 years old. The PRKY gene is similar to the protein kinase, X-linked gene in the pseudoautosomal region of the X chromosome. The gene is classified as a transcribed pseudogene because it has lost a coding exon that results in all transcripts being candidates for nonsense-mediated decay (NMD) and unlikely to express a protein. Abnormal recombination between this gene and a related gene on chromosome X is a frequent cause of XX males and XY females, making it a plausible biomarker for genomic instability in age related diseases.

962M

Genetic admixture and proliferative diabetic retinopathy in Latinos. *X. Gao[1], W.J. Gauderman[2], P. Marjoram[2], M. Torres[3], Y.I. Chen[4], K.D. Taylor[4], J.I. Rotter[4], R. Varma[3].* 1) Dept Ophthalmology & Visual Sci, Univ Illinois, Chicago, Chicago, IL; 2) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) USC Eye Institute, Department of Ophthalmology, University of Southern California, Los Angeles, CA; 4) 3Institue for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA, Torrance, CA.

Diabetic retinopathy is a leading cause of blindness in working age adults, and proliferative diabetic retinopathy (PDR) is the most advanced stage of the disease. Epidemiology studies have observed that Latinos have a higher prevalence of severe diabetic retinopathy than whites. The purpose of this study is to test the association between genetic admixture and PDR in Latinos with type 2 diabetes mellitus (T2DM). We conducted a case-control study using 647 T2DM subjects (552 controls and 95 PDR cases) selected from the tails of the diabetic retinopathy distribution from the Los Angeles Latino Eye Study. Genotyping was performed on the Illumina OmniExpress BeadChip (730,525 markers). We estimated genetic ancestry in Latinos using STRUCTURE with the HapMap reference panels. Univariate and multivariate logistic regression analyses were used to test the relationship between the proportions of genetic ancestry in Latinos and PDR. Native American ancestry in Latino T2DM subjects is significantly associated with PDR (P = 0.004) in our univariate analysis. The association remained significant after adjusting for age, sex, hemoglobin A1c, body mass index, systolic blood pressure, education and income. Risk factors for PDR in Latinos in our multivariate analysis include duration of diabetes (P < 0.0001), hemoglobin A1c (P = 0.0003), systolic blood pressure (P = 0.0004), and Native American ancestry (NAA, P = 0.006). The empirical p-value for NAA from permutation tests was 0.005. We also validated the NAA estimates in Latinos using ADMIXTURE with the 1000 Genomes Project reference panels and obtained consistent results. Our results demonstrate for the first time that NAA is a significant risk factor for PDR in Latinos. We are further performing local ancestry analysis, which may identify specific genomic regions associated with PDR in Latinos.

963T

GWAS of 89,283 individuals identifies genetic variants associated with being a morning person. *Y. Hu[1], A. Shmygelska[1], D. Tran[1,2], N. Eriksson[1], J. Tung[1], D. Hinds[1].* 1) 23andMe, Inc, Mountain View, CA, 94043; 2) Department of Biological Sciences, San Jose State University, San Jose, CA, 95112.

Circadian rhythms are a nearly universal feature of living organisms and affect almost every biological process. Our innate preference for mornings or evenings is determined by the phase of a circadian rhythm. We conducted a genome-wide association analysis of self-reported morningness, followed by analyses of biological pathways and its relationship to other phenotypes. We identified a total of 15 loci that were genome-wide-significant. Of those, seven were close to known circadian genes (RSG16, VIP, PER2, HCRTR2, RASD1, PER3 and FBXL3), four were near genes that could plausibly play a role in circadian rhythms (PLCL1, APH1A, FBXL13 and NOL4) and another four were near genes not previously associated with circadian rhythms (TOX3, AK4, DLX5 and ALG10B). We identified circadian and phototransduction related pathways enriched in our results and found morningness associated with many sleep phenotypes such as insomnia and sleep duration. Morningness is also associated with BMI and depression but in a Mendelian randomization analysis we did not find statistical evidence for a causal relationship. Our findings reinforce currently known circadian genes and will guide future studies of circadian rhythm, sleep and related disorders.

964S

Genome-wide Copy Number Scan Identifies Involvement of IRF6 in an Indian Family with Van der Woude Syndrome. *D.S. Manjegowda[1], M. Prasad[2], A.M. Veerappa[3], N.B. Ramachandra[3].* 1) Department of Biomedical Sciences, K. S Hegde Medical Academy, Nitte University, Deralakatte, Mangalore-575018, Karnataka, India; 2) Centre for Cleft Services, St. Joseph Hospital, DCKH-Cleft Centre, Mysore-570015. Karnataka, India; 3) Genomics Laboratory, Department of Studies in Zoology, University of Mysore, Manasagangotri, Mysore-570006, Karnataka, India.

Van der Woude syndrome (VWS) is an autosomal dominant developmental malformation presenting with bilateral lower lip pits related to cleft lip, cleft palate, and other malformations. We performed a whole genome copy number variations (CNV) scan in the members of a VWS family in India using 2.6 million combined SNP and CNV markers. We found CNVs affecting IRF6, a known candidate gene for VWS in all the three cases, while none of the non-VWS member showed any CNVs in the IRF6 region. The duplications and deletions of the chromosomal critical region in 1q32-q41, confirms the involvement of CNVs in IRF6 in south Indian VWS patients. Molecular network analysis of these and other Cleft lip/palate related module genes suggests them to be associated with cytokine-mediated signaling pathway, response to interferon-gamma mediated signaling pathway. This is a maiden study indicating the involvement of CNVs in IRF6 in causing VWS in the Indian population.

965M

Genetic determinants of healthspan - analysis of Wellderly dataset. *W. Sikora-Wohlfeld[1], M. Sirota[1], E. Scott[2], A. Torkamani[2], E. Topol[2], A.J. Butte[1].* 1) Department of Pediatrics, Stanford University, Stanford, CA; 2) Scripps Translational Science Institute, La Jolla, CA.

Understanding the mechanisms of aging would potentially enable detecting and preventing debilitating diseases and expanding the life expectancy. Numerous research efforts have focused on searching for 'longevity genes'. Development of genotyping and genome sequencing techniques enabled direct investigation of the genetic background of longevity. In recent years a number of studies have been conducted, where the genomes of people who have lived to a very advanced age in good health, were analyzed to search for variants for longevity. However, the mechanisms governing long healthspan remain to be explained.

In this study we have analyzed genome sequences of 534 participants of the Wellderly project, who have lived to 80 years or beyond. We compared this cohort to a control group, derived from the 1000 Genomes Project, assuming that the 1000 Genomes participants represent a sample from an ordinary population with average life expectancy. Using principal component analysis to estimate genetic ancestry, we identified a group of 416 Wellderly individuals who matched 174 individuals from CEU and GBR populations in the 1000 Genomes Project.

We performed an association test for a selected set of SNPs known to be associated with common human diseases. We derived this list of SNPs from VARIMED, a database of human disease-SNP associations. At the time of this work, VARIMED contained 466,890 unique SNPs associated with 6,691 disease and related phenotypes, manually curated from 17,088 publications.

Among the top variants differentiating the Wellderly and 1000 Genomes cohorts, we identified four markers in the T cell receptor alpha constant gene (*TRAC*). In addition, we found a marker in the immunoglobulin heavy locus gene (*IGH*). A link between aging and immunity has been established before and the performance of the immune system has been suggested to affect healthspan. While still preliminary, our findings suggest an association between immune pathway genetics and longevity, as well as enabling the exploration of other longevity-related markers.

966T

Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *H. Aschard, B. Vilhjálmsson, A. Joshi, A. Price, P. Kraft.* Department of Epidemiology, Harvard School of Public Health, Boston, MA.

Adjustment for covariates in genome-wide association studies (GWAS) has a dual purpose: (a) to account for potential confounding factors that can bias SNP effect estimates, and (b) to improve statistical power by reducing residual variance. Recently, researchers have conducted GWAS of human traits and diseases while adjusting for other heritable covariates with another motivation: identifying genetic variants associated only with the primary outcome. We show that this objective is fulfilled when the tested variants have no effect on the covariate or when the correlation between the covariate and the outcome is fully explained by a direct effect of the covariate on the outcome (i.e. mediation). For all other scenarios, an unintended bias is introduced with respect to the primary outcome as a result of the adjustment, and this bias may lead to false positives. We illustrate this point by providing examples from published genome-wide association studies, including a large meta-analysis of waist-to-hip ratio (WHR) adjusted for body mass (BMI) index from the GIANT consortium, where genetic effects may be biased as a result of adjustment for BMI.

We first show that the expected bias of the effect estimate in the covariate adjusted analysis is proportional to the correlation between the covariate and the outcome, and the association between the covariate and the genetic variant tested. We then derive a statistical test to evaluate the presence of such bias using GWAS summary information and the correlation between the covariate and the phenotype. When applied to the GIANT summary statistics, we observed that half of the reported associations with WHR adjusted for BMI are likely influenced by a genetic association with BMI. Finally, we show that heritability estimation of covariate adjusted traits is also subject to the same bias. In particular, we demonstrate that the heritability of an outcome adjusted for a heritable covariate may not necessarily represent the genetic component of the outcome, but rather a heterogeneous mixture of outcome-specific, covariate-specific and shared genetic components.

In summary we highlight that adjusting for heritable human traits such as BMI to remove mediated genetic effect through these variables, as commonly done in many genetic studies, should be done with care as it can induce false signal.

967S

Meta-analysis of genome-wide association studies in alopecia areata reveals new susceptibility loci and resolves HLA associations. *R. Betz[1], L. Petukhova[2,3], S. Ripke[6,7], H. Huang[6,7], S. Redler[1], T. Becker[8,9], S. Heilmann[1,24], T. Yamany[2], M. Duvic[10], M. Hordinsky[11], D. Norris[12], V. Price[13], J. Mackay-Wiggan[2], A. Menelaou[14], G.M. Destano[5], M. Bohm[16], U. Blume-Peytavi[17], H. Wolff[18], G. Lutz[19], R. Kruse[20], C.I. Amos[21], A. Lee[22], P.K. Gregersen[22], B. Blaumeiser[23], D. Altshuler[6,7], P.I.W. de Bakker[14,15], M.M. Nothen[1,24], R. Clynes[2,4], M. Daly[6,7], A.M. Christiano[2,5].* 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Dermatology, Columbia University, NY, NY; 3) Department of Epidemiology, Columbia University, NY, NY; 4) Department of Medicine, Columbia University, NY, NY; 5) Department of Genetics & Development, Columbia University, NY, NY; 6) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 7) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 8) German Center for Neurodegenerative Diseases, Bonn, Germany; 9) Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany; 10) Department of Dermatology, MD Anderson Cancer Center, Houston,TX; 11) Department of Dermatology, University of Minnesota, Minneapolis, MN; 12) Department of Dermatology, University of Colorado, Denver, CO; 13) Department of Dermatology, UCSF, San Francisco, CA; 14) Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands; 15) Department of Epidemiology, University Medical Center Utrecht, Utrecht, The Netherlands; 16) Department of Dermatology, University of Münster, Münster, Germany; 17) Clinical Research Center for Hair and Skin Science, Department of Dermatology and Allergy, Charité-Universitätsmedizin Berlin, Berlin, Germany; 18) Department of Dermatology, University of Munich, Munich, Germany; 19) Dermatological Practice, Hair and Nail, Wesseling, Germany; 20) Dermatological Practice, Paderborn, Germany; 21) Community and Family Medicine and Genetics, Dartmouth College, Hanover, NH; 22) The Feinstein Institute for Medical Research, Manhasset NY; 23) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; 24) Department of Genomics, Life & Brain Center, University Bonn, Bonn, Germany.

Alopecia areata (AA) is one of the most prevalent autoimmune diseases, with ten known susceptibility loci so far. Here, we performed a meta-analysis in AA, combining data from two GWAS, and including replication data from Immunochip supplemented with Sequenom genotyping, for a total of 2,807 cases and 6,975 controls. The strongest region of association was the MHC, where we fine-mapped 4 independent effects, all implicating HLA-DR as a key etiologic driver, three of which alter protein sequence in the binding cleft of HLA-DRβ1. Outside the MHC, we identified two novel loci that exceed statistical significance, containing ACOXL/BCL2L11(BIM) (2q13; rs3789129, p=1.51×10-8, ORA=1.3); LRRC32(GARP) (11q13.5; rs2155219, p=1.25×10-8, ORT=1.2). A third region achieved nominal significance: SH2B3(LNK)/ATXN2 (12q24.12; p=1.3×10-7), which has been observed in other autoimmune diseases. Expression analysis of genes in these three regions provide biologic plausibility to these findings and further support the causal role of aberrant immune processes in AA, with contributions from both immune cells as well as the end organ (hair follicle). Finally, we performed a cross phenotype meta-analysis integrating our data with data from seven other autoimmune diseases, providing insight into the molecular taxonomy of autoimmune diseases and the alignment of AA within this class of disorders. Importantly, as GWAS help to resolve disease mechanisms and identify pathogenic pathways perturbed in AA and autoimmunity in general, these approaches advance the field towards precision medicine in autoimmunity.

968M

Examining the genetic basis of variation in sitting height ratio (SHR) using population cohorts. *Y. Chan[1,2,3], E.T. Lim[2,3,4], D. Strachan[5], G. McMahon[6], G. Davey-Smith[6], R.M. Salem[1,2], J.N. Hirschhorn[1,2,3].* 1) Endocrinology, Boston Children's Hospital, Boston, MA; 2) Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Genetics Dept, Harvard Medical School, Boston, MA; 4) Genetics and Genomics, Boston Children's Hospital, Boston, MA; 5) Population Health Research Institute St George's, University of London, London, UK; 6) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

Sitting height ratio (SHR) is the ratio of sitting height to total height and is a measure of body proportion. Studying SHR is important because it can be a more sensitive phenotype to detect mild growth disorders than height. While sitting height is heritable, with approximately 70% of variation within a population attributable to genetic factors, the heritability of SHR has not been well-studied. Nonetheless, there can be large differences in the average SHR between populations; for example, the average SHR of European-ancestry individuals is about 1 standard deviation greater than the average SHR of African-ancestry individuals. This difference is one of the largest reported for anthropometric traits. Here, we report the results from the first genome-wide genetic studies of SHR in large population cohorts of different ancestries. Using data from dbGAP, we first examined the association of SHR with European ancestry in 3,069 African-Americans. We observed a highly significant association ($P<5\times10^{-16}$); as expected from the average differences between populations, European ancestry is associated with increased SHR. We next performed genome-wide association studies of SHR on 11,943 European-Americans and 3,069 African-Americans. We discovered three novel loci significantly associated ($P<5\times10^{-8}$) with SHR, near *ANO10*, *PTPRM*, and *ITM2A*. Interestingly, *ITM2A* is on the X-chromosome and is associated with SHR only in women. This locus was recently reported to also be associated with height and is known to escape X-inactivation. However, these three loci do not account for differences in SHR between individuals of European and African ancestries, and there were no significant peaks from admixture mapping, suggesting that the genetic basis for the difference in SHR between populations is largely polygenic. Finally, we tested 421 independent SNPs known to be associated with height and observed that more of these SNPs are nominally associated ($P<0.05$) with SHR than expected by chance ($N=49$, $P=21\times10^{-8}$). Of the 49 height-decreasing alleles, 31 are associated with increased SHR and 18 are associated with decreased SHR, suggesting that different height loci have specific effects on skeletal growth of the spine or long bones. Using our results, we are able to dissect the role of previously reported loci associated with height, as well as new loci associated with SHR, which can potentially aid in therapeutic target discovery for skeletal disorders.

969T

Six novel loci associated with VEGF circulating levels identified by a meta-analysis of genome-wide association studies. *S. Choi[1,2,3], D. Ruggiero[4], R. Sorice[4], T. Nutile[4], C. Song[5,9], N. Ndiaye[8], M. Stathopoulou[8], C. Barbieri[7], C. Bellenguez[10,11,12], M. Concas[6], P. Fitzgerald[22], V. Guðnason[17,18], A. Leutenegger[13,14], E. Ingelsson[5], P. Kovacs[21], V. Lagou[15,16], J. Lamont[22], L. Lind[9], G. Maestrale[6], M. Pirastu[6], C. Sala[7], A. Smith[17,18], D. Toniolo[7], A. Tönjes[20], M. Traglia[7], R. Vasan[3,19], A. DeStefano[1,2,3], S. Visvikis-Siest[8], S. Seshadri[1,2,3], M. Ciullo[4].* 1) Department of Neurology, Boston University School of Medicine, Boston, MA; 2) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 3) National Heart, Lung and Blood Institute's Framingham Heart Study; 4) Institute of Genetics and Biophysics, National Research Council of Italy, Naples, Italy; 5) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 6) Institute of Population Genetics, National Research Council of Italy, Sassari, Italy; 7) Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milano, Italy; 8) UMR INSERM U 1122; IGE-PCV 'Interactions Gène-Environnement en Physiopathologie Cardio-Vasculaire'.Faculté de Pharmacie, Université de Lorraine, Nancy, France; 9) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden; 10) Institut Pasteur de Lille, Lille, France; 11) Inserm, U744, Lille, France; 12) Université Lille-Nord de France, Lille, France; 13) Inserm, U946, Paris, France; 14) Univ Paris Diderot, Sorbonne Paris Cité, IUH, UMR-S 946, Paris, France; 15) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 16) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom; 17) Icelandic Heart Association, IS-201 Kopavogur, Iceland; 18) University of Iceland, 101 Reykjavik, Iceland; 19) Section of Preventive Medicine and Epidemiology, Department of Medicine, Boston University Schools of Medicine and Public Health, Boston, MA; 20) University of Leipzig, Department of Medicine, Leipzig, Germany; 21) University of Leipzig, IFB Adiposity Diseases, Leipzig, Germany; 22) Randox Laboratories, Crumlin, United Kingdom.

Vascular Endothelial Growth Factor (VEGF) is the most important pro-angiogenic factor, implicated in both physiological and pathological angiogenesis. A previously published GWAS had identified four loci independently associated with VEGF levels, two on chromosome 6, one on chromosome 9 and one on chromosome 8 (Debette and Visvikis-Siest et al, 2011). We sought to identify additional loci associated with circulating VEGF levels measured on ~13.000 individuals from 6 cohorts using genome wide association data imputed to the 1000genomes v3 panel. A GWAS of VEGF levels was performed in each cohort and the results were meta-analyzed. Five chromosomal regions (5q14.3, 6p21.1, 8q23.1, 9p24.2, 10q21.3) containing SNPs associated with VEGF levels at genome-wide significance ($p<5\times10^{-8}$) were found. To identify independently associated variants within the genome-wide significant genomic regions, conditional analyses were carried out. These analyses revealed 10 independent signals (1 on 5q14.3, 4 on 6p12.1, 1 on 8q23.1, 3 on 9p24.2 and 1 on 10q21.3). Further, 13 loci contained variants suggestively associated at $5\times10^{-8}\leq$p-value$<1\times10^{-5}$. The lead variant for each of the 10 independent signals and the 13 suggestive loci was carried forward to in silico and de novo replication in ~2800 individuals from 4 additional independent cohorts. Ten signals, 8 out of the top 10 independent variants and 2 out of the 13 suggestive signals in the discovery sample, were successfully replicated ($p<5\times10^{-8}$) in the meta-analysis of the combined discovery and replication samples. Overall, we confirmed the association of 4 already known loci and found 6 new signals, 4 located in novel chromosomal regions (5q14.3, 10q21.3, 16q24.2, and 18q22.3) and 2 in previously identified chromosomal regions (6p21.1 and 9p24.2). These 10 variants explain about 51% of the variability of the circulating levels of VEGF. The Ingenuity Pathway Analysis software (IPA) was used to explore the relationships between all genes associated with circulating VEGF levels. Twenty-four genes located in the regions identified by the 10 replicated variants were selected as focus genes. Out of them, 17 genes were connected within a unique network of 70 molecules. The analysis revealed strong functional relations of the gene network with embryonic and organism development, and cardiovascular system development and function. Further analyses will be necessary to identify the functional variants in the identified loci.

970S

Genome-Wide Association Study in Different Stages of Clinical Progression of Alzheimer Disease. *J. Chung[1], L. Farrer[1,2,3,4,5], G. Jun[1,2,3].* 1) Biomedical Genetics, Boston University Medical Campus, Boston, MA; 2) Biostatistics, Boston University, Boston, MA; 3) Ophthalmology, Boston University, Boston, MA; 4) Epidemiology, Boston University, Boston, MA; 5) Neurology, Boston University, Boston, MA.

Background: Amyloid-beta 42 (Aβ42) and phosphorylated tau (p-tau) levels in cerebrospinal fluid (CSF) are key biomarkers for Alzheimer disease (AD) and known to become abnormal before the onset of clinically diagnosed AD. However, the discovery of new genetic risk factors in AD using CSF biomarkers as surrogate measures of disease is challenging in sample comprising a mixture of subjects with AD and mild cognitive impairment (MCI), as well as clinically normal (CN) subjects. Considering different distribution of biomarker levels among these subgroups, we conducted genome-wide association (GWA) analyses separately in these subgroups and in the total sample. Methods: A total of 783 subjects comprising with 112 AD, 483 MCI, and 188 CN participants with GWA data were available from two sets of the Alzheimer's Disease Neuroimaging Initiative (ADNI) Study. The residuals of a quantitative trait of Aβ42/p-tau ratio after adjusting for age and sex were normalized by rank transformation. GWA markers using the normalized trait in each set were analyzed in a linear regression model after accounting for population substructure. The results across sets were combined using meta-analysis. Results: We observed genome-wide significant (GWS, $p<5\times10^{-8}$) association with SNPs in the APOE region and three novel loci including MAST4 (rs80222306, minor allele frequency [MAF]= 14%, meta-analysis p [meta-p]=2.9×10^{-8}), and FBXO16 (rs10112794, MAF= 4%, meta-p=1.1×10^{-8}) in CN subjects, and CCDC3/OPTN (rs12356199, MAF=5%, meta-p=3.4×10^{-8}) in MCI subjects. Associations with several SNPs in CADM1 approached the GWS threshold (rs10891864, MAF=39%, meta-p=5.9×10^{-8}) in AD cases. Rs80222306 was also strongly associated in separate analyses of CSF Aβ42 (meta-p=8.1×10^{-6}) and p-tau (meta-p= 6.0×10^{-5}) levels in CN subjects. None of these genome-wide or near genome-wide significant SNPs outside the APOE region were even nominally significant in other subgroups (meta-p>0.05) and less significant in the total sample (meta-p>4.7×10^{-6}). Conclusion: Our results suggest that markers of AD pathogenesis are influenced by multiple loci which exert their effects at different stages of clinical progression toward AD. Further studies in large independent datasets are warranted to identify the functional variants and understand the mechanisms underlying the stage-specific genetic associations with CSF biomarkers of AD.

971M

Genome-wide association studies for dental caries in African American and Latino populations: Novel genes, heterogeneity, and replication. *J. Colavincenzo[1], J.R. Shaffer[2], E. Feingold[1,2], C. Sanchez[3], T. McHenry[3], F.W.B. Deleyiannis[4], D.W. McNeil[5], R. Crout[6], R.J. Weyant[7], M.L. Marazita[2,3,8].* 1) Department of Biostatistics, Graduate School of Publich Health, University of Pittsburgh, Pittsburgh, PA; 2) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 3) Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 4) Department of Surgery, Plastic and Reconstructive Surgery, School of Medicine, University of Colorado, Aurora, CO; 5) Dental Practice and Rural Health, West Virginia University, Morgantown, WV; 6) Department of Periodontics, School of Dentistry, West Virginia University, Morgantown, WV; 7) Department of Dental Public Health and Information Management, University of Pittsburgh, Pittsburgh, PA; 8) Clinical and Translational Sciences Institute, and Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA.

Motivation: Dental caries is the most common chronic disease worldwide, affecting a majority of children and adults. Disease burden and related comorbidities (pain, tooth loss, missed work/school, difficulty eating, hearing, and sleeping, and emergency room visits) are concentrated in vulnerable populations, including minority groups and those living in poverty or rural areas. Though dental caries experience is highly heritable, few specific risk loci have been identified and rigorously validated. To date, multiple genome-wide association studies (GWAS) in whites have nominated plausible caries risk genes, however, no GWAS in other races or ethnic groups have been performed. Methods: We performed four GWAS scans for dental caries in child and adult African American and Guatemalan cohorts. Results: Several novel caries genes were nominated, including *IGF-1* (insulin-like growth factor 1; p-value 2.7E-8), which is thought to be important for tooth development and saliva production. Additionally, loci previously implicated in GWAS of whites showed evidence of replication, including *LPO* (lactoperoxidase), an oral bactericidal enzyme. Conclusions: This study provides further support for the role of previously nominated genes, and suggests that studying cohorts with different ancestry and environment may help discover a wider range of caries genes. Grants: R01-DE014899, U01-DE018903, R01-DE016148.

972T

Genetic determinants of normal human facial variation. *J.B. Cole[1], M. Manyama[2], J. Larson[2], D.K. Liberton[2], T.M. Ferrara[1], S.L. Riccardi[1], M. Li[3], W. Mio[3], S.A. Santorico[1], B. Hallgrimsson[2], R.A. Spritz[1].* 1) Human Medical Genetics and Genomics Program, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO; 2) Department of Cell Biology & Anatomy, University of Calgary, Calgary, AB; 3) Department of Mathematics, Florida State University, Tallahassee, FL.

Facial shape exhibits perhaps the greatest range of variation of any normal human trait, yet at the same time represents the most recognizable human characteristic. Facial shape comprises a multiplicity of complex multifactorial traits with clear genetic components. Nevertheless, little is known about normal human facial development. To identify genetic determinants of normal human facial shape variation, we carried out a large genome-wide association study (GWAS), replication study, and meta-analysis in 6300 normal African Bantu children from Tanzania. We used advanced 3D imaging technology, a revolutionary automated landmarking method, and geometric morphometrics to capture and analyze quantitative facial shape phenotypes, followed by genome-wide analysis of over 2.5 million SNPs. Our GWAS identified a number of loci with association surpassing criterion for genome-wide significance in an African population (P<2.5E-08). These include a gene desert at chr11q23.1 associated with nasal ala length, *AKAP13* associated with philtrum width, *DRAM2* associated with upper facial height, *PTCH1* associated with allometry, *DHX34* associated with upper facial depth, *SCHIP1* associated with centroid size, and *SRC* associated with nasal width. *AKAP13* encodes an A-kinase anchor protein. *DRAM2* encodes an inducer of autophagy. *PTCH1* encodes a receptor for sonic hedgehog, and *PTCH1* mutations result in holoprosencephaly. *DHX34* encodes a regulator of nonsense-mediated RNA decay that is essential for embryonic head development in zebrafish. *SCHIP1* encodes a schwannomin interacting protein with limited functional information. *SRC* has an essential role in bone remodeling, and *Src*-knockout mice have broad, abnormal facies. Our GWAS of normal facial shape in an African population did not confirm previously-reported association of mid-nasal shape with *PAX3* in European-derived whites, did not identify loci that have been associated with non-syndromic cleft lip/palate, and largely did not identify loci that are associated with single-gene disorders of facial morphogenesis. Most of the loci that were associated with facial shape in our GWAS encode proteins that either participate in basic cellular processes or for which functions remain largely unknown. Our results thus provide both novel candidate genes and novel insights into early development of the human face.

973S

Contribution of common polygenic variation captured by the Immunochip to celiac disease heritability in an independent Irish population. *C. Coleman[1], EM. Quinn[1], AW. Ryan[1], RJL. Anney[2], V. Trimble[1], DW. Morris[2], G. Donohoe[2], J. Conroy[3], G. Trynka[4], C. Wijmenga[5], S. Ennis[3], R. McManus[1].* 1) 1Department of Medicine, Institute of Molecular Medicine, Trinity College Dublin, St. James's Hospital, Dublin 8, Ireland; 2) Department of Psychiatry, Trinity College Dublin, Dublin, Ireland; 3) Conway Institute, University College Dublin, Ireland; 4) Genetics Department, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 5) Genetics Department, University Medical Center and University of Groningen, The Netherlands.

Celiac disease (CD) is a chronic immune-mediated disease with a prevalence of ~1% in European populations. Following the large Immunochip study of Trynka et al (2011) the HLA and 39 other CD susceptiblity loci are known. In an independent Irish CD case-control study we examined whether reported risk alleles were similar in direction of effect and whether a weighted burden of risk alleles (or polygenic risk score) could be used to distinguish case status. The polygenic risk score (Purcell et al 2009) has been used to examine the influence of risk alleles en masse to disease susceptibility in several complex disorders. Following stringent quality control we analyzed 143,074 markers genotyped on Immunochip in 425 cases and 453 controls. To examine concordance in the observed direction of effect in our sample, we performed a binomial sign-test for LD independent markers identified as genome-wide significant by Trynka et al. Secondly, for LD independent markers we calculated the polygenic risk score for each individual in our study. Regression was performed for disease status adjusting for marker-count-per-score (missingness), gender and population covariates. Binomial sign test indicated there was significant concordance in direction of effect between studies. 83% (122/147) of genome-wide significant SNPs show effect in the same direction (Pr (K>=122)=7.9×10 -17). When restricted to non-HLA markers 10/11 (91%) show effect in the same direction (Pr (K>= 10)=0.0059). The polygene analysis showed that polygenic risk scores were significantly associated with coeliac case-control status across a range of p threshold values. Including the HLA markers up to 36% of the variance was explained by the polygenic score (SNPs P< 0.0001; P=7.86×10 -67). 12% of the variance could be explained by the non-HLA SNPs alone (SNPs P< 0.001; P=9.56×10 -22). We have replicated the findings of a large CD association study in an independent Irish population. Use of the polygene risk analysis allows for highly significant findings even in a reduced sample size. Polygenic scores explained a significant proportion of the variance in coeliac disease confirming the contribution of common SNPs to CD susceptibility.

974M

Genetic insights into primary biliary cirrhosis - an international collaborative meta-analysis and replication study. *H.J. Cordell[1], Y. Han[2], Y. Li[2], G.F. Mells[3], G.M. Hirschfield[4], G. Xie[5], B. Juran[6], M.E. Gerschwin[7], P. Invernizzi[8], K. Lazaridis[6], C.A. Anderson[9], M.F. Seldin[7], C. Amos[2], R.N. Sandford[3], K. Siminovitch[5], Canadian-US, Italian and UK-PBC Consortia.* 1) Newcastle University, UK; 2) Dartmouth College, USA; 3) University of Cambridge, UK; 4) University of Birmingham, UK; 5) Mount Sinai Hospital, Toronto, Canada; 6) Mayo Clinic, USA; 7) UC Davis, USA; 8) Humanitas, Italy; 9) Wellcome Trust Sanger Institute, UK.

Primary biliary cirrhosis (PBC) results from an interaction of genetic and environmental factors. To date, four genome-wide association studies (GWAS) and two Illumina Immunoarray studies of PBC have helped delineate the genetic architecture of this disease. These studies have confirmed associations at the human leukocyte antigen (HLA)-region and identified 27 non-HLA susceptibility loci. Candidate genes are notably involved in the IL-12 signalling cascade. To identify additional risk loci for PBC, we have undertaken genome-wide meta-analysis (GWMA) of discovery datasets from the North American, the Italian and the UK GWAS of PBC, with a combined, post-QC sample size of 2745 cases and 9802 controls. Genome-wide imputation of each discovery dataset was undertaken in MACH using HapMap3 as reference panel; GWMA was undertaken using ProbABEL and META. Following meta-analysis, the index single nucleotide polymorphisms (SNPs) at selected loci with P_GWMA<2×10-5 were genotyped in a validation cohort consisting of 3716 cases and 4261 controls. To prioritise candidate variants and genes at confirmed risk loci, we used the ENCODE and the 1000Genomes datasets to identify SNPs within regulatory elements and non-synonymous (ns) SNPs in strong linkage disequilibrium (LD) with the index variant (r2>0.8). We identified seven previously unknown risk loci for PBC. Functional annotation of these loci revealed SNPs within regulatory elements that are predicted to affect expression of DGKQ (4p16), PAM (5q14) and IL21R (16p12), that are strongly-correlated to the index variant. Other candidate genes include IL12B (5q31), which forms part of the IL-12 signalling cascade, and CCL20 (2q36), which is involved in chemo-attraction of lymphocytes and dendritic cells towards epithelia and is expressed by TH17 cells originating from Foxp3+ T cells. Pathway analysis identified several highly plausible gene sets associated with PBC, including the IL-12 and JAK-STAT signalling pathways, and implicated several other immune processes in the pathogenesis of PBC, including innate immune processes (e.g. IFN-α,β signaling). Conclusion: This uniquely powered international collaborative GWMA and replication study confirms additional immunologically relevant loci that are associated with the risk of developing PBC.

975T

A Genome Wide Association Study of peanut sensitisation in the Manchester Asthma and Allergy Study. *J.A. Curtin[1], A. Custovic[1], A. Simpson[1], E.N.C. Mills[2].* 1) Centre for Respiratory Medicine and Allergy,Institute of Inflammation and Repair, University of Manchester, Manchester, Manchester, United Kingdom; 2) Institute of Inflammation and Repair, Manchester Academic Health Sciences Centre, Manchester Institute of Biotechnology, University of Manchester, United Kingdom.

Background: Very little is known about the molecular mechanisms underlying food allergy, including the model food allergen peanut. Twin studies indicate a significant genetic influence on peanut allergy, with one study estimated the heritability of peanut allergy at 81. 6%. Within the setting of a prospective birth cohort study, the Manchester Asthma and Allergy Study (MAAS), we investigated if a Genome-wide association study (GWAS) could help explain some of the heritability of peanut sensitisation. **Methods:** DNA was genotyped using Illumina 610 quad chips. Following standard QC we imputed additional genotypes (IMPUTE version 2.2.2) with the "1000 Genomes Phase I integrated variant set" reference genotypes; we excluded SNPs with INFO <0.4 and MAF <0.01. Genome-wide association study (GWAS) was performed for children that were skin prick positive to peanut at either at age 8 or 11 (46 cases, 554 controls). **Results:** We identified 66 regions that were associated with peanut sensitisation at genome-wide significance (p<5×10-8). Some of the most significantly association regions included the *CLDN14* (p=4×10-22), *DACH1* (p=7×10-23) and *CBFA2T2* (p= 8×10-23) genes. CLDN14 is part of the claudin family which is an important component of tight junctions. DACH1 is a chromatin-associated protein that helps regulate gene expression during development. *CBFA2T2* forms part of a fused gene that can be important in myeloid leukemia. *CBFA2T2* is a parologue of *DEAF1* which is reported to control the expression of genes encoding peripheral tissue antigens in type 1 diabetes. In drosophilia DEAF1 is reported to be a regulator of the innate immune response. Interestingly *CCL5* ($p=2\times10^{-8}$) was also associated with peanut sensitisation. CCL5 is a chemokine that is involved in immunoregulatory and inflammatory processes. **Conclusion:** This is the first GWAS of peanut sensitisation in children and we provide additional evidence that peanut sensitisation is heritable. We have identified several loci associated with peanut allergy that contain plausible candidate genes. Replication of these results will provide further evidence of these associations. Further studies will also be required using cohorts of individuals with either challenge-confirmed food allergy or a clear history of severe reactions linked to peanut consumption.

976S

A genome-wide association study identifies a LEPR gene as a novel predisposing factor for childhood FPG. *M. GO, J. HWNAG, L. Heo, T. Park, B. Kim.* KNIH, Osong, South Korea.

To date, genome-wide population-based association studies have been predominantly conducted on genetic predispositions in adult individuals. In uncovering childhood FPG susceptibility, a combined meta-analysis identified a novel LEPR locus (rs17407594) reaching genome-wide significance (n = 1,260, Padd = 4.98 ' 10-8). We also observed an association with T2D risk in the AGEN consortium (n = 18,817, Padd = 2.06 ' 10-2). In conclusion, our findings might expand understanding of genetic architecture contributing to glucose regulation and T2D risk.

977M

Comprehensive curation and visualization of ethnicity information from published genome-wide association studies (GWAS): an improved GWAS Catalog. *L.A. Hindorff[1], J.A.L. MacArthur[2], J. Morales[2], E.H. Bowler[2], P. Hall[1], K. Klemm[3], H. Junkins[1], T. Burdett[2], D. Welter[2], T. Manolio[1], H. Parkinson[2].* 1) Division of Genomic Medicine, NHGRI, NIH, Bethesda, MD; 2) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; 3) Division of Policy, Communication and Education, NHGRI, NIH, Bethesda, MD.

Genome-wide association studies (GWAS) have been prolific, with over 1,900 publications identifying 13,400 genetic variants associated with a broad range of human diseases and traits. The GWAS Catalog, available at http://www.genome.gov/gwastudies and http://www.ebi.ac.uk/fgpt/gwas/ , has served as a repository and visual summary of published GWAS findings since 2008. For each published GWAS study the Catalog includes publication, trait and SNP-trait association information, including SNP identifier, gene and risk allele, P-value, and a brief cohort description. To date the majority of GWAS studies have primarily been carried out in populations of European ancestry. This raises the question of how generalizable these trait-associations are across diverse populations. GWAS in diverse populations are now becoming more common, raising the importance of systematically curating detailed ancestry information to allow this question to be answered. Efforts to curate in greater detail the race/ancestry information from Catalog papers and to integrate this expanded information into a user-friendly interface will be described. Building upon an existing framework that relies largely on author-reported descriptors of ethnicity and ancestry, the more detailed ancestry data are based on systematic criteria and extracted as semi-structured fields. Information is curated at various levels of granularity - broad ancestral categories (e.g. “European”, “Hispanic/Latin American”), specific countries of origin and recruitment, and finer geographic descriptors where available (e.g. city and state). The public availability of these data will coincide with the release of an improved web interface that will enable visualization and searchability of ancestry data. A preview of the curation criteria and visualization of the ancestry data, with opportunity for detailed comment, will be presented and available at http://www.ebi.ac.uk/fgpt/gwas/ . The incorporation of detailed ancestry information, with associated search and visualization features, will enable users to investigate the generalizability of trait-associations across diverse populations and identify those limited to specific ancestries.

978T

Variants within ADAMTS9-AS2 influence fingerprint patterns. *Y.YW. Ho[1, 2], D.M. Evans[3, 4, 5], G.W. Montgomery[1], A.K. Henders[1], J.P. Kemp[3, 4, 5], N.J. Timpson[4, 5], B. St Pourcain[4, 5], D.Z. Loesch[6], G.D. Smith[4, 5], N.G. Martin[1], S.E. Medland[1].* 1) QIMR Berghofer, Herston, QLD, Australia; 2) School of Psychology, The University of Queensland, St. Lucia, QLD, Australia; 3) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK; 4) School of Social & Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol, UK; 5) University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Australia; 6) Department of Psychology, La Trobe University, Melbourne, Australia.

Dermatoglyphics (the scientific study of dermal ridges on the distal phalanges, palms, and soles of primates) is a complex quantitative trait, which provides a model for genetic investigation. The current study sought to identify the genetic influences on fingerprint patterns across all 10 fingers, utilizing genome-wide association and meta-analysis. Data were collected from three samples, consisting of 2296 participants from the QIMR Brisbane Adolescent Twin Study, 1859 from the QIMR health and lifestyle study, and 5339 from the Avon Longitudinal Study of Parents and Children (ALSPAC). Results of meta-analyses across samples identified an effect for *rs1523452*, within *ADAMTS9-AS2* for whorls on the little fingers (left, p = 3.43×10-27, r = .124; right, p = 6.25×10-15, r = .089) and ring fingers (left, p = 7.18×10-13, r = .083; right, p = 3.87 x 10-11, r = .076) of both hands. Post-hoc TATES multivariate analyses were also conducted on the QIMR and ALSPAC samples, which showed strong association at this locus (p = 2.51 ×10-08, p = 7.42 ×10-19 respectively). As genetic variants within *ADAMTS9* have previously been associated with type 2 diabetes and waist-hip ratio, these results suggest variants influencing prenatal growth, evidenced by their effect on formation of ridge patterns, have ongoing effects on later development.

979S

Genome-wide association and local ancestry analyses of high-altitude adaptations in Tibetans. *C. Jeong[1], B. Basnyat[2], G. Childs[3], S. Craig[4], D. Witonsky[1], C. Beall[5], A. Di Rienzo[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Oxford University Clinical Research Unit, Patan Hospital, Kathmandu, Nepal; 3) Department of Anthropology, Washington University in St. Louis, St. Louis, MO, USA; 4) Department of Anthropology, Dartmouth College, Hanover, NH, USA; 5) Department of Anthropology, Case Western Reserve University, Cleveland, OH, USA.

Indigenous human populations in the Tibetan plateau show a set of physiological traits distinct from those of acclimatized lowlanders, e.g. unelevated hemoglobin concentration (Hb) up to 4,000 m altitude and extremely low arterial blood oxygen saturation level (SaO_2). Several population genomic studies of Tibetans independently identified two candidate genes, *EGLN1* (egl nine homolog 1) and *EPAS1* (endothelial PAS-domain containing protein 1), which harbor extreme allele frequency divergence and signatures of positive selection. However, the genetic basis of the distinctive Tibetan high-altitude physiology remains poorly understood. We performed a genome-wide association study of two key physiological traits, Hb and SaO_2, in a group of 880 ethnic Tibetan women born and raised at altitudes ranging from 2,982 m to 4,052 m in Nepal. Both phenotypes were controlled for known covariates. No genome-wide significant association was found applying the Bonferroni correction for the 363,954 SNPs tested. However, SNPs around *EPAS1* gene are associated with Hb (linear mixed model (LMM) $p \geq 2.27\times10^{-5}$) and those around *HIF1A* are associated with SaO_2 (LMM $p \geq 2.72\times10^{-4}$), thus confirming a role for oxygen homeostasis systems in Tibetan adaptations. Next, we conducted a local ancestry analysis across 337 unrelated Tibetans, to find out loci with excess high-altitude ancestry represented by the Sherpa in this study. Both *EPAS1* and *EGLN1* show marked enrichment in high-altitude ancestry: 73.1% (+6.1 standard deviation; SD; the top signal) for *EPAS1* and 60.1% (+3.3 SD; top 0.2%) for *EGLN1*. The genome-wide mean high-altitude ancestry is 45.1% with SD = 4.6%. Last, we explored if SNPs with excess high-altitude ancestry were preferentially associated with high-altitude phenotypes. High-altitude ancestry proportion shows a marginally negative correlation with Hb association p-value (Spearman’s rank correlation p = 0.054). However, without *EPAS1* SNPs, this correlation becomes weaker (p = 0.100). Our study presents the first genome-wide association study for two physiological traits in Tibetans and our local ancestry analysis raises the possibility that phenotypes other than hemoglobin level and oxygen saturation are important in Tibetan physiology.

980M

GWAS meta-analysis of primary sclerosing cholangitis identifies new disease loci and further clarifies the genetic relationship with inflammatory bowel disease. *S. Ji[1], B.D. Juran[2], E. Melum[3], S. Mucha[4], J.Z. Liu[1], A. Franke[4], T.H. Karlsen[3], K.N. Lazaridis[2], C.A. Anderson[1] on behalf of the International Primary Sclerosing Cholangitis Study Group (IPSCSG).* 1) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom; 2) Center for Basic Research in Digestive Diseases, Division of Gastroenterology and Hepatology, Mayo Clinic, College of Medicine, Rochester, Minnesota, USA; 3) 3.Norwegian PSC Research Center, Department of Transplantation Medicine, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway; 4) Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany.

Primary sclerosing cholangitis (PSC) is a chronic biliary disease, with a 9-39 fold sibling relative risk and marked comorbidity with inflammatory bowel disease (IBD). Previous PSC genome-wide association studies (GWASs) have identified six non-HLA risk loci. Recently, the IPSCSG performed an Immunochip-based genetic association study across 3,789 PSC cases and 25,079 controls from North America and Europe that identified nine new non-HLA risk loci. Here, we report the most highly powered genome-wide study of PSC to date, combining GWAS data from 2,871 PSC cases and 12,019 controls from Europe and North America. Following quality control and genotype imputation, over 7 million SNPs genome-wide were tested for association to PSC. We found genome-wide significant evidence of association ($p<5\times10^{-8}$) for nine novel loci, as well as at previously identified PSC risk loci including the HLA. Newly associated loci include those previously reported to be associated with other autoimmune disorders and genes with known immune-related function such as FUT2, MANBA, and UBASH3A. Analyses comparing our data to those from the largest GWAS of IBD showed high genetic correlation (0.4-0.5, estimated from genome-wide SNPs) between PSC and ulcerative colitis (UC), whereas Crohn's disease (CD) had near zero correlation with PSC patients without IBD. Furthermore, we observed an increase in the genetic correlation between PSC and UC after removing SNPs within the HLA region, suggesting that the HLA contribution to risk differs between the two diseases. We next tested for allele frequency differences between our PSC-only cases and those with comorbid UC, identifying 3 loci of suggestive significance ($p=5\times10^{-6\sim-8}$). These, and 70 loci from our standard case-control analysis are currently being tested for association in an independent replication panel of around 2,000 cases and 4,000 controls, and the results will be presented. In summary, we have identified nine novel PSC risk loci and more are anticipated post-replication. As expected, five of the new risk loci have been previously associated with other immune-mediated diseases. We show that PSC is genetically distinct to IBD, but is genetically correlated with UC (but not CD). In addition, we suggest that the HLA may play the major role in determining the presence or absence of UC in individuals who have an increased risk of PSC.

981T

Integrated analysis of known height association signals with novel signals from an East Asian GWAS decodes the genetic architecture of height through *in silico* functional candidate prioritization and gene network analysis. *T.A. Johnson[1], K.A. Boroevich[1], T. Tanaka[2], L. Qi[3], M. He[4], M. Xu[5], T. Wu[4], M. Kubo[6], T. Tsunoda[1], Asian Genetic Epidemiology Network (AGEN) Consortium.* 1) Lab for Medical Science Mathematics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 2) Lab for Cardiovascular Diseases, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 3) Dept of Nutrition, Harvard School of Public Health, Boston, MA, USA; 4) MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science & Technology, Wuhan, Hubei, China; 5) Key Lab for Endocrine and Metabolic Diseases of Ministry of Health, Shanghai Clinical Center for Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Dept of Endocrinology and Metabolism, Ruijin Hospital Affiliated to; 6) Lab for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan.

Genome-wide association studies of human height have identified over 200 loci to-date, but the vast majority were found using samples of European ancestry. To expand our understanding of how height associated variants, genes, and networks are shared between ethnicities or restricted to particular ethnic groups, we analyzed 43,418 BioBank Japan (BBJ) samples and 7.8 million variants imputed using 1000 Genomes Project (1000G) reference haplotypes. We found 65 association signals with $P\leq5\times10^{-8}$, including 23 previously unreported loci, and replicated 14 of those signals using Asian Genetic Epidemiology Network (AGEN) samples (n=36,225). To further explore inter- and intra-ethnic group associated signals, we identified 251 known loci from 484 previously reported SNPs combined with 64,515 1000G variants that were in linkage disequilibrium (LD; $r^2>0.2$); 131 of those signals had evidence for association in the BBJ dataset. We extrapolated *P*-values for 1000G SNPs based on r^2 with Genetic Investigation of ANthropometric Traits (GIANT) consortium data, and found that ten of the BBJ replicated and three non-replicated loci showed evidence of association in the European ancestry data. LD cluster analysis and stepwise model selection for the 145 BBJ associated loci found evidence for secondary signals in 31 loci. To identify candidate causal SNPs and underlying genes, we intersected the variants from 265 known or novel loci with genomic annotations and classified them based on the level of evidence for protein-coding changes (PolyPhen2) or regulatory function (transcription factor binding/motifs, DNase hypersensitivity, promoter/3'-UTR/5'-UTR, micro-RNA target sites). Using association strength and functional class, we extracted 1,333 candidate causal SNPs and 470 genes; only 2.5% of candidate SNPs impacted coding sequence. To identify a smaller set of candidate causal genes, we used STRING-DB 9.1 data to perform a gene network randomization analysis and step-wise gene prioritization. The final gene network highlights an interconnectedness between a number of pathways related to growth and development including Estradiol synthesis and estrogen receptor signaling, Insulin-like growth factor/Growth hormone axis, and Indian Hedgehog signaling. These findings significantly add to our current understanding of genes and pathways that impact variation in human height, and likely represent variants that are important for other common human diseases and phenotypic variation.

982S

Genome-wide association study of comorbidity of alcohol and nicotine dependences. *J. Jung, H. Zhang, B. Grant.* Laboratory of epidemiology and biometry, National Institute on Alcohol Abuse and Alcoholism/National Institutes of Health, Bethesda, MD.

Alcohol and nicotine dependences are common psychiatric disorders that are often developed together. Both disorders are genetically influenced with heritability of 55% and 60% respectively, and co-occurrence of both lifetime dependences is influenced by a substantial genetic correlation between both disorders. We analyzed Study of Addiction: Genetics and Environment (SAGE) to perform a genome-wide association study of a comorbidity of both alcohol dependence (AD) and nicotine dependence (ND), which were diagnosed by the Diagnostic and Statistical Manual of Mental Disorders-IV. Samples in SAGE were genotyped by illumine Human 1M BeadChip Array. European American (EA) samples of 1180 cases and 1378 controls of AD and 1163 cases and 1392 controls of ND were analyzed. Correlated binary model with bivariate odd ratio were utilized to test association of co-occurrence with SNPs after controlling for age, sex and two population stratification scores. We found that rs12439549 in Gamma-aminobutyric acid (GABA) A receptor, gamma 3 (GABRG3) is associated with co-occurrence of both disorders (p-value =1.56e-05). We replicate the results by African American (AA) samples with 709 cases and 516 controls of AD and 659 cases and 692 controls of ND (p-value =9.5e-04). The gene GABRG3 is known to be linked with alcohol dependence. In addition we found that in both EA and AA groups the additional common 109 genes are associated with the comorbidity by a threshold of 10e-02. This finding is very significant to understand the etiology of the co-occurrence of both disorders.

983M

Mapping variation in response to vitamin D in the immune system. *S.N. Kariuki, C. Jeong, J. Maranville, S. Baxter, D. Witonsky, A. Di Rienzo.* Department of Human Genetics, University of Chicago, Chicago, IL, USA.

The active hormonal form of vitamin D, 1,25-dihydroxyvitamin D ($1,25(OH)_2D_3$) is a potent modulator of immune response. Inhibition of lymphocyte proliferation is one of the effects of $1,25(OH)_2D_3$ on the immune system and it varies across individuals. Our objective was to understand the genetic basis for variation in this cellular phenotype, which we refer to as Iv. We conducted a genome-wide association study (GWAS) of Iv in peripheral blood mononuclear cells (PBMCs) from 88 healthy African American individuals. They were genotyped at 511,618 single nucleotide polymorphisms (SNPs) using Illumina arrays, and additional genotypes were imputed using Impute2, resulting in a total of 3,952,978 SNP genotype data available for these individuals. The PBMCs were treated with $1,25(OH)_2D_3$ for 48 hours, and Iv was determined by subtracting proliferation in vehicle-treated cells from proliferation in $1,25(OH)_2D_3$-treated cells. Associations between SNPs and Iv were tested using a likelihood-ratio test. We also measured transcriptome-wide gene expression in PBMCs from the same individuals treated with $1,25(OH)_2D_3$ for 6 hours. *cis*-eQTL mapping was performed using the Bayesian Regression for Identifying Gene-Environment interactions (BRIdGE) program. We found strong signals of association between intergenic SNPs in chromosomes 5 (rs7724571, p=3.57×10-8) and chromosome 18 (rs7228926, p=4.97×10-8) with Iv. The top SNPs in chromosome 5 are near the chemokine (C-C motif) ligand 28 (*CCL28*) gene, which encodes a cytokine that plays an important role in the chemotactic activity of $CD4^+$ and $CD8^+$ T cells. We also found 17 SNPs which are significantly associated with expression of genes in response to $1,25(OH)_2D_3$ treatment. These genes include two pore segment channel 2 (*TPCN2*), which is expressed in multiple cell types and has been implicated in blonde versus brown hair pigmentation, and interleukin 4 receptor (*IL4R*), which is expressed in immune cells and bronchial epithelial cells, and has been implicated in asthma risk. In summary, we found that SNPs in chromosomes 5 and 18 are associated with inter-individual variation in the Iv phenotype, accounting for ~46% of this variation. We also found 17 *cis*-eQTLs associated with differential expression of genes in response to vitamin D treatment. Further functional studies of the top GWAS SNPs and the *cis*-eQTLS will enable us to elucidate how these genetic variants influence the action of $1,25(OH)_2D_3$ in the immune system.

984T

Comparing of GWAS data for the personality in four Korean cohort. *B. Kim[1], H. Kim[1], H. Cho[1], S. Roh[1], M. Lee[2], S. Yang[2], N.H. Cho[3], C. Shin[4], J. Sung[2], H. Kim[1].* 1) Department of Biochemistry, School of medicine, Ewha Womans University, Seoul, Korea; 2) Complex Disease and Genetic Epidemiology Branch, Department of Epidemiology and Institute of Environment and Health, School of Public Health, Seoul National University, Seoul, Korea; 3) Department of Preventive Medicine, Ajou University School, Suwon, Korea; 4) Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Korea University Hospital, Ansan, Korea.

Personality is a determinant of behavior and lifestyle associated with health and human diseases. Despite high heritability ranging from 33 to 60%, the understanding of the genetic origins of personality trait variation is extremely limited. To identify genetic variants associated with each of the five dimensions of personality, we performed a genome wide association (GWA) meta-analysis of four cohort, which are the rural Ansung (1,126 adults) and urban Ansan cohort (1,683 adults), healthy Twin Study (979 adults) and young women cohort (1,089 adults) in Korea. Personality traits were measured with the Revised NEO Personality Inventory for the five-factor model of personality, using the Korean short version of the original NEO-PI-R, a 90-item measure of the five factors of personality. Genomic DNA was extracted from whole-blood samples using a commercial isolation Kit according to the manufacturer's protocols. We used PLINK for the quality control procedure and PLINK, BEAGLE, SHAPEIT, and IMPUTE2 for SNP imputation, due to different chip platform. After imputation, we used PLINK and GenABEL for association analysis and METAL software for meta-analysis of Ansung and Ansan cohort (1,387,466 SNPs), young women cohort (1,581,609 SNPs), and healthy Twin Study (1,387,466 SNPs). There are no SNPs reached genome-wide thresholds of 5×10E-08 for statistical significance. In the result of this study, however, we found the association between Neuroticism and DRD1 (p= 1×10E-05), Extraversion and CDKAL1 (p=2×10E-06), Openness to experience and ERBB4 (p=4×10E-06), Agreeableness and FAM110B (p=1×10E-05), and Conscientiousness and IGF2BP3 (p=4×10E-07). Among five-factor personality, we speculated that Neuroticism could associate with D1 subtype of the dopamine receptor. It will explain the relationship of personality and some behavioral responses. In addition, Openness to experience could associate with cancer and Conscientiousness could associate with insulin-like growth factor II and RNA synthesis or metabolism. But these suggestions should be replicated and confirmed by further studies. [This research was supported by a grant of Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI14C0072)].

985S

Genome-wide association study (GWAS) of atopic dermatitis in Korean children. *K.W. Kim[1,2], R.A. Myers[1], J.H. Lee[3], EJ. Kim[4], D. Yoon[4], JS. Lee[4], KE. Kim[2], D.L. Nicolae[1], C. Ober[1], M.H. Sohn[2].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Pediatrics, Yonsei University College of Medicine, Seoul, Korea; 3) Department of Oral Biology, Yonsei University College of Dentistry, Seoul, Korea; 4) Research Center for Standardization of Allergic Disease, Allergy TF, Department of Immunology and Pathology, Korea National Institute of Health, Osong, Korea.

Atopic dermatitis (AD) is a heritable chronic inflammatory skin disease, which commonly presents during childhood when it is strongly associated with allergic sensitization. Previous GWAS of AD have not included Korean populations. Here, we conducted the first GWAS of moderate to severe AD in Korean children. Our study included 246 children with both AD and allergic sensitization (specific IgE to ≥1 food or airborne allergen) and 551 adult controls with a negative history of both AD and allergic sensitization (negative skin prick tests to common allergens). DNA from these individuals was genotyped using the Affymetrix Axiom or 5.0 chip; a set of common SNPs were imputed using minimac and the 1000 Genome Asian reference panel. 2.7 million SNPs remained after quality control (QC) checks, which included allele frequency comparisons between controls genotyped on different chips and inclusion of only high imputation accuracy (r^2>0.9) SNPs, in addition to standard QC criteria. Associations with SNPs at a locus on 13q21.31 were associated with AD at genome-wide levels of significance ($p<1.8\times10^{-8}$). This locus is in a gene desert and >1Mb from the closest genes, *PCDH9* and *PCDH20*. SNPs at four additional loci had p-values $<1\times10^{-6}$, including SNPs at 15q24.3 that are eQTLs for *PSTPIP1* in monocytes after treatment LPS (Fairfax BP et al. Science2014;343:1246949) and at 6q22.33 in the *THEMIS* gene. *PSTPIP1* is involved in aberrant inflammation in the skin, a primary feature of AD, and *THEMIS* modulates regulatory T cell function, a key cell type in inflammatory responses. We also identified SNPs at 2p24.3 in the *NBAS* gene and at 10p14, and replicated previous AD associations with SNPs in *KIF3A* (chr5q31.1, p=0.002), *RAD50* (5q31.1, p=0.03) and *ZNF365* (10q21.2, p=0.03). In summary, our study of Korean children identified new susceptibility loci for AD that are related to the potential dysregulation of immune cells in this disease, and extends our understanding of the pathophysiology of AD.

986M

The genetic landscape of pediatric autoimmune diseases. *Y.R. Li[1,2], J. Li[1], J.A. Ellis[3], S. Kugathasan[4], M.L. Becker[5], A. Latiano[28], E. Perez[7], R.K. Russell[8], D.C. Wilson[9], M.S. Silverberg[10], V. Annese[6], B.A. Lie[11], M. Punaro[12], M.C. Dubinsky[13], C. Strisciuglio[14], A. Staiano[14], E. Miele[14], C. Wise[15], H. Chapel[16], C. Cunningham-Rundles[17], J.S. Orange[18], A.M. Griffiths[19], J. Satsangi[20], T. Finkel[21], C. Polychronakos[22], R.N. Baldassano[23,24], E.T. Luning Prak[25], H. Li[26], B.J. Keating[1,23], H. Hakonarson[1,23,27].* 1) The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Medical Scientist Training Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Genes, Environment and Complex Disease, Murdoch Children's Research Institute, Parkville, Victoria, Australia; 4) Department of Pediatrics, Emory University School of Medicine and Children's Health Care of Atlanta, Atlanta, GA, USA; 5) Division of Clinical Pharmacology, Children's Mercy, Kansas City, MO, USA; 6) Unit of Gastroenterology, Department of Medical and Surgical Specialties, Careggi University Hospital, Florence, Italy; 7) Division of Pediatric Allergy and Immunology, University of Miami Miller School of Medicine, Miami, FL; 8) Yorkhill Hospital for Sick Children, Glasgow, Scotland; 9) Paediatric Gastroenterology and Nutrition, Royal Hospital for Sick Children, Edinburgh and Child Life and Health, University of Edinburgh, EH9 1UW, UK; 10) Mount Sinai Hospital IBD Centre, University of Toronto, 441-600 University Avenue, Toronto, Ontario M5G 1X5, Canada; 11) Department of Immunology, Oslo University Hospital, Rikshospitalet, 0027 Oslo, Norway; 12) Texas Scottish Rite Hospital for Children, Dallas, TX, USA; 13) Department of Pediatrics, Pediatric IBD Center, Cedars Sinai Medical Center; 14) Department of Translational Medical Science, Section of Pediatrics, University of Naples "Federico II", Naples, Italy; 15) Sarah M. and Charles E. Seay Center for Musculoskeletal Research, TX Scottish Rite Hospital for Children, Dallas, TX, USA; 16) Nuffield Department of Medicine, University of Oxford, UK; 17) Institute of Immunology and Department of Medicine, Mount Sinai School of Medicine, New York, NY; 18) Section of Immunology, Allergy, and Rheumatology, Department of Pediatric Medicine, Texas Children's Hospital, Houston, TX; 19) The Hospital for Sick Children, University of Toronto, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada; 20) Gastrointestinal Unit, Division of Medical Sciences, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK; 21) Department of Pediatrics, Nemours Children's Hospital, Orlando, FL, USA; 22) Departments of Pediatrics and Human Genetics, McGill University, Montreal H3H 1P3, Quebec, Canada; 23) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 24) Division of Gastroenterology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 25) Department of Pathology and Lab Medicine, Perelman School of Medicine University of Pennsylvania, Philadelphia, PA, USA; 26) Department of Biostatistics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 27) Division of Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 28) 'IRCCS 'Casa Sollievo della Sofferenza', Division of Gastroenterology, San Giovanni Rotondo, Italy.

Autoimmune diseases affect seven to ten percent of individuals living in the Western Hemisphere, and represent a significant cause of chronic morbidity and disability. High rates of comorbidity and familial clustering suggest that strong genetic predisposition underlies autoimmune disease susceptibility, and genome wide association studies (GWAS) have identified hundreds of susceptibility genes associated with autoimmune diseases with some shared across clinically-distinct disease groups. To investigate the genetic architecture of pediatric autoimmune diseases (pAIDs), we performed a heterogeneity-sensitive GWAS (hsGWAS) across 10 pAIDs in a nested case-control study including over 5,200 cases and 11,000 controls. We identified 86 independent pAID association loci reaching GWS ($P < 5 \times 10\text{-}8$), including lead SNPs mapping to candidate genes with established immunoregulatory functions (e.g., CD40LG; $P < 3.08 \times 10\text{-}11$ and NFATC3; $P < 1.18 \times 10\text{-}8$). Of the 147 lead GWS (86) and marginally significant (61) loci, 97% were supported by functional (n=30), regulatory (n=55), conservational (n=30) or literature-reported (n=40) data, which is enriched as compared to that observed at random across the genome ($p<0.021$), particularly for DNAse hypersensitivity sites ($p < 0.01$) as well as for genetic association annotations, eQTLs, and coding variants ($p<0.001$). In addition, we extensively characterized the expression profiles of the candidate genes mapped by the lead loci in human tissues and murine immune-specific cell typing, providing evidence to support a disease-specific gene expression signature across subsets of immune cell lineages. Integration of multiple in silico analytical approaches identified highly shared autoimmune signals (e.g., IL2-IL21 $P < 6.24 \times 10\text{-}12$) and converging roles for JAK-STAT, innate, and TH1-TH2/TH17 mediated T-cell signaling across pAIDs and as molecular pathways representing attractive pharmacological targets for pAIDs.

987T

Genomewide associaiton study identifies six novel loci associated with plasma carnitine levels. *H. Li, Y. Hu, R. Zeng, X. Lin.* Chinese Academy of Sciences, Shanghai Institutes for Biological Sciences, Shanghai, China.

Carnitine, a nutrient that plays important role in energy and fatty acid metabolism, has recently been showed to have protective effect on type 2 diabetes, obesity and cardiovascular diseases in both animal studies and clinical trials. Carnitine is mainly synthesized by the liver and the kidneys from the essential amino acids lysine and methionine. To identify genetic variants that affect circulating levels of various types of carnitines, a genomewide association study was performed in population of Chinese Hans with 2228 individuals from the Nutrition and Health of Aging Population in China (NHAPC) cohort. Association analyses were performed in Beijing and Shanghai subpopulations separately and summary statistics were calculated using inverse variance weighted fixed effect meta-analyses. We identified five novel loci at MCCC1, CPS1, FADS1-FADS2, DLG4 and LRP8-CPT2 that were associated with carnitines at genomewide significance. Of which, common variants in MCCC1, CPS1, FADS1-FADS2 and DLG4 were significantly associated with plamsa levels of C5OH ($P=1.47\times10\text{-}13$), C6DC ($P=3.78\times10\text{-}10$) and C20:4(n-6) ($P=6.80\times10\text{-}18$), and C12/C10 ratio ($P=7.96\times10\text{-}11$), respectively, while genetic variants in LRP8-CPT2 were significnatly associated with both lower C6DC ($P=7.22\times10\text{-}10$) and higher C16 ($P=1.06\times10\text{-}12$) levels. We observed an independent locus (rs7516477) at ACADM that showed associations with C8 ($P=6.51\times10\text{-}14$) level and C12/C10 ratio ($P=6.92\times10\text{-}21$), and is not in linkage disequilibrium (LD) ($r2\leq0.018$).with the previously reported genetic variants (rs11163924 and rs211718) in ACADM in populations of European origin. We also confirmed the previously reported associations between ACADS-rs2066938 and C4 level ($P=8.97\times10\text{-}20$) and between ACADS-rs2014355 and C3/C4 ratio ($P=9.20\times10\text{-}89$) in our study. In conclusion, we have identified six novel loci that were associated with various types of carnitines at genomewide significance in this study of Chinese Hans, and also confirmed the previously reported ACADS locus for C4-carnitine. Our findings provide novel insight into the genetic basis of carnitine biosynthesis.

988S

***CLEC16A* associates with human common variable immunodeficiency and influences murine B cell survival and function.** *J. Li[1], S.F. Jørgensen[2], S.M. Maggadottir[1,3], M. Bakay[1], K. Warnatz[4], J. Glessner[1], U. Salzer[4], R.E. Schmidt[5], E. Resnick[6], S. Goldacker[4], M. Buchta[4], T. Witte[5], L. Padyukov[7], F. Atschekzei[5], J.T. Elder[8,9], R.P. Nair[8], K.E. Sullivan[3,10], J.S. Orange[11], B. Fevang[2,12], S. Schreiber[13], W. Lieb[14], P. Aukrust[2,12], H. Chapel[15], C. Cunningham-Rundles[6], A. Franke[13], T.H. Karlsen[2,16,17], B. Grimbacher[4], H. Hakonarson[1,18], L. Hammarström[19], E. Ellinghaus[13].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, USA; 2) K.G. Jebsen Inflammation Research Centre, Research Institute of Internal Medicine, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway; 3) Division of Allergy and Immunology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 4) Center for Chronic Immunodeficiency, University Hospital of Freiburg, Freiburg, Germany; 5) Clinic for Immunology and Rheumatology, Hannover Medical School, Hannover, Germany; 6) Institute of Immunology and Department of Medicine, Mount Sinai School of Medicine, New York, , USA; 7) Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden; 8) Department of Dermatology, University of Michigan, Ann Arbor, Michigan, USA; 9) Ann Arbor Veterans Affairs Hospital, Ann Arbor, Michigan, USA; 10) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 11) Section of Immunology, Allergy, and Rheumatology, Department of Pediatric Medicine, Texas Children's Hospital, Houston, TX, USA; 12) Section of Clinical Immunology and Infectious diseases, Oslo University Hospital Rikshospitalet, Norway; 13) Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Germany; 14) Institute of Epidemiology and Biobank popgen, Christian-Albrechts-University of Kiel, Kiel, Germany; 15) Department of Clinical Immunology, Nuffield Department of Medicine, University of Oxford, UK; 16) Norwegian PSC Research Center, Division of Cancer, Surgery and Transplantation, Oslo University Hospital, Oslo, Norway; 17) Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 18) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 19) Department of Laboratory Medicine, Division of Clinical Immunology and Transfusion Medicine, Karolinska University Hospital, Huddinge, Stockholm, Sweden.

With a prevalence of 1:25,000 common variable immunodeficiency disorder (CVID) is a clinically important form of primary immune deficiency, where inadequate quantity and quality of immunoglobulins results in susceptibility to bacterial infections. CVID is a heterogeneous disorder with variable comorbidities and the underlying genetic mechanisms are poorly understood. A previous genome-wide association study reported association between human leukocyte antigen (*HLA*) complex and CVID. To explore whether further common genetic variants outside the *HLA* complex contribute to CVID susceptibility, we conducted an association analysis of 123,127 common single nucleotide polymorphisms (SNPs) in a cohort of 778 cases and 10,999 healthy controls of European ancestry through dense genotyping of immune-related loci on the immunochip (iCHIP). In this largest CVID genetic study performed to date, we replicated the *HLA* association and further identified the first non-HLA CVID risk locus at *CLEC16A* (rs17806056, P = 2.0×10-9). Although the function of *CLEC16A* is largely unknown, it encodes a C-type lectin-like domain protein that is expressed in B-cells, dendritic cells and natural killer cells. Because of the critical role of B-cell function in CVID development, we further examined the B cell phenotype in Clec16a knock down (KD) mice. We found Clec16a KD mice manifested a reduced percentage of B-cells and increased IgM secretion. Taken together, our results suggest *CLEC16A* affects B-cell survival and function. The reported association of *CLEC16A* with multiple autoimmune disorders presents a compelling link between these disorders and autoimmunity commonly seen in CVID .

989M

Common variants near ABCA1, AFAP1 and GMDS confer risk of primary open-angle glaucoma. *S. Macgregor[1], P. Gharahkhani[1], R. Fogarty[2], S. Sharma[2], A.W. Hewitt[4], S. Martin[2], MH. Law[1], K. Cremin[5], J.N. Cooke Bailey[6,7], S.J. Loomis[8], L.R. Pasquale[8,9], J.L. Haines[6,7], M.A. Hauser[10], A.C. Viswanathan[11], P. McGuffin[12], F. Topouzis[13], P.J. Foster[11], R.A. Mills[2], J.J. Wang[15], G.W. Montgomery[1], N.G. Martin[1], G. Radford-Smith[1,16], M.A. Brown[5], J.L. Wiggs for NEIGHBORHOOD cons.[8], D.A. Mackey[3,17], P. Mitchell for WTCCC cons.[15], K.P. Burdon[3,14], J.E. Craig[2].* 1) QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia; 2) Department of Ophthalmology, Flinders University, Adelaide, SA 5042, Australia; 3) Menzies Research Institute Tasmania, University of Tasmania, Hobart, TAS, 7000, Australia; 4) Centre for Eye Research Australia (CERA), University of Melbourne, Royal Victorian Eye and Ear Hospital, Melbourne, Victoria, Australia; 5) University of Queensland Diamantina Institute, Translational Research Insitute, Princess Alexandra Hospital, Brisbane, QLD 4102, Australia; 6) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, Tennessee, USA; 7) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio, USA; 8) Department of Ophthalmology, Harvard Medical School and Massachusetts Eye and Ear Infirmary, Boston, Massachusetts, USA; 9) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; 10) Departments of Medicine and Ophthalmology, Duke University Medical Center, Durham, North Carolina, USA; 11) NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; 12) MRC Social Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, King's College, De Crespigny Park, London, UK; 13) Department of Ophthalmology, School of Medicine, Aristotle University of Thessaloniki, AHEPA Hospital, Thessaloniki, Greece; 14) School of Medicine, Flinders University, Adelaide, SA 5001, Australia; 15) Centre for Vision Research, Westmead Millennium Institute for Medical Research, Westmead, NSW 2145, Australia; 16) School of Medicine, University of Queensland, Herston Campus, Brisbane, QLD, Australia; 17) Centre for Ophthalmology and Visual Science, Lions Eye Institute, University of Western Australia, Perth, Australia.

Primary open-angle glaucoma (POAG) is a major cause of irreversible blindness worldwide. We performed a genome-wide association study in an Australian discovery cohort comprising 1,155 advanced POAG cases and 1,992 controls. Association of the top SNPs from the discovery stage was investigated in two Australian replication cohorts (total 932 cases, 6862 controls) and two US replication cohorts (total 2616 cases, 2634 controls).

Meta-analysis of all cohorts revealed three novel (genome-wide significant) loci associated with development of POAG. These loci are located upstream of *ABCA1* (rs2472493 [G] OR=1.31, P= 2.1×10^{-19}), within *AFAP1* (rs4619890 [G] OR=1.20, P= 7.0×10^{-10}) and within *GMDS* (rs11969985 [G] OR=1.31, and P= 7.7×10^{-10}). Additionally there was strong evidence for association at SNP rs2276035 within the gene *ARHGEF12* in some samples, although heterogeneity of effect in other samples meant this result did not reach genome-wide significance overall. Using RT-PCR, we showed that all three genomewide significant genes are expressed within both human retina and other ocular tissues. For *ABCA1* and *AFAP1*, using immunolabelling we demonstrated protein distribution in the eye consistent with the expression findings.

ABCA1 is a membrane-bound receptor involved in phospholipid and cholesterol efflux from cells. ABCA1 has been reported to regulate neuroinflammation and neurodegeneration through co-ordinated activity in various cell types in mouse brain and it may be involved in glaucoma through a similar function in the retina. *AFAP1* encodes a protein that binds to actin filaments and allows their crosslinking. Actin cytoskeleton-modulating signals have been shown to be involved in the regulation of aqueous outflow and intraocular pressure, which are important parts of glaucoma pathogenesis.

These new loci, in addition to the previously identified risk loci, will help improve risk profiling for glaucoma with better opportunities for management of high-risk individuals.

990T

CTNNA3 and SEMA3D: Promising loci for asthma exacerbation identified through multiple cohorts. *M. McGeachie[1], A.C. Wu[1], S.M. Tse[1], G.L. Clemmer[1], J. Sordillo[1], P. Weeke[2], C. Shaffer[2], J. Denny[2], D.M. Roden[2], B.A. Raby[1], S.T. Weiss[1], K.G. Tantisira[1].* 1) Channing Div Network Med, Dept of Medicine, Brigham and Women's Hospital, Boston, MA; 2) Office of Personalized Medicine, Vanderbilt University School of Medicine, Nashville TN.

INTRODUCTION Asthma affects 300 million people worldwide and asthma exacerbations are a major cause of morbidity and medical cost. The objective of this study was to identify genomic predictors of exacerbations in subjects with asthma. METHODS In the Childhood Asthma Management Program (CAMP) and Childhood Asthma Research and Education (CARE) trials, acute asthma exacerbation was defined by treatment with a five-day prednisone course. We first performed a GWAS in CAMP, and to verify that the top hits from the GWAS were not artifacts of a non-normally distributed phenotype, we further performed permutation testing on the top 50 GWAS SNPs. We combined p-values from the CAMP permutation tests with the GWAS results from 205 participants of CARE. From BioVU, the DNA biobank for Vanderbilt University Medical Center, we identified a cohort of 786 subjects who had asthma, with 191 experiencing an asthma-related exacerbation, defined as a hospitalization or emergency department visit. We used CD4+ lymphocyte genome-wide mRNA expression profiling from the ABRIDGE project to identify associations of top SNPs with mRNA abundance of nearby genes. RESULTS In CAMP and CARE we identified a locus in CTNNA3 reaching genome-wide significance (rs7895190, p = 4.44E-09); we also prioritized other top hits for replication (rs746578 in TTN, p = 2.15E-06; rs993312 in SEMA3D, p = 7.63E-04). In replication in BioVU, rs993312 was significantly associated with exacerbations (p = 0.008). Rs746578 was in strong LD with a variant, rs10997296, identified as an eQTL for CTNNA3 in CD4+ cells from the ABRIDGE cohort (p = 0.00079). DISCUSSION The CTNNA3 SNP, rs7915690, is intronic. It is in strong LD (r2 = 0.97) with a predicted regulatory variant, rs10997296, that is a strong enhancer in human mammary epithelial cells (ENCODE); and in a DNAse I hypersensitivity site in small airway epithelial cells (ENCODE). This variant was identified as an eQTL for CTNNA3. The SEMA3D SNP, rs993312, is intronic and in strong LD (r2 = 0.99) with a regulatory SNP, rs55834466. The ENCODE project identifies rs55834466 as a strong enhancer in epidermal keratinocytes and human mammary epithelial cells, in addition SEMA3D is relatively overexpressed in the lung (Illumina Human BodyMap). CONCLUSIONS We identified two exacerbation-associated regulatory SNPs from GWAS of asthma clinical trials. One met genome-wide significance thresholds while the other replicated in a clinical Biobank database.

991S

Genome-wide association study of exfoliation syndrome/exfoliation glaucoma in a Japanese population. *M. Nakano[1], Y. Ikeda[2], Y. Tokuda[1], H. Adachi[1], M. Ueno[2], K. Imai[2], R. Sato[1], N. Omi[1], K. Mori[2], S. Kinoshita[2], K. Tashiro[1].* 1) Dept Genomic Med Sci, Kyoto Prefectural Univ Med, Kyoto, Japan; 2) Dept Ophthal, Kyoto Prefectural Univ Med, Kyoto, Japan.

Purpose: The common variants in lysyl oxidase-like 1 gene (*LOXL1*) at chromosome 15q24.1 are well known to be associated with exfoliation glaucoma (XFG) patients developed through exfoliation syndrome (XFS[MIM 177650]). However, LOXL1 by itself seemed to be insufficient for explaining the molecular mechanism of XFG pathogenesis. Moreover, the risk allele of a variant (rs1048661) in exon 1 of *LOXL1* has found to be inverted between Asian and Caucasian populations, suggesting that the variants in other genes should also be contributing to the genetic heritability of XFS/XFG. Therefore, we newly conducted a genome-wide association study (GWAS) using a Japanese population in an attempt to discover genetic markers for XFS/XFG. **Methods:** We first conducted a GWAS using the Japanese subjects and analyzed 652,792 autosomal common single-nucleotide polymorphisms (SNPs) in 201 XFS/XFG patients and 697 controls. We then replicated the results of GWAS using an independent population consisted of 121 XFS/XFG patients and 263 controls. We also performed a conditional analysis to analyze the combinational effect of the variants identified from the different genes. We finally confirmed the expression of the identified genes in human ocular tissues. **Results:** As a result of GWAS, 34 genome-wide significant SNPs ($P = 5.56 \times 10^{-8}$ - 3.46×10^{-54}) were indeed identified as a cluster only from chromosome 15q24.1. However, the significant SNPs were distributed in not only *LOXL1* but also two other adjacent genes in the locus. The significance of these SNPs was confirmed by the replication analysis. We also identified a suggestive association between the variants of one of the two genes and *LOXL1* by the conditional analysis. Both two genes were expressed in the ocular tissues. **Conclusions:** We have successfully identified some novel variants associated with XFS/XFG. The results suggested that the combination of newly discovered variants in these genes would be useful for precise XFG risk assessment especially in Asian population, as well as for elucidating the molecular mechanism of XFG pathogenesis through XFS.

992M

Genome-wide association study identifies two distinct risk haplotypes at LBX1, and links ITPR1 and SOX5 to adolescent idiopathic scoliosis. *L. Nelson[1], R. Chettier[2], J.W. Ogilvie[2], H.M. Albertsen[2], K. Ward[2].* 1) Affiliated Genetics, Salt Lake City, UT; 2) Juneau Biosciences, Salt Lake City, UT.

Background: Adolescent idiopathic scoliosis (AIS) is a common pediatric disorder that affects 2-4% of children. Genetic factors have been suspected for decades and recent studies have identified possible proteins involved in the condition. In spite of studies showing high heritability in AIS, the genetic risk factors for susceptibility to AIS across different ethnic groups remain largely unknown. A few AIS related genome-wide association study studies (GWAS) in Asian populations have revealed promising candidates. In particular, a recent genome-wide association study conducted in a Japanese population has identified LBX1 as a strong susceptibility locus for AIS. In this study we explore genetic association to AIS in subjects of European descent. **Methods:**A GWAS comparing 853 female Caucasian patients with adolescent idiopathic scoliosis with 1,368 ethnically matched unaffected females from the general population. **Results:** Using stringent quality filters we replicated the LBX1 association and identified two novel loci, ITPR1 and SOX5, associated with AIS at a genome-wide significance level below 5×10^{-8}. Our strongest association was found on LBX1 locus [rs11190878 (OR= 0.59,p=2.23×10^{-14}) and rs7893223 (OR=0.63,p=1.01×10^{-10})]. Using imputation, the previously reported marker rs11190870, also showed association (OR=0.65, p=3.64×10^{-11}) with AIS in this study. In haplotype analysis, the LBX1 haplotype ATT comprising the risk alleles for rs11190878, rs11190870 and rs7893223 showed the strongest association to AIS (OR=1.62, p= 9.48×10^{-15}) as a recessive model. The GCC haplotype was a protective haplotype for the same markers (OR=0.5, p=2.03×10^{-11}). These results suggest that the mutations on LBX1 loci are quite old and LBX1 gene is a potential susceptibility loci for AIS across many populations. We also identified ITPR1 and SOX5 as novel candidates for AIS. Together, LBX1, ITPR1 and SOX5 explain 4.2% of the phenotypic variance. **Conclusion:** This is the first large GWAS study of European descent in AIS to show a significant effect of a risk haplotype that spans the LBX1 locus and extends into the LBX1 antisense RNA1 locus. In addition, we report ITPR1 and SOX5 as new candidate risk loci for AIS patients of European descent. These markers identified will help in elucidating the pathogenesis of AIS.

993T

Genome-wide association of 44,714 subjects of African ancestry imputed to the 1000 Genomes reference panel identified two novel loci influencing body mass index. *M.C.Y. Ng[1], M. Graff[2], A. Justice[2], Y. Lu[3], P. Mudgal[1], K. Rand[4], Y. Li[5], B.E. Cade[6], J. Brody[7], M.K. Wojczynski[8], L.R. Yanek[9], J. Smith[10], M.A. Nalls[11], L.A. Lange[5], S. Vedantam[12], X. Guo[13], D. Siscovick[7], S.R. Patel[6], B.M. Psaty[7], I.B. Borecki[8], D.M. Becker[9], L.F. Bielak[10], Y. Liu[14], J.G. Wilson[15], J.N. Hirschhorn[16], J.I. Rotter[13], C.A. Haiman[4], R.J.F. Loos[3], K.E. North[2], African Ancestry Anthropometry Genetic Consortium.* 1) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA; 2) Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 3) The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 4) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 5) Departments of Genetics, Biostatistics, Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 6) Division of Sleep Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 7) Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA; 8) Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO, USA; 9) The GeneSTAR Research Program, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD, USA; 10) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA; 11) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 12) Divisions of Endocrinology and Genetics and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA; 13) Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; 14) Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA; 15) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA; 16) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA.

Obesity is a major public health problem with disproportionally high prevalence in African Americans. Our previous genome-wide association (GWA) meta-analysis in individuals of African ancestry, imputed to the HapMap 2 reference panel, confirmed 5 loci previously established in European ancestry populations and identified 2 novel loci for association with body mass index (BMI). As individuals of African ancestry display greater genetic variation than European or Asian ancestry populations, we hypothesized that a GWA study imputed to the 1000 Genomes reference panel might prove particularly useful for identifying additional variants associated with BMI in African ancestry individuals. As such, The African Ancestry Anthropometry Genetic Consortium (AAAGC) imputed data to the 1000 Genomes Phase 1 cosmopolitan reference panel in 44,714 individuals of African ancestry from 19 GWA studies. Following single variant association for BMI in individual studies under an additive model, meta-analyses were performed for 22 million variants using the fixed-effects inverse variance method. Analyses were performed in all individuals combined as well as stratified by sex. We observed genome-wide significant (*P*<5E-8) associations at 7 established BMI loci (near *SEC16B*, *TMEM18*, *GNPDA2*, *GALNT10*, *KLHL32*, *FTO* and *MC4R*) and 2 novel loci (near *PFKP* and *SYCP2L*) in the overall sample, as well as 2 novel loci in sex-stratified analyses (near *INTS10* in men, near *LRFN5* in women). Among the established loci, the index SNPs in this study were in strong linkage disequilibrium (LD), measured in r^2, with the index SNPs in both African and European ancestry studies imputed to the HapMap reference panel, enable fine mapping of associated signals shared between the two populations. Among the novel loci except for *LRFN5*, all index SNPs are of low-frequency (minor allele frequency 1% - 6%), were not present in the HapMap reference panel, and have even lower frequencies (0% - 0.6%) in other non-African ancestry populations. *PFKP* encodes the platelet isoform of phosphofructokinase that is a key regulator in glycolysis. Follow-up analyses to fine-map loci and identify causal genes and variants are ongoing. Our results suggest that a high-density reference panel including low-frequency variants improves the coverage of the genome and unmasks novel loci that may be missed due to lower LD in African ancestry populations. Further replication studies are required to confirm our novel findings.

994S

The Genetics of Erectile Dysfunction Risk in Men with Type 1 Diabetes. *M.R. Palmer[1], J.M. Hotaling[2], A.D. Paterson[3], H. Wessells[1].* 1) Urology, University of Washington, Seattle, WA; 2) Urology, University of Utah, Seattle, WA; 3) Genetics & Genome Blology, University of Toronto, Toronto, Canada.

Erectile dysfunction (ED) affects greater than 40% of men over 40 years old. Major risk factors include age, hypertension, and heart disease. However, there is also a genetic component to ED risk. A twin study showed that ED has 30% heritability. In addition, several small genome-wide association studies have linked genetic variants to increased risk of ED in men with diabetes, as well as in patients who have undergone radiation therapy for prostate cancer. We have ED phenotype data for men from a cohort of men with type 1 diabetes. Longitudinal data includes a yearly single-item survey question over 20 years, and a more detailed follow up study at year 17. A pilot GWAS in this population at year 10 showed association of ED with two SNVs near ALCAM (activated leukocyte cell adhesion molecule). Over the 20 years, 445 individuals had at least one report of ED, and 290 had none. In this population, ED is significantly associated with age, BMI, and HBA1C (all $p < 0.01$, t-test). We plan to use genetic data and new phenotype information from these men to replicate the association of variants that have previously been found to be significant or sub-significant. We will perform a cross-sectional case-control analysis, as well as a time-to-event analysis using a Cox proportional hazards model. The models will stratify for BMI, age, HBA1C, and other factors that associate with ED status, and use PCA to correct for ancestry. Identification of variants associated with ED risk will be useful for prediction and early treatment in diabetic men.

995M

Common genetic variation explains a substantial fraction of nicotine and cotinine glucuronidation in multiple populatons. *Y.M. Patel[1], D.O. Stram[1], L. Le Marchand[3], S.S. Hecht[2], C.A. Haiman[1], S.E. Murphy[2].* 1) Department of Preventive Medicine and Norris Comprehensive Cancer Center Keck School of Medicine University of Southern California Los Angeles, CA 90032; 2) University of Minnesota VFW Cancer Research Center 406 Harvard Street SE Minneapolis, MN 55455; 3) University of Hawaii Cancer Center 701 Ilalo St #600, Honolulu, HI 96813.

Evaluating the role of genetics in the metabolism of nicotine may be important in understanding racial/ethnic differences in lung cancer risk among smokers. Noted racial disparities in risk may be related to differences in internal carcinogen dose resulting from faster metabolism including the glucuronidation of nicotine and cotinine. We conducted a genome-wide association study (GWAS) in search of common genetic variants that may be predictive of nicotine and cotinine glucuronidation in a multiethnic sample of 2,239 current smokers representing 5 ethnic populations. The phenotypes analyzed were cotinine and nicotine glucuronidation. Cotinine glucuronidation is a ratio derived from the difference of total and free cotinine, divided by total cotinine. Nicotine glucuronidation is similarly created. In an analysis of 11,892,802 genotyped and imputed variants appearing in the thousand genomes project, 1,241 were found to be strongly associated with cotinine glucuronidation at $p<5\times10\text{-}8$. The vast majority were within the region *4q13*, near the previously reported gene *UGT2B10*. Fifteen different SNPs in 9 different regions were globally significant and contributed independent information genome-wide ($p<1\times10\text{-}3$ in stepwise regression). Together, these 15 SNPs explain 33.2% of variation observed in cotinine glucuronidation which ranged from 55% for African Americans to 19% among Japanese Americans. The strongest single SNP association was rs115765562 ($p= 1.60\times10\text{-}155$), near *UGT2B10*, which is highly correlated with a known splice site variant, rs116294140, in *UGT2B10*. This splice variant in combination with an Arg67Try nonsynonymous SNP rs61750900 explain 24.3% of variation in cotinine glucuronidation, indicating effects seen through putative functional missense and splice site variants might be the underlying explanation for the associations. Our analysis has determined that a high fraction of individual variation of cotinine glucuronidation is explained by genetic differences. We also noted that in our data African Americans and Japanese Americans have the lowest cotinine glucuronidation levels and they also tend to smoke fewer cigarettes per day and have lower nicotine equivalents than do Europeans. However, their lung cancer risks (higher in African Americans and lower in Japanese Americans) are quite different. This observation therefore does not strongly support an important role of glucuronidation in determining the factors underling these risk differences.

996T

Search for new risk gene for Stevens-Johnson Syndrome independent of *HLA* risk allele. *H. Sawai[1], M. Ueta[2, 3], Y. Hitomi[1], C. Sotozono[2], S. Kinoshita[2], K. Tokunaga[1]*. 1) Department of Human Genetics, The University of Tokyo, Tokyo, Japan; 2) Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan; 3) Research Center for Inflammation and Regenerative Medicine, Faculty of Life and Medical Sciences, Doshisha University, Kyoto, Japan.

Stevens-Johnson syndrome (SJS) and its severe form, toxic epidermal necrolysis (TEN), are acute inflammatory vesiculobullous reactions of the skin and mucous membranes including the ocular surface, oral cavity, and genitals. These reactions are very rare but are often associated with inciting drugs and infectious agents. We previously revealed the association between a specific *HLA-A* allele (*HLA-A*02:06*) and SJS/TEN with severe ocular surface complications (SOC) in Japanese population. We also performed genome-wide association study (GWAS) using Japanese SJS/TEN with SOC cases and healthy controls, and found that the *HLA-A* region showed the strongest association with susceptibility to SJS/TEN with SOC (odds ratio [OR] = 4.40, $p = 3.5 \times 10^{-14}$). To identify host genetic factor(s) contributing independently of *HLA-A*02:06* to SJS/TEN with SOC, we conducted a GWAS with *HLA-A*02:06* negative individuals.

We genotyped a total of 820 samples including 118 Japanese SJS/TEN patients and 702 Japanese healthy controls using the Affymetrix AXIOM Genome-Wide ASI 1 Array. We excluded one case sample with an overall call rate less than 97%, and recalled the remaining 819 samples by using Genotype Console v4.1.4 software. All samples used for GWAS passed a heterozygosity check, and five related samples identified by descendent testing were excluded. A principal component analysis found six outliers to be excluded by the Smirnov-Grubbs test, and we showed that all 117 cases and 691 controls formed a single cluster with the HapMap JPT samples but not with CHB samples. We further excluded samples with *HLA-A*02:06* and finally used 59 cases and 597 controls for the GWAS. The chi-square test was applied to an allele frequency model.

Four loci showed genome-wide marginal associations with SJS/TEN (OR = 5.75, $p = 1.19 \times 10^{-7}$ for *IKZF1*; OR = 2.80, $p = 2.95 \times 10^{-7}$ for Gene A; OR = 2.71, $p = 2.07 \times 10^{-7}$ for Gene B; OR = 5.34, $p = 2.50 \times 10^{-7}$ for Gene C). One of four loci (*IKZF1*) showed a significant association with SJS/TEN with SOC in the previous GWAS using all the patients and controls followed by a replication study using Korean samples (Ueta *et al.*).

997S

Investigation of genetic variation underlying central obesity among Indian Asians. *W.R. Scott[1,5], W. Zhang[1], M. Loh[1], S-T. Tan[4,5], B. Lehne[1], U. Afzal[1], P. Elliott[1,3], J. Scott[2,5], J. Chambers[1,2,4], J. Kooner[2,4,5]*. 1) Epidemiology and Biostatistics, Imperial College London, Norfolk Place, London, UK; 2) Imperial College Healthcare NHS Trust, DuCane Road, London, UK; 3) MRC-PHE Centre for Environment and Health, Imperial College London, Norfolk Place, London, UK; 4) Ealing Hospital NHS Trust, Southall, Middlesex, UK; 5) National Heart and Lung Institute, Imperial College London, Hammersmith Hospital, Ducane Road, London, UK.

Indian Asians are 1/4 of the world's population and have increased susceptibility to central obesity and related metabolic disturbances. Previous studies investigating the genetic basis of central obesity have been carried out in predominantly European populations. Few studies have investigated the contribution of genetic variation to increased susceptibility to central obesity in Indian Asians.

We studied Indian Asian and European men and women, aged 35-75 years, participating in the London Life Sciences Prospective Population (LOLIPOP) study. Waist hip ratio (WHR) was higher in Indian Asians compared to Europeans (0.94 [95% C.I. 0.86-1.02] v. 0.91 [95% C.I. 0.83-0.99] $P<0.0001$). To investigate genetic variation underlying central obesity among Indian Asians we examined: i) known genetic variants, comprising the 46 SNPs associated with body mass index and WHR in European genome-wide association studies (n=10,318); ii) common genetic variants through genome-wide association (n=10,318); iii) exonic variants through Illumina HumanExome array (n=2,637). Our sample size provides 80% power to detect SNPs explaining >0.3% of trait variance at $P<5\times10\text{-}8$.

We observed no systematic differences in risk allele frequencies or effect sizes of the 46 known adiposity SNPs in Indian Asians compared to Europeans (Sign test P=0.185; P=0.145 respectively). Combined these variants explained <1% of population variation in WHR among Indian Asians. Next we carried out a genome-wide association study of WHR in Indian Asians. Variants from 1000 Genomes haplotypes were imputed into available genome-wide SNP data. >6M SNPs (MAF >5%) passed quality control. No new variants were associated with WHR at $P<5\times10\text{-}8$. We then evaluated exonic variants (247,870 SNPs; 97.4% call rate) for association with WHR in Indian Asians. Again we found no novel associations with WHR ($P<2\times10\text{-}7$; Bonferroni corrected for the number of independent tests). These findings argue against a role for cosmopolitan variants underlying central obesity in Indian Asians. Future efforts aimed at identifying the genetic basis of central obesity in Indian Asians should be focused on other mechanisms, including population-specific variants and transgenerational epigenetic modifications.

998M

Common variants at c11orf30 and CAPN14 are associated with eosinophilic esophagitis. *P.M.A. Sleiman[1,2], M-L. Wang[2,3], A. Cianferoni[2,4], S. Aceves[5], N. Gonsalves[6], K. Nadeau[7], G.T. Furuta[8], J. Spergel[2,4], H. Hakonarson[1,2]*. 1) The Center for Applied Genomics, The Children's Hospital of Philadelphia, PA, USA; 2) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania Philadelphia, PA, USA; 3) Division of GI, Hepatology, and Nutrition, The Children's Hospital of Philadelphia, PA, USA; 4) Division of Allergy and Immunology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 5) Division of Allergy, Immunology, 9500 Gilman Drive MC-0760, Department of Pediatrics and Medicine, University of California, San Diego and Rady Children's Hospital, San Diego, CA, USA; 6) Division of Gastroenterology & Hepatology, Northwestern University - The Feinberg School of Medicine, Chicago, IL, USA; 7) Stanford University School of Medicine, Lucile Packard Children's Hospital, Stanford Hospital and Clinics, Division of Allergy, Immunology, and Rheumatology, CA, USA; 8) Digestive Health Institute, Section of Pediatric Gastroenterology, Hepatology and Nutrition, Children's Hospital Colorado, Gastrointestinal Eosinophilic Diseases Program, Department of Pediatrics, Mucosal Inflammation Program, University of Colorado.

Eosinophilic esophagitis (EoE) is an inflammatory disorder of the esophagus histologically characterized by accumulation of eosinophils in the esophageal epithelium. Clinical symptoms of EoE include dysphagia, failure to thrive, vomiting and epigastric or chest pain. A diagnosis of EoE is made following endoscopy and biopsy upon finding isolated eosinophils in the esophagus having ruled out gastroesophageal reflux. The stringent diagnostic criteria for EoE, that include biopsy proven eosinophilic infiltration of the esophagus, result in a phenotypically homogenous case series that is well powered for GWAS and a potentially powerful model to study the genetics of food allergy and atopy in general. We had previously reported association of the TSLP/WDR36 locus with EoE, here, we report the results of an expanded GWAS totaling 934 cases and 6512 controls in an imputed dataset that included ~2.3M variants. The dataset was split into discovery and replication sets based on the Illumina arrays on which the samples were genotyped (HH550/HH610 or OmniExpress). Following GWAS of the discovery cohort (n= 601 cases and 5837 controls) by logistic regression of the binary EoE phenotype adjusting for sex and the first 10 eigenvectors of the principal component analysis, three loci remained genome wide significant (cutoff $P \leq 5\times10\text{-}8$) following multiple testing correction. The previously reported TSLP/WDR36 locus and two novel loci, c11orf30, which has been previously associated with both atopic and autoimmune disease, and CAPN14, that encodes a calpain whose expression is highly enriched in the esophagus. The same variants at all three loci were also associated with EoE in the replication cohort (n= 333 cases 675 controls). RNAseq of primary epithelial cells derived from esophageal biopsy of 9 EoE patients and 3 controls confirmed expression of TSLP, c11orf30 and CAPN14 in esophageal epithelial cells. Analysis of differential expression between cases and controls showed CAPN14 expression was almost 4 fold increased in EoE cases. c11orf30 and TSLP showed subtle, albeit not statistically significant expression level changes. TSLP and c11orf30 variants have now been associated with a number of atopic conditions and asthma. CAPN14 may be specific to EoE explaining the tissue specificity, it appears to be expressed exclusively in the esophagus and our results also indicate CAPN14 is overexpressed in EoE esophageal epithelial cells compared with controls.

999T

Genome-wide association study imputed to 1000 genomes identifies novel loci associated with lung function. *M. Soler Artigas[1,2], L.V. Wain[1,2], N. Shrine[1,2], J. Huffman[3], I. Sayers[4], D. Strachan[5], I.P. Hall[4], M.D. Tobin[1,2], UK BiLEVE consortium, SpiroMeta consortium.* 1) Departments of Health Sciences and Genetics, Adrian Building, University of Leicester, Leicester, UK; 2) National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester; 3) MRC Human Genetics, MRC IGMM, University of Edinburgh, Edinburgh, Scotland, UK; 4) Division of Respiratory Medicine, University Hospital of Nottingham, Nottingham, UK; 5) Division of Population Health Sciences and Education, St George's, University of London, London, UK.

Lung function measures predict morbidity and mortality, and are used in the diagnosis of chronic obstructive pulmonary disease (COPD), which affected 64 million people worldwide in 2004. Currently genome-wide association studies have identified 26 loci associated with lung function, however they explain only a small proportion of the variance of the lung function measures. In order to identify new loci associated with lung function, in the SpiroMeta consortium we have undertaken the largest meta-analysis of 1000 genomes imputed GWAS of lung function measures to date, analysing around 24 million variants in 38,194 individuals of European ancestry. We studied three quantitative lung function measures: forced expiratory volume in 1 second (FEV_1), forced vital capacity (FVC) and the ratio of FEV_1 to forced vital capacity (FEV_1/FVC). These were adjusted for age, sex, height and ancestry principal components, and stratified by ever smoking status. Genomic control was applied at study level and after meta-analysing the results across studies. After undertaking quality control checks on the top signals, we selected 56 independent loci with P-value<5×10^{-6} for any of the three traits for follow-up; 53 of which were novel and 3 within known regions but still significant when conditioning on the previously discovered sentinel SNPs. The follow-up stage is being undertaken in ~55,000 individuals, including the UK BiLEVE consortium which has genotyped a subset of individuals from UK Biobank. Pending additional follow-up data, our interim findings show that 16 loci meet the genome-wide significance threshold when meta-analysing the discovery and the follow-up stages, and 13 also replicate independently if using a Bonferroni corrected threshold for 56 tests. Currently, the strongest signals are two new signals within known loci for FEV_1/FVC: an intergenic SNP ($P<10^{-25}$) between *GPR126* and *LOC153910* which is in strong LD with a non-synonymous variant and an intronic SNP in *NPNT* ($P<10^{-22}$); as well as a new locus for FVC, an intergenic insertion ($P<10^{-17}$) upstream *PTHLH* and *CCDC91*. This research has been conducted using the UK Biobank Resource.

1000S

Family based association study for pulmonary function in North-east Asian population. *H.Y. Son[1], S.W. Shon[2], S.H. Im[3], H.J. Kim[4], M.K. Lee[6], B. Gombojav[4, 6], H.S. Kwon[7], D.S. Park[8], H.L. Kim[9], K.U. Min[10], J. Sung[6], J.S. Seo[1,4,5], J.I. Kim[1,4,5].* 1) Biochemistry and Molecular Bio, Seoul National University College of Medicine, Seoul, South Korea; 2) Department of Internal Medicine, Dongguk University Ilsan Hospital, Goyang, Republic of Korea; 3) Department of Obstetrics and Gynecology, College of Medicine, Chung-Ang University, Seoul, Republic of Korea; 4) Genomic Medicine Institute, Medical Research Center, Seoul National University, Seoul, Republic of Korea; 5) Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Republic of Korea; 6) Department of Epidemiology and Institute of Environment and Health, School of Public Health, Seoul National University, Seoul, Republic of Korea; 7) Department of Internal Medicine, Asan Medical Center, Seoul, Republic of Korea; 8) Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA; 9) Department of Biochemistry, Ewha Womans University, School of Medicine, Seoul, Republic of Korea; 10) Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea.

The spirometric measurement of pulmonary function forced expiratory volume in one second (FEV1) is a heritable trait that reflects physiological condition of the lung and airway. We conducted a family-based association test with 706 GENDISCAN study participants from 72 Mongolian families to find candidate genetic determinants of pulmonary function. We identified 7 SNPs at 5 different loci showing significant association with FEV1. We chose nine candidate genes from the 5 loci, and tested 381 SNPs on or near the candidate genes for association with FEV1 in 2,729 subjects of Korea Healthy Twin study. Finally, we could identify TMEM132C as a potential candidate gene at 12q24.3, a previously reported locus of asthma and spirometric indices. We also showed significant association with the FEV1 phenotype and two adjacent candidate genes (UNC93A and TTLL2) in the 6q27 region which has been previously identified as a pulmonary function locus in the Framingham cohort study. Our findings suggest novel candidate genes of pulmonary function in North-east Asians.

1001M

Genome-wide association study implicates *LEKR1* at 3q25.31 and an intergenic region at 8q24.21 in Primary Spontaneous Pneumothorax risk. *I. Sousa[1,2], P. Abrantes[1,2], V. Francisco[1,2], G. Teixeira[3], M. Monteiro[4], J. Neves[4], A. Norte[5], C. Robalo Cordeiro[5], J. Moura e Sá[6], P. Santos[1,2], M. Oliveira[7], S. Sousa[8], M. Fradinho[9], F. Malheiro[10], E. Reis[4], L. Negrão[11], S. Feijó[12], S.A. Oliveira[1,2].* 1) Soliveira lab - Genomics of Complex Diseases, Instituto de Medicina Molecular, Faculdade de Medicina da Univ de Lisboa, Lisboa, Portugal; 2) Instituto Gulbenkian de Ciência, Oeiras, Portugal; 3) Hospital Infante D. Pedro, Aveiro, Portugal; 4) Hospital de Santo António (Centro Hospitalar do Porto), Porto, Portugal; 5) Centro de Pneumologia, Faculdade de Medicina da Universidade de Coimbra, Coimbra, Portugal; 6) Centro Hospitalar de Vila Nova de Gaia, Vila Nova de Gaia, Portugal; 7) Universidade de Évora, Évora, Portugal; 8) Hospital de S. Bernardo (Centro Hospitalar de Setúbal, E.P.E.), Setúbal, Portugal; 9) Hospital Egas Moniz (Centro Hospitalar de Lisboa Ocidental), Lisboa, Portugal; 10) Hospital da Luz, Lisboa, Portugal; 11) Instituto Português do Sangue e da Transplantação, Centro Regional de Sangue de Lisboa, Lisboa, Portugal; 12) Hospital de Santa Maria, Lisboa, Portugal.

Primary spontaneous pneumothorax (PSP) is characterised by the presence of air in the pleural cavity that occurs without preceding trauma or known cause in individuals with no lung disease. Despite elevated incidence and recurrence rates, little is known about its aetiology. So far, the genetics of PSP remains largely unresolved and virtually no research has been dedicated to the identification of genetic factors for risk or recurrence of sporadic PSP.

To identify genetic variants contributing to sporadic PSP risk, we conducted the first PSP genome-wide association study (GWAS). Two replicate pools of 92 Portuguese PSP cases and of 129 age- and sex-matched controls were allelotyped in triplicate on the Affymetrix Human SNP Array 6.0 arrays. Tested markers were ranked by relative allele score difference between cases and controls ($|RAS_{diff}|$), through a novel cluster method and a combined Z-test. 101 out of the 108 most significant SNPs were technically validated by individual genotyping ($P < 5.00 \times10^{-2}$). Replication of validated SNPs was carried out in an independent Portuguese dataset of 100 cases and 425 controls.

Two single nucleotide polymorphisms (SNPs), rs2101167 in *LEKR1* (leucine, glutamate and lysine rich 1) at 3q25.31 and the intergenic variant rs4733649 at 8q24.21 were strongly associated ($P < 5.00 \times10^{-4}$) with sporadic PSP in the combined datasets. The study reported herein provides for the first time genetic clues for sporadic PSP risk, but future studies are warranted to uncover the function of these polymorphisms.

1002T

Genomewide association for rotator cuff disease identifies two significant SNPs. *C.C. Teerlink[1], J.M. Farnham[1], L.A. Cannon-Albright[1,2,3], R.Z. Tashjian[4].* 1) Dept Internal Medicine, Univ Utah Sch Med, Salt Lake City, UT; 2) George E. Wahlen Dept of Veterans Affairs Medical Center, Salt Lake City, UT; 3) Huntsman Cancer Institute, Salt Lake City, UT; 4) Dept of Orthopaedics, University of Utah School of Medicine, Salt Lake City, UT.

Previous evidence suggests a familial predisposition for the development of rotator cuff tearing. There is limited data identifying specific genes associated with rotator cuff tearing. We collected 238 blood samples for DNA extraction from patients with MRI-confirmed full-thickness rotator cuff tears for the purpose of conducting a genome-wide association analysis. Samples were genotyped on the illumina Omni5.0 SNP platform. We used 2756 genetically matched Caucasian controls from the illumina icontrols database genotyped on the illumina 550K platform. The intersection of the two platforms contains 272,704 markers across the genome that passed typical quality control metrics. We used GEMMA software to conduct tests of association at each marker. GEMMA software empirically estimates a kinship covariance matrix for all pairs of individuals, which is used in the analysis to mitigate potential effects of population stratification and cryptic relatedness. Our analysis identified two SNPs that exceeded the threshold for significance ($p < 1.8\text{-}7$) after adjusting for multiple testing. One SNP (rs11939479; p = 6.9-10; OR = 2.8) resides on 4q32.3 in the SPOCK3 gene. SPOCK3 codes for a proteoglycan which are typically expressed in connective tissue. The other significant SNP (rs7897503; p =1.2-7; OR = 2.0) resides on 10q21.1, approximately 200 Kb downstream of PCHD15, which is a member of the cadherin family of genes that are generally involved in cell-cell adhesion. To our knowledge, these findings represent the first attempt to identify genetic factors influencing rotator cuff disease via a genome-wide association. Identification of potential genes or genetic variants associated with rotator cuff tearing may help in identifying individuals at-risk for the development of rotator cuff tearing as well as identifying possible genetic targets to for biologic augments to improve rotator cuff repair healing.

1003S

A genome-wide meta-analysis of hyper- and hypothyroidism and thyroid function. *A. Teumer[1], M. Medici[2], H. Völzke[1], R. Peeters[2], the CHARGE thyroid function working group.* 1) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; 2) Department of Internal Medicine, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands.

Thyroid hormones play a key role in cellular growth, development and metabolism. Overt thyroid dysfunction leads to common endocrine disorders affecting approximately 10% of individuals over their life span. Milder changes in thyroid function are associated with weight changes, coronary heart disease, atrial fibrillation, osteoporosis, psychiatric disorders, and mortality. Circulating concentrations of thyroid hormones have a strong heritable component. Using genome-wide association studies (GWAS) in predominantly population-based cohorts, we and others have identified 26 genetic loci associated with serum TSH and free T4 levels. However, the variance explained by these loci is still low, i.e. 5.64% and 2.30% for TSH and free T4, respectively. To find novel loci associated with thyroid function, we have increased our sample size substantially to conduct a new GWAS in more than 33,000 individuals. In contrast to former studies, these analyses will be performed by imputing against a recent 1000 genomes panel, which gives us the opportunity to fine-map known loci and reveal additional loci due to better imputation and tagging of causal SNPs. Sex-stratified analyses of TSH and free T4 will be performed in individuals with TSH values within the reference range, allowing the analysis of interactions by sex. Additionally, we are performing a GWAS for hypo- and hyperthyroidism which will address genetic influences for more overt forms of thyroid disorders. Preliminary results in up to 1800 cases showed that 11 of the 26 SNPs known to be associated with serum TSH and free T4 in euthyroid subjects, including PDE8B and PDE10A, are also associated with hypothyroidism ($p < 4E-4$) or hyperthyroidism ($p < 4E-5$). In this 1000 genomes based GWAS, we expect to reveal additional loci influencing thyroid function as well as hypo- and hyperthyroidism.

1004M

Genome-wide Mega-Analysis on Myopia and Refractive Error in CREAM and 23andMe. *V.J.M. Verhoeven[1,2,3,4], R. Wojciechowski[5,6], P.G. Hysi[7], Q. Fan[8], A.K. Kiefer[9], N. Eriksson[9], C.J. Hammond[7], N.A. Furlotte[9], C.C.W. Klaver[1,2], Consortium for Refractive Error and Myopia (CREAM).* 1) Deparment of Ophthalmology, Erasmus Medical Center, Rotterdam, the Netherlands; 2) Deparment of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; 3) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore; 4) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, Netherlands; 5) 5.Department of Twin Research and Genetic Epidemiology, King's College London School of Medicine, London, United Kingdom; 6) 6.Inherited Disease Research Branch, National Human Genome Research Institute, US National Institutes of Health, Baltimore, Maryland, United States; 7) 7.Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States; 8) 8.Saw Swee Hock School of Public Health, National University Health Systems, National University of Singapore, Singapore, Singapore; 9) 23andMe, Mountain View, California, United States of America.

Purpose: Myopia is widely recognized as a multifactorial, complex genetic disorder. Recently, multiple loci for refractive phenotypes were identified separately by the Consortium for Refractive Error and Myopia (CREAM) and investigators from 23andMe, Inc. We aimed to identify additional genetic loci that explain the genetic architecture of refractive error using higher power and denser imputation by meta-analyzing data from CREAM and 23andMe. Methods: We conducted a genome-wide association study (GWAS) meta-analysis of refractive error including 63,697 individuals (49,808 Caucasians; 13,899 Asians) from the CREAM consortium using a linear regression model. Age-at-onset of myopia for 104,294 individuals from the 23andMe dataset was analyzed using survival analysis (Cox proportional hazards model). GWAS regression results from both studies were meta-analyzed by z-scores using a fixed effects model. We performed pathway analyses using Ingenuity IPA and public databases. Results: Over 100 regions across the genome reached genome-wide significance at P-value $< 5.0\times10^{-8}$. The most significant P-value was 2.74×10^{-54} for rs670352 near the known refractive error gene *GJD2*. We confirmed association with all but 2 from 36 previously reported CREAM and 23andMe genes. We identified 71 new genome-wide significant associations with refractive error and myopia. These SNPs explain ~12% of the variance of all common SNPs for SE. The results confirm over-representation of known pathways, such as extracellular matrix, ion channel activity, and glutamate signaling but also suggested potentially new mechanisms, including signaling of calcium, VEGF and TGF-β, mitochondrial function, and apoptosis. Conclusions: This study is the largest meta-analysis on refractive error known to date. This large catalogue of genetic variants opens up new insights in myopigenesis.

1005T

Genetic discovery in the 23andMe participant cohort. *D.A. Hinds, C.A.M. Northover, M.H. McIntyre, C. Wilson, K.E. Huber, A. Kleinman, F. Sathirapongsasuti, R.K. Bell, E. Pierson, K. Bryc, A.S. Shmygelska, N.A. Furlotte, Y. Hu, C. Tian, E.Y. Durand, C.Y. McLean, B. Naughton, J.L. Mountain, N. Eriksson, J.Y. Tung.* 23andMe, Inc., Mountain View, CA.

The 23andMe cohort now includes more than 550,000 genotyped subjects who have consented to participate in research, including more than 100,000 individuals with substantial non-European ancestry. Most participants provide phenotypic information through web based surveys covering a wide range of topics. The scale of this cohort has required us to develop specialized solutions for many tasks where conventional approaches have inconvenient scaling properties. These include methods for ancestry inference and detection of identity by descent; phasing and imputation; and genome-wide association testing. While most of our data is collected on the "sample of convenience" of 23andMe customers, we have recruited cohorts in several disease areas, including Parkinson's disease, myeloproliferative neoplasms, and inflammatory bowel disease. Over the next year, we expect to take advantage of much deeper imputation panels to extend the range of allele frequencies we can effectively interrogate.

Our status as a commercial company poses challenges with respect to our participation in the scientific community and our responsibilities to protect the privacy of our customers, many of whom did not join 23andMe with the intent to participate in research. We have tried to balance these demands through publication of GWAS findings and participation in collaborations to enable discoveries using the data our customers have shared, while controlling access to potentially identifiable aggregated data, and in rare cases, only providing access to individual-level data with explicit customer consent. We have over 25 active academic collaborations in atopic disease, infectious disease, autoimmune disease, fertility, migraine, Parkinson's disease, and morphological phenotypes, among others.

Cohort growth has expanded the range of phenotypes for which we are able to make discoveries, to where we can now make meaningful scientific contributions even for phenotypes that have been intensively investigated by large consortia. The dataset is now also sufficiently large to identify robust associations with categories of phenotypes that have until now been essentially intractable to genetic analysis, including cognition and behavior. The data provide unique insights into the genetic architecture of human phenotypic variation.

1006S

A genome-wide regulatory haplotype analysis of asthma. *D.C. Croteau-Chonka[1], W. Qiu[1], S.T. Weiss[1,2], B.A. Raby[1,3], EVE Consortium.* 1) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Partners Center for Personalized Medicine, Partners Health Care, Boston, MA; 3) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Genome-wide association studies (GWAS), including two large meta-analyses of >35,000 individuals (the GABRIEL and EVE asthma consortia), have identified more than ten highly-reproducible asthma susceptibility loci. We can now start to look beyond single-marker main effects on asthma to epistasis and the role of functional regulatory haplotypes, namely the joint effect of pairs of regulatory and deleterious coding single nucleotide polymorphisms (SNPs) on disease susceptibility. Such associations may give further insight into potential underlying mechanisms at asthma GWAS loci.

We have performed a pilot genome-wide haplotype association analysis of asthma susceptibility in 359 cases from the Childhood Asthma Management Program (CAMP) along with 846 healthy Illumina iControlDB controls. We constrained our analyses to a set of 526 genes with significant evidence within 50 kb windows of expression quantitative trait loci (eQTL) (false discovery rate (FDR) < 5%) in disease-relevant CD4+ T-lymphocytes (Murphy *et al.* (2010), *Hum Mol Genet.*). The dataset thus consisted of 1,609 eQTL SNPs matched with 831 coding SNPs in their respective target genes. We performed an H - 1 degrees-of-freedom case-control omnibus haplotype test of each of the SNP pairs (where $H = 3$ or 4 observed haplotypes).

Among 2,538 omnibus tests, we identified 149 associations significant at nominal $P < 0.05$, but no associations significant at FDR < 5%. A visual inspection of the distribution of omnibus associations suggested a moderate departure from the expected distribution, though there was no substantial evidence of inflation ($\lambda_{GC} = 1.05$). One of the most significant omnibus tests ($P = 0.0087$) included a quinone oxidoreductase gene called *CRYZ*, which is near a GWAS SNP associated with circulating levels of resistin. A pro-inflammatory hormone, resistin is thought to be a potential biomarker for asthma.

The next stage of the analysis will be to include all 4,903 cases, 3,371 controls, and 1,677 pseudo-controls among 12 cohorts from the entire EVE Consortium. We will use meta-analysis to integrate all of the individual cohort haplotype results together using Stouffer's method, setting the cohort weights to be their effective sample sizes. The marked increase in sample size will substantially improve our ability to detect regulatory haplotypes of asthma.

1007M

Genetic analysis of central serous chorioretinopathy and polypoidal choroidal vasculopathy. *J. Ahn[1,2], N.K. Ryoo[1,3], S.J. Woo[1,3], H.S. Cheong[4], K.H. Park[1,3].* 1) Department of Ophthalmology, Seoul National University, College of Medicine, Seoul, Korea; 2) Department of Ophthalmology, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, Korea; 3) Department of Ophthalmology, Seoul National University Bundang Hospital, Seongnam, Korea; 4) Department of Genetic Epidemiology, SNP Genetics Incorporation, Seoul, Korea.

Central serous chorioretinopathy (CSC) and polypoidal choroidal vasculopathy (PCV) are both common retinal disorders affecting the macula. Although PCV is often considered a subtype of exudative age-related macular degeneration (AMD), CSC and PCV share many clinical characteristics such as choroidal vascular hyperpermeability, choroidal thickening demonstrated on enhanced-depth imaging optical coherence tomography and inner choroidal vasculature abnormalities seen on indocyanine angiography. In this study, we performed genetic analysis of CSC and PCV using genome-wide association study (GWAS) and exome chip data to compare the genetic makeup of the two disease entities. This was a case-control analysis consisting of 168 PCV, 167 CSC patients and 543 control subjects. A subgroup analysis of 50 CSC patients diagnosed as the "chronic" phenotype, with persistent or progressive visual symptoms for more than 6 months, was performed additionally. PCV and control subjects were genotyped using Illumina OmniExpress and CSC patients were genotyped with the HumanExome beadchips. Integrated analysis of GWAS and exome array data was done and a total of 12,777 SNPs were included. The well-known ARMS2 rs10490924 SNP showed significant association with PCV compared to the control (P=1.78E-13). A novel SNP rs2455512 was found to be associated with all and chronic CSC patients compared to the control group (P=2.13E-11 and 6.92E-11, respectively). There were 3 SNPs (rs8098316, rs1363098, rs1214752) that showed borderline significant association with all CSC patients compared to control (P=1.85E-05, 4.90E-05, and 8.74E-05, respectively). In the analysis of chronic CSC versus the control group, 4 SNPs (rs12843815, rs6843141, rs11158685, rs4800452) reached borderline significance (P=1.46E-05. 7.18E-05. 7.19E-05, and 8.89E-05). This is the first study to perform comparative genetic analysis of CSC and PCV and we found differences in the significant SNPs for each disease entity compared to the control group. The novel SNPs that were found to be associated with CSC require further replication studies.

1008T

Known Age-Related Macular Degeneration Risk Variants Are Not Associated with Rapid Disease Progression or Good Treatment Response. *M.D. Courtenay[1], W. Cade[1], G. Wang[1], S.G. Schwartz[2], J.L. Kovach[2], A. Agarwal[3], M.A. Brantley[3], W.K. Scott[1], J.L. Haines[4], M.A. Pericak-Vance[1].* 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 3) Ophthalmology and Visual Sciences, Vanderbilt University School of Medicine, Nashville, TN; 4) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

Age-related macular degeneration (AMD) is the leading cause of irreversible vision loss among older adults in developed countries. Genome-wide association meta-analysis has implicated common variations in 19 genes as AMD risk factors. Given the strong genetic influence on the risk of developing AMD, and the substantial variation in progression to advanced AMD and in treatment response, we investigated whether these clinical endpoints are also influenced by the known genetic variation. In an existing dataset, 16 of the 19 GWAS loci were used to calculate a genetic risk score (summed number of risk alleles weighted by effect sizes) for each individual as outlined in Fritsche et al., 2013. To maximize our power for discovery, we created two "clinically extreme" case/control datasets with the following characteristics: 1) Rapid clinical progression of 36 individuals with early AMD or geographic atrophy to neovascular AMD within one year vs. 31 early AMD patients who did not progress over five or more years, and 2) 20 patients with bilateral neovascular AMD who were treated with intravitreal injections of anti-angiogenic agents for one year and demonstrated bilateral good treatment response (three lines of vision gain on the Snellen vision chart, prolonged or persistent (at least 6 months) absence of subretinal fluid (SRF) and absence of intraretinal fluid (IRF) on OCT, presence of unchanged small intraretinal cysts with no recurrence of SRF) vs. 22 neovascular AMD patients who demonstrated a poor response in one or both eyes (three lines of vision loss, persistent SRF and IRF, formation of disciform scar). The genetic risk score was tested for association with rapid AMD progression or good treatment response with logistic regression adjusting for age and sex. Risk score was not associated in either dataset in our study (progression P=0.43 and treatment P=0.63). Therefore, the genetic effects of the 19 known AMD risk loci are not associated with rapid progression or good treatment response in this small dataset, but perhaps a larger sample size would have sufficient power to demonstrate an effect. Nevertheless, unknown genetic factors may still be regulating variability in progression and treatment response. Whole exome sequencing is currently underway in this dataset to test if novel rare or common variants are associated with these clinical endpoints.

1009S

Large association study of exonic variants in idiopathic achalasia. *J. Becker[1,2], T. Schwämmle[1,2], M.M. Wouters[3], S. Niebisch[4], E. Mangold[1], H.G. Schulz[5], G.E. Boeckxstaens[3], M.M. Nöthen[1,2], I. Gockel[4], M. Knapp[6], J. Schumacher[1].* 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany; 3) Translational Research Center for Gastrointestinal Disorders, Catholic University of Leuven, Leuven, Belgium; 4) Department of General, Visceral and Transplant Surgery, University Medical Center, University of Mainz, Mainz, Germany; 5) Department of General and Abdominal Surgery, Protestant Hospital Castrop-Rauxel, Castrop-Rauxel, Germany; 6) Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany.

Idiopathic achalasia is a severe disorder of the lower esophageal sphincter (LES) with a lifetime prevalence of 1:10,000. The disease is characterized by the degeneration of neurons in the myenteric plexus leading to the development of a megaesophagus with irreversible loss of LES function. On the etiological level, achalasia is a multifactorial disorder with environmental and genetic factors being risk-associated. By testing markers in immune-relevant loci using the Illumina's Immunochip and subsequent HLA imputation, we already identified strong association signals reaching genome-wide significance within the HLA-DQ complex indicating that autoimmune processes contribute to the pathophysiology of achalasia. The aim of the present study was to determine the role of exonic variants in the development of achalasia. We therefore performed an association study using Illumina's Exomechips which have been developed based on the data of 12,000 exomes. The chip contains more than 240,000 - mainly functional-relevant - markers. We genotyped 677 patients with idiopathic achalasia and 2,316 population-based controls from Central Europe and after quality control (QC) steps 106,417 markers remained for association testing. The analysis yielded a strong association signal within the HLA region ($P<5\times10^{-08}$). We carried out a conditional analysis adjusting for the variants within the HLA-DQ complex identified before. The analysis revealed that the HLA signal on the Exomechip is not independent of the already known achalasia risk variants within this region. Next, we focused on variants outside the HLA region and chose 57 markers ($P<10^{-03}$) for a successive replication step in 383 patients and 1,017 population-based controls of Central European descent. Of 51 variants surpassing QC steps, 5 markers reached a *P*-value below 0.1 in the replication step with the same allelic direction as seen in the initial Exomechip study. These 5 markers are either low-frequent (N=4, $MAF_{Controls}<5\%$) or common (N=1) variants and are located within the genes *CRISP3*, *HBS1L*, *IGSF22*, *KRTAP12-1* and *LMOD3*. Our study provides evidence that low-frequent and common exonic variants play a role in the pathophysiology of achalasia. Currently, we test the 5 associated markers of the replication step in a third cohort of achalasia patients and controls in order to confirm the contribution of these variants to the development of achalasia.

1010M

Identification of novel and rare coding variants associated with free fatty acids and serum fatty acid profile. *X. Sim[1], R.P. Welch[1], A.U. Jackson[1], H.M. Stringham[1], T.M. Teslovich[1], P.S. Chines[2], N. Narisu[2], C. Fuchsberger[1,3], J.R. Huyghe[1], A.E. Locke[1], M. Cannon[4], M. Ala-Korpela[5], M. Boehnke[1], K.L. Mohlke[4], M. Laakso[6].* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 2) Genome Technology Branch, National Human Genome Research Institute, Bethesda, Maryland, USA; 3) Center for Biomedicine, European Academy of Bozen/Bolzano (EURAC), Bolzano, Italy; 4) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA; 5) Computational Medicine Research Group, Institute of Health Sciences, University of Oulu, Finland; 6) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

Fatty acid (FA) levels play a critical role in insulin sensitivity, and are often elevated in obese or individuals with type 2 diabetes. Previous GWAS studies have implicated common genetic variants with FAs, however, genetic contributions at coding and low frequency variants remain largely unknown. We assayed exomic variation in 6,373 non-diabetic Finnish males from population-based METSIM study with Illumina HumanExome array (N=~242,000) and Illumina OmniExpress genotypes imputed to GoT2D reference panel (N=~19.3M). Free FA (FFA) was measured on enzymatic colorimetric method while FA profile (including n-3, n-6, n-7, n-9, mono/poly-unsaturated, phosphatidylcholines, and sphingomyelins) were measured using proton nuclear magnetic resonance spectroscopy. Inverse normalised residuals were generated after adjusting for age, BMI, and smoking as a primary model. We identified genome-wide associations at 19 unique loci with multiple FAs. An intergenic variant (MAF=17%, $P=2\times10^{-9}$) was associated with FFA, near *PPP1R3B*, a glycogen-targeting regulatory subunit regulating liver glycogen metabolism. The remaining 18 loci included previously identified FA (e.g. *FADS1-2-3*, *CPT1A*), lipids and glycemic loci. As FAs are correlated with lipids and glycemic traits, we repeated association analyses on an adjusted model, conditioning these traits from FA levels. We identified 2 additional loci with the adjusted model only: (i) common intronic variant rs4985154 at *PDXDC1* associated with polyunsaturated FAs other than linoleic acid [MAF=34.4%, $P=5\times10^{-8}$], and (ii) rare missense variant rs199717050 at *LCAT* associated with both phosphatidylcholine and other cholines (PC) and total phosphoglycerides (TotPG) [MAF=.46%, $P_{PC}=7\times10^{-10}$, $P_{ToTPG}=3\times10^{-8}$]. We also note that associations at 9 loci were substantially diminished after conditioning on lipids and glycemic traits, suggesting that they were mediated by the major classes of liproproteins and insulin measures. Finally, we aggregated rare (<1%) protein-altering variants and carried out gene-level association analysis using SKAT-O. We identified gene-level associations at *LIPG* with PC [$P=2\times10^{-7}$] and TotPG [$P=5\times10^{-8}$], combining evidence across 3 missense variants and a splice acceptor on the primary model. We demonstrate that rare coding genetic variants influence FAs, and could be informative about biological pathways in lipid metabolism and insulin sensitivity.

1011T

Replication of the association signals of thyrotoxic periodic paralysis (TPP) at chromosome 17q24.3. *P. Chen[1,2,3,4,5], T. Chang[2,6].* 1) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan, Taiwan; 2) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 3) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 4) Graduate Institute of Clinical Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 5) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan; 6) Department of Internal Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan.

Thyrotoxic periodic paralysis (TPP) is a characteristic disease featured by episodic attacks of hypokalemia and muscle weakness in certain thyrotoxic individuals. Epidemiology data of TPP show male predominance and huge inter-population prevalence difference. Up to 10% of thyrotoxic Asian males develop TPP, while only less than 0.1% thyrotoxic Caucasian males do. In 2010, Ryan et al. (Cell 140(1): 88-98) found that *KCNJ18* mutations could be found in 33% of their patients but not in controls; those mutations were mainly found in patients of European descent. In 2012, Cheung et al. (Nat Genet 44(9): 1026-1029) published a GWAS (genome-wide association study) in ethnic southern Chinese and reported a single susceptibility locus at chromosome 17q24.3 near *KCNJ2*; the odds ratio was estimated to be ~3.3. It is not clear whether these two loci are genuine risk loci for TPP. Furthermore, if these two loci do confer susceptibility, it is not clear about the real causative genes and the risk alleles. To address these issues, we conducted a GWAS with 41 TPP patients as cases and 725 thyrotoxic, non-TPP individuals as the control individuals. All of the participants are of ethnic Chinese Han population background in Taiwan. We used Axiom Genome-Wide CHB Array Plate from Affymetrix as the genotyping platform, which contains 642,832 common SNPs (MAF>5%) in the Chinese Han genome. Our results showed several SNPs with p values smaller than 5×10^{-7} (trend test and/or allelic test) at chromosome 17q24.3. Association signals at 2q31.2, 2q36.3, 9q22.1 and 18q21 were also identified. We further perform Sanger sequencing of the coding regions of the *KCNJ2* gene; however, no disease causing variants could be found. On the other hand, our GWAS did not show association signals at or near *KCNJ18*. Sanger sequencing of the coding regions of the *KCNJ18* also showed negative results. Our results demonstrate that 17q24.3 is the true TPP susceptibility locus in ethnic Chinese Han individuals. However, the genuine causative variants are still unidentified yet.

1012S

The PhenX Toolkit: A Genomic Resource for Standard Measures of Phenotypes and Exposures. *W. Huggins[1], H. Pan[1], D.S. Nettles[1], E.B. Bullard[1], T. Hendershot[1], J.G. Pratt[1], D.R. Maiese[1], D.C. Brown[1], M. Phillips[1], T.H. Brandon[2], S.T. Tiffany[3], G.K. Brown[4], P.H. Byers[5], M.L. Marazita[6], C.A. McCarty[7], K.P. Conway[8], G.K. Farber[9], K.L. Wanke[10], E.M. Ramos[11], C.M. Hamilton[1].* 1) RTI International, Research Triangle Park, NC; 2) Moffitt Cancer Center, Tampa, FL; 3) University at Buffalo, State University of New York, Buffalo, NY; 4) University of Pennsylvania, Philadelphia, PA; 5) University of Washington School of Medicine, Seattle WA; 6) University of Pittsburgh, Pittsburgh, PA; 7) Essentia Institute of Rural Health, Duluth, MN; 8) National Institute on Drug Abuse, Bethesda, MD; 9) National Institute of Mental Health, Bethesda, MD; 10) Office of Disease Prevention, National Institutes of Health, Bethesda, MD; 11) National Human Genome Research Institute, Bethesda, MD.

The PhenX (consensus measures for **Phen**otypes and e**X**posures) Toolkit (https://www.phenxtoolkit.org/) is a well-established, Web-based catalog of measures for collaborative biomedical research. PhenX Working Groups (WGs) consider measures that are suitable for a variety of study designs, including longitudinal and clinical studies as well as genomic studies. Expert Review Panels are being convened to ensure that the current Toolkit measures are responsive to the evolving needs of the biomedical community. PhenX phase 1 focused on measures of phenotypes and exposures relevant to genomic studies of common complex diseases and released measures for 21 research domains (including Demographics, Nutrition, Diabetes, and Environmental Exposures) in the PhenX Toolkit. PhenX phase 1 also added depth to the Toolkit in the area of Substance Abuse and Addiction by adding one Core collection and six Specialty collections. PhenX phase 2 expands the scope of the Toolkit to address four new domains and also adds depth in the areas of Mental Health Research and Tobacco Regulatory Research. The first new domain to be addressed is Rare Conditions; preliminary measures and Toolkit annotations will be presented. “PhenX Measures for Mental Health Research,” funded by the National Institute of Mental Health, aims to add several Specialty collections (Suicide, PTSD, and Eating Disorders) and one Core collection to the Toolkit. “PhenX Measures for Tobacco Regulatory Research,” funded by the Tobacco Regulatory Science Program, aims to add several Specialty collections (Social/Cognitive, Bio-behavioral, Agent, Vector, and Environment) and one Core collection to the Toolkit. Preliminary measures for the Suicide and Social/Cognitive WGs will be presented. To support investigators who want to collect data via the Web, Web-based versions of PhenX protocols will be made available as Research Electronic Data Capture (REDCap) modules. Study variables in the database of Genotypes and Phenotypes (dbGaP) are being classified as identical, comparable, or related to PhenX variables. The dbGaP advanced search tool uses these mappings to help investigators discover variables and studies that are suitable for cross-study analysis. The PhenX Toolkit provides the biomedical research community with easy access to recommended, low-burden measures that have the potential to increase the overall impact of individual studies. *Funding provided by U01 HG004597, 3U01 HG004597-02S1, and U41 HG007050.*.

1013M

Analysis of Interleukin 10 haplotypes with soluble IL-10 levels and autoantibody production in Mexican patients with primary Sjögren’s syndrome. *M. Vázquez-Villamar[1], CA. Palafox-Sánchez[1,2], JF. Muñoz-Valle[1], Y. Valle[1], D. Salazar-Camarena[1], BA. Treviño-Talavera[1], G. Orozco-Barocio[2], E. Oregon-Romero[1].* 1) Instituto de Investigación en Ciencias Biomédicas (IICB), Universidad de Guadalajara, Guadalajara, México; 2) Servicio de Reumatología, Hospital General de Occidente, Zapopan, México.

Background: Primary Sjögren's syndrome (pSS) is a systemic autoimmune disease characterized by lymphocytic infiltration of salivary and lachrymal glands, leading to xerostomia (dry mouth) and xeroftalmia (eye dryness). Infiltrating monocytes produce interleukin 10 (IL-10) that is capable of promoting the activation and differentiation of B cell as well as autoantibodies production, such as anti-Ro and anti-La which have an important role in the pathogenesis of pSS. Three polymorphic sites in positions -1082(A/G), -819(C/T), and -592(C/A) in the IL10 promoter can form haplotypes that may influence the IL-10 production. Objective: To determine the frequency of IL10 haplotypes and serum levels of IL-10 (sIL10), anti-Ro, and anti-La antibodies in patients with SSp. Methods: The study included 111 patients with pSS, from the Department of Rheumatology (HGO, SSJ, Zapopan, Jalisco, Mexico). As a control group, 260 subjects were included. IL10 polymorphisms were genotyped by PCR/RFLP technique. The haplotype frequencies were inferred by EmapFre software. Levels of sIL-10, anti-Ro, and anti-La autoantibodies were performed by ELISA. Rheumatoid factor (RF) levels were determined by nephelometry. Statistical analysis was performed with Stata 9.0 and GraphPad Prims 6 software. Results: The polymorphisms were in Hardy-Weinberg equilibrium and showed a high linkage disequilibrium (100%, pc=3673^{-18}). The most frequent haplotypes were: ACC (44%), ATA (23%) and GTA (16%). The ATC haplotype was associated with increased risk for pSS (OR 3.42, CI:1.06-11.04, p=0.0395). Higher sIL10 levels were observed in patients with pSS than in the control group (3.0 vs 2.4 pg/mL, p<0.0001). sIL-10 correlated with anti-Ro(r=0.3615, p= 0.0053) and anti-La(r=0.2, p=0.0401) autoantibodies. Patients positive to anti-Ro showed higher levels of sIL-10 than negative patients (3.2 vs 2.9 pg/mL, p=0.0243). Similar results were found to pSS positive to anti-La antibodies (3.4 vs 2.9 pg/mL, p=0.0027). Conclusions: Our results suggest that carriers of the haplotype ATC have 3.42-folds more likely to develop SSp, however, due to their low frequency in our population lacks clinical utility. Serum IL-10 levels are increased in pSS patients and are positively correlated with autoantibodies which suggest its role as a pathogenic cytokine in pSS. However, the low associations of IL-10 with the clinical parameters suggest that it should not be considered as a biomarker of the disease.

1014T

Large scale meta-analysis of 1000G imputed genotypes in 95,061 subjects reveals 7 novel loci for loss of kidney function. *M. Gorski[1,2], the CKDGen Consortium.* 1) Department of Genetic Epidemiology, University of Regensburg, Regensburg, Bayern, Germany; 2) Department of Nephrology, University Hospital Regensburg, Regensburg, Germany.

Meta-analyses of genome-wide association studies (GWAS) using Hapmap imputed genotypes have successfully revealed several susceptibility loci for estimated glomerular filtration rate (eGFR). 1000G imputed genotypes promise to reveal novel loci and to give additional insights into the genetic architecture of kidney function measured by eGFR. A large scale meta-analysis of 1000G imputed genotypes has not yet been performed. The objective was to answer if a meta-analysis of 1000G imputed GWAS can identify loci for the quantitative measure eGFR, and to evaluate how potentially new loci compare to previous meta-analysis based on Hapmap imputed genotypes. GFR was estimated by the four-variable MDRD Study Equation. GWAS were performed on genotype dosages, imputed with reference panels of up to 30 million variants in 32 studies including 95,061 subjects. We identified 7 new susceptibility loci for reduced eGFR (p values range from 2.89E-08 to 6.21E-11). The loci were in or near the genes LPHN2, ACVR2A, KIAA1715, ARL15, SLC17A1, ASTN2 and SLC7A6. The lead variants of the novel loci were either new in the 1000G reference panel or were previously included in Hapmap panels but did not show genome wide significance. One of the loci, SLC7A6, was suggested in previous work based on pathway derived candidates. We also replicated 31 of 53 lead SNPs in known loci with genome wide significance (p-values < 5E-8). Our large-scale meta-analysis of 1000G imputed genotypes in 95,061 subjects identified novel susceptibility loci for the kidney function measure eGFR, mostly due to the higher coverage of variants in the latest reference panels. Lead SNPs in 2 of the novel loci (SLC17A1 and KIAA1715) are non-common (minor allele frequency 2.8% and 3.6%).

1015S

Association analysis of *PPARG* (p.Pro12Ala) polymorphism with type 2 diabetic retinopathy in patients from north India. *N. Kaur, V. Vanita.* Department of Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

Purpose: To determine whether p.Pro12Ala polymorphism in *PPARG* is associated with type 2 diabetic retinopathy in patients from north India. **Material & Methods:** Present case-control association study included total 1613 individuals, of which 717 were diagnosed as diabetic retinopathy (DR) patients due to presence of neovascularisation, microaneurysms and hemorrhage in their retina. 608 T2DM patients without any sign of retinopathy (CDR) and 288 healthy controls (NDC) without diabetes and any sign of retinopathy, were also collected and analyzed. The mean duration of diabetes in DR and CDR group was 12.440±6.356 and 8.050±5.772, respectively. The genotyping for p.Pro12Ala polymorphism of *PPARG* in these total 1613 cases and controls was performed by Taqman Drug Metabolism Genotyping Assays using Real time PCR. Further, few genotypes were confirmed by direct DNA sequence analysis of the amplified products. Statistical analysis was performed using SPSS for windows version 16 (SPSS Inc). Binary logistic regression analyses were used to determine odds ratio (OR). Hardy-Weinberg equilibrium was tested using Cochran-Armitage trend test (2 x 3 contingency table) based on linear regression model. **Results:** The genotype distribution and allele frequency of p.Pro12Ala (*PPARG*) polymorphism did not differ significantly (p>0.05) between DR and CDR. However, statistically significant differences in genotype (p=0.002) and allele distribution (p= 0.0326) were observed on comparing DR with NDC group. A significantly higher frequency of heterozygous CG genotype was observed in DR group as compared to NDC group (19.66% vs 10.76%). On comparison of CDR and NDC groups statistically significant differences in the genotype distribution were observed (p=0.0005) whereas distribution of allele frequency did not differ significantly (p=0.091) among these two groups. The homozygosity of dominant allele CC was found to be more common in NDCs as compared to CDRs (86.11% vs 78.94%) which suggested the evidence of the association of this genotype in dominant mode of inheritance (OR; 1.65, 95% CI: 1.72-2.43, p=0.009). **Conclusion:** These findings suggest a statistically significant association of *PPARG* (p.Pro12Ala) polymorphism with T2DM when compared with healthy controls. However, no significant association has been observed on comparing DR cases with T2DM patients in tested north Indian patients.

1016M

To study the association of -106 C/T polymorphism in the aldose reductse (*AKR1B1*) gene with diabetic retinopathy. *Vanita. Kumar, N. Kaur.* Department of Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

Purpose: Present study aimed to assess the association of -106C/T polymorphism of the aldose reductase gene (*AKR1B1*) with retinopathy in type 2 diabetic cases from North India. **Material and Methods:** The present study included 213 patients diagnosed with diabetic retinopathy and 233 type 2 diabetes mellitus (T2DM) cases without any sign of retinopathy, taken as controls. All the cases and controls were recruited from Dr. Daljit Singh Eye Hospital, Amritsar. All the participating individuals underwent complete ophthalmologic examination that included visual acuity testing, Humphrey's perimetry, optical coherence tomography (OCT) and fundus examination after pupil dilation. The genotyping was performed by direct DNA sequence analysis of the amplified products. Statistical analysis was performed using SPSS for windows (Version. 16.0). Genotype distributions and allele frequencies were evaluated using statistical calculator (StatPac V. 3. 0). MedCalc (V 9. 3. 9. 0) was used to determine the odds ratios (OR) and 95% confidence intervals (CI). **Results:** The alleles (p=0.8842) and genotypes frequencies (p=0.6304) showed no statistical differences when DR cases were compared with controls (T2DM). OR analysis also revealed no significant co-relation of -106 C/T polymorphism with DR. **Conclusions:** Present findings suggest lack of association of -106 C/T polymorphism in *AKR1B1* with diabetic retinopathy in tested patients from north India.

1017T

Menopausal age shares a common genetic background with diabetes and lipid traits: a study on 13.484 Finns. *A. Joensuu[1, 2], J. Kettunen[3, 2], S. Ripatti[1, 2, 4], J. Sinisalo[5], M.S. Nieminen[5], M-L. Lokki[6], A. Jula[7], J.G. Eriksson[8], V. Salomaa[2], M. Perola[2, 1, 9], K. Auro[2, 1].* 1) Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland; 2) National Institute for Health and Welfare (THL), Helsinki, Finland; 3) University of Oulu, Finland; 4) Hjelt Institute, University of Helsinki, Finland; 5) Heart and Lung Center HUCH, Helsinki University Central Hospital, Finland; 6) Haartman Institute, University of Helsinki, Finland; 7) National Institute for Health and Welfare (THL), Turku, Finland; 8) Department of General Practice and Primary Health Care, University of Helsinki, Finland; 9) University of Tartu, Estonia.

Background: The average age for menopause in western countries is 51 years, but individual variation is high with natural menopause occurring between 40-60 years of age. Later age at menopause has been associated with decreased risk of cardiovascular diseases and osteoporosis as well as increased life expectancy. However, associations between menopausal age and diabetes are controversial and no genetic links have been reported. Smoking has been shown to have a lowering effect on menopausal age whereas obesity raises the age at natural menopause, possibly due to the secretion of hormones from adipose tissue. These and other confounding environmental effects might disturb some of the reported associations between menopausal age and diseases. Methods: Stolk et al. performed a large meta-analysis of genome-wide association studies reporting 17 genetic variants which associate with age at natural menopause (Nat.Genet. 2012). We constructed a genetic score of menopausal age by summing the reported effect sizes (years/allele) of these polymorphisms in five Finnish cohorts (FINRISK1997, PredictCVD, Health2000, Corogene and HBCS) and studied its associations to prevalent and incident diabetes. Logistic and survival analysis corrected by age and geographical variables were performed in individual cohorts and combined in meta-analysis.

Results: One-year increase in the genetic score (range -2.30 - 3.13 years) associated significantly with future diabetes in women (N=2831, hazard ratio=1.46, P=0.0083) but not in men (P=0.3). Interestingly, the genetic score associated nominally with prevalent diabetes in men (N=6249, odds ratio=0.86, P=0.016) but no association was seen in women (P=0.7). In women nominal association of higher menopausal age score was seen also in linear analysis with higher triglyceride levels (P=0.021) and lower HDL cholesterol (P=0.025), but no associations were detected in men. Conclusions: The timing of menopause may share a common genetic background with metabolic endpoints. Our results suggest that genetic variants previously associated with menopausal age have a gender-specific impact on future risk of diabetes: in women these genetic variants increased the risk of future diabetes 1.5-fold, whereas in men the effect was slightly protective on prevalent diabetes. These results may provide valuable insights on biomolecular mechanisms underlying metabolic traits.

1018S

eSNP regulators of genes underlying Mendelian diseases are enriched among T2D-associated variants. *J. Torres, K. Shah, N. Cox.* Medicne/Genetic Medicine, The University of Chicago, Chicago, IL, IL.

Thousands of associations between Mendelian and complex diseases have been recently detected through extensive data mining of medical records from more than 110 million patients and constitute a "non-degenerate Mendelian code" (Blair et al., Cell 2013). Given that common variants within Mendelian disease (MD) genes implicated by this code were found to be significantly associated with common diseases (Blair et al. 2013), we hypothesized that regulators of MD gene expression would be overrepresented among top signals from GWAS on T2D. We evaluated single nucleotide polymorphisms associated with gene expression (eSNPs) mapped in human adipose, skeletal muscle, lymphoblastoid cell lines (LCLs) and nine additional tissues mapped by the GTEX Consortium for enrichment among T2D-associated variants in the Wellcome Trust Case Control Consortium T2D GWAS dataset. We excluded eSNPs for genes underlying monogenic forms of diabetes to ensure that any observed enrichment would not be attributable to effects from established diabetes genes. The proportion of eSNPs for MD genes with false discovery rate (FDR) q-values $\leq 5\%$ is 2×10^{-4} whereas the proportion for all GWAS-interrogated SNPs is 3×10^{-5}. The MD eSNPs most associated with T2D correspond with myeloid differentiation primary response gene (*MyD88*). MyD88 deficiency caused by loss of function mutations has been observed in patients with increased susceptibility to pyogenic bacterial infections. Moreover, the L265P mutation is common in patients diagnosed with Waldenstrom macroglobulinemia. Interestingly, we did not observe an enrichment of eSNPs for MD genes previously associated with T2D. The narrow-sense heritability explained by eSNPs of MD genes is disproportionate relative to the proportion of all SNPs in this set by a factor of 2.18. Taken together, these results support an important yet complex role for genetic regulators of MD genes in T2D susceptibility.

1019M

Association analysis of a *SLC16A11* variant with type 2 diabetes in 12,811 American Indians and evidence for its association with *RNASEK* expression. *M. Traurig[1], A. Marinelarena[1], S. Kobes[1], S. Cole[2], R.L. Hanson[1], W.C. Knowler[1], C. Bogardus[1], L.J. Baier[1].* 1) NIH, Phoenix, AZ; 2) Texas Biomedical Research Institute, San Antonio, TX.

A recent GWAS in Mexican and Latin American populations identified 5 highly concordant exonic SNPs, rs75493593, rs75418188, rs13342232, rs13342692, and rs117767867, in the *SLC16A11* gene that were associated with type 2 diabetes. In the present study, we examined whether these SNPs are also associated with type 2 diabetes in American Indians. Whole genome sequencing data for 234 Pima Indians showed that all 5 of the SNPs were in near complete concordance (r^2=0.99), therefore, rs75493593 was selected as the representative SNP for genotyping in 7,710 American Indians (3,625 were full-heritage Pima Indians and the remaining 4,085 were on average 1/2 Pima and 3/4 American Indian). No association with diabetes was observed (*P*=0.38, OR=1.04 [0.95-1.14] adjusted for age, sex, heritage, and family relationship). To gain more power, rs75493593 was genotyped in two additional samples of American Indians consisting of various tribes who were participants of the Family Investigation of Nephropathy and Diabetes study (*n*=3,095) and Strong Heart Study (*n*=2,421). Analysis of all American Indian samples combined (*n*=12,811) showed a nominal association between rs75493593 and type 2 diabetes (*P*=0.001, OR=1.11 [1.04-1.19] adjusted for age, sex, and study center). To determine a possible mechanism for the association with diabetes, genotypic data for rs75493593 was merged with genome-wide adipose tissue and skeletal muscle gene expression data obtained in a prior study in Pima Indians using the Affymetrix Human Exon 1.0 ST Array. The top genome-wide expression quantitative trait loci (eQTL) for both adipose tissue and skeletal muscle was *RNASEK* located 33 kb upstream of *SLC16A11* (adipose, *n*=187, *P*=0.0004; skeletal muscle, *n*=196, *P*=0.001). The association between rs75493593 and *RNASEK* expression levels in adipose tissue was validated by qRT-PCR in a subset of the microarray subjects (*n*=159, *P*=0.01). These results suggest that rs75493593 or a highly concordant variant may affect *RNASEK* transcription. In conclusion, our study shows that either *SLC16A11* or *RNASEK* could potentially play a minor role in type 2 diabetes for some American Indian populations.

1020T

Asthma susceptibility genetic variants are more strongly associated with phenotypically similar subgroups of patients. *E. Lavoie-Charland[1], J.-C. Bérubé[1], L.-P. Boulet[1], Y. Bossé[1,2].* 1) IUCPQ, Québec, Canada; 2) Department of Molecular Medicine, Laval University, Québec, Canada.

Large scale genome-wide association studies (GWAS) identified polymorphisms reproducibly associated with asthma. Most GWAS were performed with a broadly defined asthma phenotype, leaving the role of susceptibility variants/genes largely undefined. We recently identified four homogeneous subgroups of asthma patients defined by factor and cluster analysis: 1) older patients with low atopy and low lung functions (n=104), 2) patients with high atopy (n=138), 3) young non-smoking women with low blood cell count (n=175), and 4) patients with smoking history (n=105). The goal of this study is to evaluate if GWAS-nominated variants are more strongly associated with asthma patients sharing the same clinical characteristics, which may help reveal the role of recently identified genes. Genotyping of GWAS-nominated variants were performed in the Quebec City Case-Control Asthma Cohort (QCCCAC). The QCCCAC consists of 982 unrelated French Canadian subjects (566 cases and 416 controls) with data collected on lung functions, blood cell counts, atopy, disease history and medication. Genetic association analyses were performed to compare allele frequencies of GWAS-nominated variants between cases and controls. SNPs in or near genes *GSDML, IL13, IL18R1, IL33, LRRC32, RAD50, RORA* and *ZNFN1A3* demonstrated some evidence of replication (p-value < 0.05) for association with asthma. Allele frequencies were then compared between subgroups of phenotypically similar asthma patients and controls. Thirty SNPs were more strongly associated with a subgroup than with asthma in general. Rs10197862, an intronic variant in *IL1RL1/IL18R1*, was the most strongly associated SNP with the high atopy subgroup (p = 0.0009). SNPs located in regions encoding interleukins (*IL1RL1/IL18R1, IL13, IL33*) were associated with the high atopy subgroup or with the young non-smoking women with low blood cell count subgroup. Two SNPs were more strongly associated with the smoking history subgroup than with asthma or any other subgroups: rs1544791 (*PDE4D*) and rs3806932 (*TSLP*). Compared to previous GWAS, sixteen SNPs were only replicated in subgroups, while two were only replicated in the whole cohort. This study shows that a large part of GWAS-nominated SNPs are more strongly associated with more homogeneous subgroups of asthma than broadly defined asthma.

1021S

Pooled targeted resequencing to identify genomic variants associated with crohn's disease in Korea. *C. PARK[1,2], S. HONG[3], S. LEE[1,2,4], J. CHAE[1,2], J. KIM[1,2,4], Y. KIM[4].* 1) Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, South Korea; 2) Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul, Korea; 3) Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; 4) Medical Research Center, Genomic Medicine Institute (GMI), Seoul National University, Seoul, Korea.

Background/Aim: Crohn's disease (CD) is an inflammatory bowel disease. The cause is unknown but recent genome-wide association studies have identified over 100 CD susceptibility loci in Caucasian populations. However, genetic susceptibility loci for CD differ between Asian and Western populations. To identify the genetic susceptibility loci for CD, Next generation sequencing of a pooled DNA samples (pooled sequencing) was performed. **Method:** We performed the pooled sequencing for exons and UTRs of 135 genes associated with CD in 500 CD patients and 1000 controls (pooled 50 DNA samples at equal concentration into 30 pools). We performed sequencing with the Illumina HiSeq 2000. To call variants from the sequencing data, reads were aligned to the human reference genome (HG19) using the Bowtie 2 alignment program, and the variants were called using CRISP with the default parameters. Allele frequency differences between case and control groups were calculated using a chi-squared test or a fisher exact test. For the replication study, 500 CD patients and 1000 controls (no pooled sequencing samples) were genotyped by TaqMan SNP genotyping assay. SNPs with deviation from Hardy-Weinberg equilibrium (p<0.001) were excluded. Association analysis was performed using PLINK. **Results:** In pooled sequencing result, we identified a total of 5,854 variants. Significant difference in allele frequencies between cases and controls were observed for 33 common variants in 12 genes (quality score>15000, p<0.01) and 4 rare exonic variants in 4 genes (quality score>300, p<0.02). For the replication study, 11 common and 4 rare variants were selected. In replication study, we confirmed two previously reported SNPs associated CD in Korean: rs3810936 in *TNFSF15* (OR=1.89, p=1.11×10-15) and rs76418789 in *IL23R* (OR=0.517, p=6.75×10-6). We identified one SNP rs2241880 in *ATG16L1* (OR=1.37, p= 0.00015) that failed to be replicated previously in Asian population. We also found five novel CD susceptibility loci: rs3181374 in *TNFSF8* (OR=1.639, p=2.39×10-10), rs28362680 in *BTNL2* (OR=1.452, p=6.76×10-06), rs3208181 in *HLA-DQA2* (OR=1.38, p=0.0009), rs1053004 in *STAT3* (OR=1.29, p=0.002) and rs2273650 in *NFKBIA* (OR=0.81, p= 0.0192), and one rare CD susceptibility locus: rs200735402 in *CARD9* (OR=0.07, p= 0.0002). **Conclusion:** Using pooled sequencing, we identified new susceptibility variants for CD as well as confirmed previously reported loci in GWAS.

1022M

Association of HLA-DPB1 alleles with CHB infection and HBV related HCC in Asia. *N. Nishida[1,2], H. Sawai[2], K. Kashiwase[3], M. Sugiyama[1], Y. Mawatari[1], K. Tokunaga[2], M. Mizokami[1].* 1) Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Chiba, Japan; 2) Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo, Japan; 3) HLA Laboratory, Japanese Red Cross kanto-Koshinetsu Block Blood Center, Koutou-ku, Tokyo, Japan.

In the previous report, HLA-DPA1*01:03-DPB1*04:02 and HLA-DPA1*02:02-DPB1*05:01 showed significant associations with chronic hepatitis B (CHB) infection in Japanese as a protective and a risk haplotype, respectively (Kamatani et al., 2009). However, there have been no reports of HLA-DP genes to be associated with disease progression from CHB to liver cirrhosis (LC) or hepatocellular carcinoma (HCC). We conducted HLA-DP genotyping using a total of 3,167 samples (including Japanese, Korean, Hong Kong and Thai) for HBV patients (including CHB, LC and HCC), healthy controls and resolved individuals (HBsAg-negative and anti-HBc-positive), based on PCR-SSO system according to manufacturer's protocol. The Fisher's exact test in a two-by-two cross table was applied to acquire exact P values. For a meta-analysis in four populations, we used the DerSimonian-Laird method, which assumes a random-effects model to calculate the pooled effect estimate. A total of 2,895 samples were successfully genotyped. There were a total of 272 dropped samples in HLA-DP genotyping, a large part of which were belonging to Korean HBV patients and Hong Kong resolved individuals, presumably due to low concentration or low quality of DNA samples. Consequently, we identified one high-risk haplotype (HLA-DPA1*02:01-DPB1*09:01) and one protective haplotype (HLA-DPA1*01:03-DPB1*04:01) to be associated with CHB infection over the previously reported HLA-DP haplotypes in Asian populations. Moreover, a significant association of DPB1*02:01 was identified to be protective not only against HBV infection, but against disease progression from CHB to HCC. Trans-ethnic association study of HLA-DP in Asian populations revealed that specific HLA-DPB1 alleles (i.e. DPB1*02:01, *04:01, and *04:02) worked to be protective against HBV infection, and different alleles (i.e. DPB1*05:01, *09:01) worked to be susceptible to HBV infection. To determine all the associated DPB1 alleles for HBV infection would enable HBV infected individuals to divide into two groups who need treatment or not.

1023T

New mutations in the *MYOC* gene in patients with juvenile open-angle glaucoma. *J.P.C. Vasconcellos[1], P.V. Svidnicki[2], C.A. Braghini[2], M.B. Melo[2].* 1) Dept Ophthalmology, Univ Campinas - UNICAMP, Campinas SP, Brazil; 2) CBMEG, Univ Campinas - UNICAMP, Campinas SP, Brazil.

Glaucoma is the leading cause of irreversible blindness and is characterized by progressive optic disc cupping with corresponding visual field loss. Both intraocular pressure (IOP) and positive family history are risk factors for the development of the disease. Juvenile open angle glaucoma (JOAG) is a more severe form of open angle glaucoma characterized by an early onset (10 to 35 years of age). Mutations in the myocilin gene (*MYOC*) account for most cases of JOAG. This study was approved by the Ethics Committee of the University of Campinas. DNA samples from 98 unrelated Brazilian patients with JOAG and 92 normal individuals were screened for mutations in the three exons and intron/exon junctions of the *MYOC* gene through direct sequencing. In order to evaluate clinical aspects, ophthalmologic examination included evaluation of the visual field and optic disc, intraocular pressure measurement, and gonioscopy. Mutation screening revealed two new glaucoma causing mutations: the frame-shift mutation 1356delT and X505W. Besides, the previously described glaucoma-causing mutations: C433R (27 patients), Q368X (2 patients), P370L (1 patient) and K423E (1 patient) were also identified. Among neutral variants, we observed the previously reported G122G (1 patient), R76K (14 patients), P13P (1 patient), 730+35g>a (53 patients) and the unreported variants *15 t>a (2 patients), *54g>t (1 patient) and *393 c>g (1 patient). This study provides an overview of *MYOC* mutation spectra and frequencies in JOAG Brazilian patients. These findings may be important for the increase of *MYOC* mutations panel and their effects on phenotype. Financial Support: Fapesp 02/11575-0.

1024S

Association of ADIPOQ rs266729, rs17300539, rs2241766 and rs17846866 with Type 2 Diabetes and Diabetic Retinopathy in North-West Indian population. *A. BHANWER, R. SIKKA, K. MATHAROO.* Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

ADIPOQ is an adipokine gene present on chromosome#3 (3q27.3) and is known to be associated with Type 2 diabetes (T2D) and metabolic syndromes. It is specifically and abundantly expressed in the adipose tissue and sensitizes the body to insulin. Adiponectin act as an anti-inflammatory adipokine with anti-diabetic properties. However, the role of ADIPOQ polymorphism has not been clarified but it seems to play a role in vascular damage by interfering with anti-inflammatory and anti-diabetic effects of adiponectin, which may contribute to worsening or acceleration of microvascular complications. Polymorphisms in ADIPOQ gene influences the protein function leading to acceleration of microvascular complications through vascular damage. Therefore, in the present study, we aimed to explore the role of ADIPOQ rs266729, rs17300539, rs2241766 and rs17846866 with the aetiology of T2D and diabetic retinopathy (DR). A total of 454 individuals (207 DR and 247 T2D) and 241 gender matched above 50 years of age healthy controls were included from the region of Amritsar. Genotyping was done by using PCR-RFLP for rs266729, rs17300539, rs2241766 and PCR-ARMS for rs17846866. rs266729 CG+GG and GG genotypes tend to give 1.8 and 1.9 fold risk, respectively towards DR when compared with T2D cases and 1.7-1.9 fold risk towards DR when compared to healthy control. For rs17300539, AA genotype tends to give 4 fold risk towards DR when compared with T2D cases however; AA+GA and GA genotypes tend to give 2-4 fold risk towards DR when compared with healthy controls. In rs2241766, wild genotype TT gives 2 and 2.5 fold risk towards DR when compared with T2D and controls, respectively. However, TG genotype tends to give protection towards DR when compared with T2D [OR=0.39 (0.21-0.73)] and healthy controls [OR=0.35 (1.9-0.65)]. In case of rs17846866, GG+TG and GG genotypes tend to give 2 fold risk towards DR when compared with T2D and healthy controls. However, no association was observed for rs266729, rs2241766 and rs17846866 with T2D. In conclusion, rs266729, rs17300539, rs2241766 and rs17846866 seem to be associated with risk of DR when compared with T2D cases as well as with healthy controls in the population of Punjab.

1025M

More evidence for association of a rare TREM2 variant (R47H) with Alzheimer's disease risk. *S.L. Rosenthal[1], M.N. Bamne[1], X. Wang[1], S. Berman[2], B.E. Snitz[2], W.E. Klunk[3], R.A. Sweet[3], F.Y. Demirci[1], O.L. Lopez[2], M.I. Kamboh[1].* 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 2) Department of Neurology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 3) Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA.

Over twenty risk loci have been identified for late-onset Alzheimer's disease (LOAD). Among them, only one, *APOE*, produces a strong effect size for this neurodegenerative disease. Individuals homozygous for the *APOE*4* allele have a 60% increased lifetime risk of developing LOAD. Recent sequencing efforts have detected a rare variant in *TREM2*, R47H, that has an effect size similar to that of *APOE*. In our study, we genotyped this variant in a case-control population of Caucasian descent derived from two study cohorts: 1283 cases (mean age-at-onset, AAO=73.26 $\pm$ 6.76) and 996 controls (mean age= 75.67 $\pm$ 6.37) from the University of Pittsburgh Alzheimer's Disease Research Center, and 338 cases (mean AAO=84 $\pm$ 3.98) and 1950 controls (mean age=83.15 $\pm$ 9.93) from the Gingko Evaluation of Memory study. In the total sample, carriers of the R47H variant had a significantly increased risk of having LOAD (OR=7.40, p=3.66E-06). We also examined the effect of this variant on psychosis in LOAD and Aβ deposition (as measured by Pittsburgh compound B uptake) and found no significant association between this variant and either phenotype. We have successfully replicated the association of the *TREM2* R47H variant with LOAD risk in Caucasians. Combined with other studies, our results provide further evidence for the involvement of *TREM2*, specifically the R47H variant, in modulating LOAD risk.

1026T

The *IL10* ACA haplotype is associated with rheumatoid arthritis in patients from Western Mexico. *J. Hernandez-Bello[1], M. Vazquez-Villamar[1], E. Oregon-Romero[1], Y. Valle[1], J.R. Padilla-Gutierrez[1], G. Martinez-Bonilla[2], J.F. Muñoz-Valle[1].* 1) Instituto de Investigación en Ciencias Biomedicas, CUCS, Universidad de Guadalajara. Guadalajara, Jalisco, Mexico; 2) Servicio de Reumatologia, Hospital Civil de Guadalajara, Fray Antonio Alcalde. Guadalajara, Jalisco Mexico.

Background: Rheumatoid arthritis (RA [MIM 180300]) is a chronic autoimmune disease of unknown etiology affecting approximately 1% of the population worldwide. RA is characterized by a cartilage and bone erosion resulting from chronic synovial inflammation. Interleukin 10 (IL-10) is an immunomodulatory cytokine involved in the regulation of synovitis, the B cells activation and in a subsequent production of autoantibodies such as anti-cyclic citrullinated peptide (anti-CCP) and rheumatoid factor (RF). Three polymorphisms of *IL10* gene at -1082 A>G (rs1800896), -819 C>T (rs3021097) and -592 C>A (rs1800872) form three major haplotypes (GCC, ACC, and ATA). These haplotypes have been associated with high, medium, and low IL-10 production and have been reported as a risk factor for autoimmune diseases. However, their distribution and association with RA in population from Western Mexico is unknown. **Aim:** Evaluate the association of *IL10* haplotypes with susceptibility to RA in patients from Western Mexico. **Materials and methods:** We studied 260 RA patients classified according to ACR/EULAR 2010 criteria and 260 control subjects (CS), both study groups from Western Mexico. The -1082 A>G, -819 C>T and -592 C>A polymorphisms of the *IL10* gene were genotyped by using PCR-RFLP. The haplotype frequencies were inferred using the HEMHAPFRE software and the statistical analysis was performed using Stata software, release 9.0. **Results:** In RA and CS eight haplotypes were constructed after performing haplotypic analysis by using maximum likelihood method and their frequencies were: ACC (44.25% vs 46.53%), ATA (25.4% vs 23.18%), GCC (12.64% vs 14.23%), GTA (13.42% vs 14.13%), ATC (1.9% vs 1.05 %), ACA (1.53% vs. 0.19%), GTC (0.44% vs. 0.69%) and GCA (0.43% vs 0%). We observed significant difference in the distribution of genotypic frequencies between both study groups ($p = 0.04$), with an OR 8.7 (95 % CI 1.04-67.83) for the ACA haplotype, which indicates an association of the ACA haplotype with an increased risk of RA. Moreover, no significant difference was observed between the distribution of the three polymorphisms in RA and controls. **Conclusions:** The ACA haplotype of the *IL10* gene may be a marker for the susceptibility to RA in populations from Western Mexico.

1027S

Serum Uric Acid Levels Are Associated with Polymorphism in the SLC2A9, SF1 and GCKR Genes in Chinese Subjects. *C. Hu, X. Sun, R. Zhang, W. Jia.* Shanghai Diabetes Inst, Shanghai, Shanghai, China.

Recent reports on genome2wide association studies focusing on serum uric acid concentrations identified several novel associated loci in European descent samples. In the current study, we aimed to evaluate the association between these loci and concentrations of serum uric acid in the Chinese. We genotyped fourteen single nucleotide polymorphisms (SNPs) from eleven loci mapped in or near PDZK1, GCKR, LRP2, SLC2A9, ABCG2, LRRC16A, SLC17A1,SLC17A3, SLC22A11, SLC22A12 and SF1 in the Shanghai Chinese, including 2329 participants from a community based study. As uric acid metabolism differs between genders, all of the variants were analyzed for gender differences. SNPs including rs780094 in GCKR, rs11722228 in SLC2A9 and rs606458 in SF1 showed association to serum uric acid levels in males after adjustment for age and BMI as confounders (p= 0.016, 0.0011 and 0.030, respectively); SLC2A9 rs3775948 demonstrated a trend towards association to uric acid (p=0.071). In females, only rs506338 in SLC22A12 was detected to potentially associate with uric acid (p=0.0569). After combined analysis between males and females, we only detected that SLC2A9 rs3775948 and SF1 rs606458 were related to uric acid(p=0.036 and 0.043, respectively). The interaction between rs11722228 in SLC2A9 and gender on serum uric acid levels existed in our samples. In conclusions: We showed that GCKR,SLC2A9 and SF1 variants modulated uric acid levels in the Shanghai Chinese. Moreover, gender difference between GCKR and SLC2A9 polymorphisms and uric acid levels was also found in our study. The findings provide a perspective that genetic variation is one of the key influences upon regulation of serum uric acid levels in humans.

1028M

Transferrin receptor and hereditary hemochromatosis gene variants interact to modify childhood leukemia risk. *A.E. Kennedy[1], E. DeRycke[1], G.Y. Lai[1], K.Y. Kamdar[2], P.J. Lupo[2], M.F. Okcu[2], M.E. Scheurer[2], M.K. Baum[3], D. Seminara[1], M.T. Dorak[4].* 1) Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH, Rockville, MD; 2) Section of Hematology-Oncology, Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) Department of Dietetics and Nutrition, Robert Stempel College of Public Health & Social Work, Florida International University, Miami, FL; 4) School of Health Sciences, Liverpool Hope University, Liverpool, United Kingdom.

An interaction between the transferrin receptor (*TFRC*) rs3817672 (S142G) and hereditary hemochromatosis gene (*HFE*) rs1800672 (C282Y) is associated with increased risk for multiple myeloma, breast, colorectal and hepatocellular cancers as well as childhood acute lymphoblastic leukemia (ALL) risk in European studies. *HFE* C282Y and H63D correlate with body iron levels. Since HFE and TFRC proteins biologically interact in iron transfer across membranes, the observed statistical interaction suggests that the involvement of *HFE* variants in cancer risk modification is mediated via their synergistic effect on body iron levels. We sought to replicate the *HFE*x*TFRC* interaction in childhood ALL in a US-based study, and used bioinformatic tools to assess the contribution of *TFRC* S142G to this association. Genotyping was conducted in a multi-ethnic sample from Houston, TX (161 incident cases with childhood ALL and 231 controls, all <18 yr). Being positive for either *HFE* variant yielded an elevated odds ratio for childhood ALL risk in males (1.4, 95% CI=0.8 to 2.4), which increased to 3.0 (95% CI=1.3 to 6.8) in the presence of the S142G allele A homozygosity ($P_{interaction}$=0.04). A comprehensive screening of empirical and bioinformatic data showed that allele A shows allele-specific expression, the non-conservative amino acid substitution is within the helix portion of the protein and predicted to alter the secondary structure (but not in the transferrin or HFE binding sites), and the SNP alters a splicing site. S142G itself is not associated with any disease individually, except that it is in linkage disequilibrium ($r^2 > 0.8$) with rs9859260, which correlates with a red cell index. None of the other 21 variants in LD with S142G has a higher functionality score or a trait association. One plausible mechanism may be that S142G generates a splice variant that interacts with HFE differently. Thus, it appears that S142G is the causal SNP interacting with HFE in cancer risk associations. The interaction of *HFE* with *TFRC* in modification of risk for cancers has now been replicated in several studies and in multiple cancers. Although *HFE* and *TFRC* variants do not show individual associations with cancer risk in GWAS, their interaction should be explored in existing GWAS data in cancer to ultimately establish it as a possible risk factor. The biological basis of this association has implications on public health and policy on unregulated use of iron supplements.

1029T

Pro12Ala polymorphism in *PPARG* gene in mexican patients with systemic lupus erythematosus. *G.M. Mimendi Aguilar[1,2], J. Topete[3], I. Feldhoffer[2,4], J. Soto[3], M. Medina[3], I. García[5], R. Parra[6], I. Dávalos[1,7], M. Salazar[2], M. Morán[1,2].* 1) Doctorado en Genética Humana, Universidad de Guadalajara, Centro Universitario de Ciencias de la Salud. Sierra Mojada 550, Colonia Independencia, Guadalajara, Jalisco, Mexico; 2) División de Medicina Molecular, Centro de Investigación Biomédica de Occidente. Sierra Mojada 800, Colonia Independencia, Guadalajara, Jalisco, Mexico; 3) UMAE, Hospital de Especialidades, CMNO, IMSS. Belisario Domínguez 1000, Guadalajara, México; 4) Universität Hohenheim , Alemania; 5) Hospital General de Occidente, SSJ. Av Zoquipan 1050, Jalisco, México; 6) Hospital General Regional 46. IMSS. Lázaro Cárdenas 44100; 7) División de Genética, Centro de Investigación Biomédica de Occidente, IMSS. Sierra Mojada 800, Colonia Independencia, Guadalajara, Jalisco, México.

Introduction: Systemic lupus erythematosus (SLE) is a heterogeneous disease involving several immune cell types and pro-inflammatory signals.- Peroxisome proliferator-activated receptor (PPAR) polymorphisms have been associated with a variety of metabolic and kidney disorders. The Pro12Ala polymorphism (rs1801282) in PPARG gene, a Pro-to-Ala exchange that results in the substitution of proline with alanine at codon 12, was previously associated with diabetic nephropaty; we therefore hypothesized that Pro12Ala PPARG polymorphism might be involved in the pathophysiology of lupus nephropaty in Mexican patients with SLE. Aim. To analyze the allelic and genotype frequencies of Pro12Ala polymorphism in the PPARG gene in general population and patients with systemic lupus erythematosus (SLE) with or without lupus nephritis (LN) from Western Mexico. Material. Cross-sectional study. Fifty mestizo Mexican individuals from general population from Norwestern México, 17 patients from the nephrology clinic with a diagnosis of SLE and LN and 20 patients with SLE without evidence of LN (according the American College of Rheumatology (ACR), and the International Society of Nephrology and Renal Pathology Society (ISN / RPS) criteries) were included. Methods: Genomic DNA was extracted from a blood sample using a conventional method. A PCR-RFLPs protocol was used to amplify a fragment of 306 base pairs (bp) of PPARG gene, and then digested with the Hga I restriction enzyme. The digestion products were identified by gel electrophoresis on 10% polyacrylamide (29:1) stained with silver nitrate. Results: The general population is in Hardy Weinberg Equilibrium. We found only one individual in the group of patients without the variant submitted NL homozygous genotype (Ala12Ala). The genotype and allele frequencies in the three groups are shown in Table 1. Statistical analysis was performed using the Finetti diagram and only statistically significant differences were found when comparing the alleles in SLE patients with or without NL, where the risk allele is 12Ala. (OR = 9.6, C.I. 2.115-43.583 p=0.00197). (*) Conclusions: This is the first report of an association analysis of PPARG Pro12Ala gene polymorphism with SLE in Mexican population and the results suggest that the presence of the variant allele (Ala) confers risk for LN; however it is necessary to increase the sample size to confirm this assertion.

1030S

Quantification of Hirschsprung disease susceptibility from common polymorphisms in relation to gender, segment length of aganglionosis and familiality. *A. Kapoor[1], Q. Jiang[1, 2], C. Berrios[1], A. Chakravarti[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 2) Department of Medical Genetics, Capital Institute of Pediatrics, Beijing, China.

Hirschsprung disease (HSCR), a congenital developmental defect of the enteric nervous system, is a multifactorial disorder in which common, low penetrance, noncoding variants at *RET*, *NRG1* and *SEMA3* underlie disease risk, primarily in male, short segment simplex (~70%) cases. Rare, high penetrance, coding variants in at least a dozen genes collectively underlie disease risk in <25%, enriched for female, long segment multiplex cases. We have used the largest collection of HSCR cases and relatives of European ancestry ascertained in the USA (~1100 individuals: 365 probands, 40 affected relatives and 710 unaffected relatives) to directly assess the joint contribution of *RET* (rs2435357, rs2506030), *NRG1* (rs7835688, rs16879552) and *SEMA3* (rs1583147, rs12707682, rs11766001) common variants and risk covariates to HSCR susceptibility. Case-control analysis in 342 cases and 379 controls, showed significant associations at *RET* at rs2435357 (odds ratio (OR) = 3.7 and $P=1.2\times10^{-31}$) and at rs2506030 (OR= 1.9 and $P=3.5\times10^{-9}$), and at *SEMA3* rs11766001 (OR=1.5 and P=0.002). These results were corroborated by transmission disequilibrium tests in 243 trios demonstrating that these associations were not due to cryptic population structure (OR=4.7, $P=5.1\times10^{-25}$; OR=1.9, $P=3.6\times10^{-6}$; OR=1.7, P=0.002 for rs2435357, rs2506030 and rs11766001, respectively). We failed to find any evidence of a reported *NRG1* effect, indicating that this association is likely restricted to Asians where it was first reported. HSCR disease risk was directly related to and highly significant by dosage of 3+ risk alleles at rs2435357, rs2506030 and rs11766001 (OR=1.85, P=0.001; OR=3.29, $P=5.95\times10^{-7}$; OR=10.35, $P=6.56\times10^{-5}$ for 3, 4 and 5+ alleles, respectively); fewer alleles were protective. These risks translated into a significantly higher incidence (penetrance) by allelic dosage: 24.6, 42.2 and 141.3 cases per 100,000 live-births as compared to the population average of 15/100,000. Common allele dosage-based disease risk was significantly higher in males (P=0.005), and higher but not significant in short segment and simplex cases. Finally, despite our over-ascertainment of familial cases, sibling recurrence risk is 29% in sibships of probands with <3 risk alleles as compared to 19% when probands had ≥3 risk alleles. Consequently, these common polymorphisms delineate a class of frequent HSCR cases that have lower recurrence rates within sibships, a feature that should inform genetic counseling.

1031M

Association between common variant near *CAV1* and *CAV2* genes and phenotypic features of primary open-angle glaucoma. *F. Mabuchi[1], Y. Sakurada[1], S. Yoneyama[1], K. Kashiwagi[1], Z. Yamagata[2], H. Iijima[1].* 1) Dept Ophthalmology, Univ Yamanashi, Chuo, Yamanashi, Japan; 2) Dept Health Sciences, Univ Yamanashi, Chuo, Yamanashi, Japan.

Purpose: It was reported that common variants near caveolin 1 (*CAV1*) and caveolin 2 (*CAV2*) genes were associated with primary open-angle glaucoma (POAG). This study was performed to assess the association between the common variant near these genes and the phenotypic features in patients with POAG, including normal tension glaucoma (NTG) and high tension glaucoma (HTG). Methods: Four hundred and one Japanese patients with POAG, including 176 patients with NTG and 225 patients with HTG, and 191 control subjects without glaucoma were analyzed for the common variant (rs1052990) near *CAV1* and *CAV2* genes. The genotype and allele frequencies were compared between the patients with NTG or HTG and the control subjects. Demographic and clinical features, including age at diagnosis of glaucoma, gender, family history of glaucoma, refractive error, maximum intraocular pressure (IOP), vertical cup-to-disc ratio, and history of glaucoma surgery, were compared between the genotypes in patients with POAG. A multiple linear regression analysis was carried out with the maximum IOP as a dependent variable and age, gender, and the common variant as independent variables. Results: There was a significant difference (P = 0.024, analysis of variance) of the maximum IOP among GG (21.8 ± 4.6mmHg, mean ± standard deviation), GT (22.7 ± 6.6mmHg), and TT (24.9 ± 9.2mmHg) genotypes in patients with POAG, although no significant differences of the genotype and allele frequencies could be found between the patients with NTG (GG: 4.0%, GT: 36.4%, TT: 59.6%, P = 0.90; G allele: 22.2%, T allele: 77.8%, P = 0.99) or HTG (GG: 4.9%, GT: 31.1%, TT: 64.0%, P = 0.76; G allele: 20.4%, T allele: 79.6%, P = 0.61) and the control subjects (GG: 4.7%, GT: 34.6%, TT: 60.7%, G allele: 22.0%, T allele: 78.0%). Based on multiple linear regression analysis, a significant relationship was confirmed between the common variant and the maximum IOP ($\beta = 0.12$, standard error = 0.70, P = 0.017). Conclusion: The association of the common variant near *CAV1* and *CAV2* genes with a maximum IOP in patients with POAG suggests that this locus may contribute to the pathogenesis of POAG as a genetic risk factor associated with IOP elevation.

1032T

GSTT1 and GSTM1 gene frequency in Punjabi population exposed to pesticides. *M. Ahluwalia, A. Kaur.* Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

The glutathione-S-transferase (GST) enzyme system play an important role in defence mechanism against environmental chemicals and endogenous substances. Many of these enzymes show variability among individuals due to their polymorphic character and thus show difference in predisposition to various diseases such as cancers, neurological disorders, endocrine system disruption, and etc. The aim of the study was to determine GSTT1 and GSTM1 gene deletions in 69 individuals from different regions of Punjab involved in agricultural activities such as preparation, loading, spraying and storage of pesticides, harvesting of crops etc. Genotyping was done by Multiplex Polymerase chain reaction approach. The deleted GSTT1 and GSTM1 genes show frequency of 14.5% and 50.7% respectively. The results suggest an increased tendency of exposure related diseases due to complete gene deletion of either of the two genes. However, 5.7% population has shown deletion of both the genes and are expected to have significantly reduced detoxification tendency and thus an increased risk to diseases caused by exposure to environmental carcinogens. The study thus provides the basis for gene-environment association studies to be carried out. The detection of oxidative stress in serum by biochemical methods is in progress.

1033S

Relationships between genetic ancestries and nicotine and tobacco carcinogen metabolisms in the Multiethnic Cohort. *H. Wang[1], S. Park[2], C. Haiman[2], S. Murphy[3], S. Hecht[3], B. Henderson[2], L. Kolonel[1], D. Stram[2], L. Le Marchand[1].* 1) UH Cancer Center, Honolulu, HI; 2) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) Masonic Cancer Center, University of Minnesota, Minneapolis, MN.

It has been reported in the Multiethnic Cohort (MEC) study that major (2-to-5-fold) differences exist among US racial/ethnic minorities in the risk of lung cancer associated with cigarette smoking, even after taking into account self-reported dose and duration. Native Hawaiians and African Americans have a higher risk for lung cancer than European Americans, who in turn have a higher risk than Japanese Americans and Latinos. Here we tested the hypothesis that genetic ancestries are related to nicotine metabolism (up to 80% through C-oxidation by CYP2A6) and internal dose of tobacco smoke carcinogens, thus contributing to the observed differences in lung cancer risk across populations. Genetic ancestries were estimated using STRUCTURE based on 121 ancestral informative markers for 2,170 MEC African Americans, European Americans, Japanese Americans, Latinos, and Native Hawaiians, along with 5 HapMap reference populations, namely, ASW, CEU, YRI, JPT and MEX. Percent ancestries, i.e. African, American Indian, Native Hawaiian and East Asian ancestries were obtained and categorized according to deciles. Nicotine metabolism (total trans-3'-hydroxycotinine: free cotinine) and exposure to and detoxification of nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), a major tobacco-specific lung carcinogen, were regressed on percent ancestries in a linear regression model, adjusting for age, sex, assay batch, creatinine and internal dose of nicotine. African ancestry was related to a lower level of CYP2A6 activity (P = 0.12) and a higher level of exposure to (P = 0.03) and non-significant lower level of detoxification of (P = 0.93) NNK. East Asian and Native Hawaiian ancestries were related to lower levels of CYP2A6 activity (P's ≤ 0.002), lower levels of exposure to (P's ≤ 0.04) and of detoxification of (P's <0.0001) NNK. American Indian ancestry (in Latinos) was related to a higher level of CYP2A6 activity (P < 0.0001), a non-significant lower level of exposure to (P = 0.81) and significant lower level of detoxification of (P = 0.002) NNK. These results are consistent with and could partly explain the observed differences in lung cancer risk due to smoking in African Americans, East Asians, Latinos, compared to European Americans, but could not explain the high disease risk of Native Hawaiians, suggesting more factors contribute to risk in this recently admixed population.

1034M

Haplotype association and synergistic effect of Renin-angiotensin aldosterone system gene polymorphisms causing susceptibility to essential hypertension in Indian patients. *C. Bhupatiraju[1], MV. Uppuluri[1], P. Gunda[1], D. Pandharpurkar[2], P. Tirunilai[1]*. 1) Dept of Genetics, Osmania University, Hyderabad, India; 2) Gandhi Medical College and Hospital, Hyderabad, India.

The present study investigates association of essential hypertension (EHT) with candidate gene polymorphisms involved in vasoconstriction (*REN, AGT, ACE, AGTR1* and α_{2B}*-ADR*) and vasodilation (*BDKRB2*); haplotype association and epistatic interactions between them. Evaluation of 316 hypertensives and 293 controls for various epidemiological parameters confirmed the association of gender, obesity, family history and non-vegetarian diet with EHT. Study of candidate gene polymorphisms suggests significant association of -g.5434C>T and g.18-83G>A of *REN*; g.-217G>A of *AGT*; ±9bp of *BDKRB2* and I/D of α_{2B}*-ADR* with EHT. The study suggests high risk for developing EHT for a) CT genotypes and allele C carriers of g.-5434C>T, which was high in females b) females with II genotype of α_{2B}*-ADR* I/D and in those consuming non-veg and males with ID genotype and alcoholics c) males with g.-217A variant of *AGT* and d) +9bp allele carriers of *BDKRB2* ±9bp especially in males. The positive association of g.-217A variant and +9bp variant with EHT may be correlated with varied transcriptional activities of the respective genes resulting in vasoconstriction. Analysis for epistatic interactions suggest a high risk for individuals with i) g.-5434CC and g.18-83GA, ii) g.-5434TT and g.18-83GA and iii) g.-5434TT and g.18-83GG genotypic combinations of *REN* gene causing susceptibility to EHT. High risk with respect to epidemiological factors was observed for females with positive family history of EHT. The gene-environmental interaction analysis suggests strong synergisism between α_{2B}*-ADR* I/D polymorphism and gender. Estimates of the degree of linkage disequilibrium (LD) revealed a complete linkage disequilibrium between g.-152G>A and p.M235T and between p.T174M and g.-5312C>T polymorphisms in both hypertensives and controls. Among hypertensives complete LD was observed between g.-152G>A and g.-58C>T, g.-20C>A and g.1166A>C, p.T174M and g.1166A>C and g.-5434C>T and g.18-83G>A polymorphisms, while LD between these polymorphisms was absent or weaker among the controls. In general the study suggest the need for evaluation of life style habits like smoking, alcohol and diet at an early age and monitoring the BMI levels regularly. The genes showing polymorphisms with significant risk of developing EHT may provide clues for designing drugs whose mode of action would be genotype based. Any such effort to ameliorate the suffering due to the disease improves the quality of life of affected.

1035T

Genetic variation and Insulin Resistance in middle age Chinese men. *r. villegas[1], r. Delahanty[1], s. Williams[2], r. o'bryan[1], y. Gao[3], q. cai[1], H. li[3], y.x. Xiang[3], w. zheng[1], xo. Shu[1]*. 1) Vanderbilt University Medical Center, 2525 , Nashville TN, 37203-1738 USA; 2) 2Geisel School of Medicine, Dartmouth College, USA; 3) Department of Epidemiology, Shanghai Cancer Institute, 2200/25# Xie Tu Road, Shanghai 200032, P.R. China.

Introduction: Insulin resistance is a necessary condition for T2D to develop and results from an interaction between genetic and environmental factors. Understanding factors associated with insulin resistance and its genetic controls is of particular importance for the prevention of T2D, because insulin resistance is reversible. In this project we evaluated the association of genetic polymorphisms in the first three genes in the insulin signaling pathway (the insulin receptor, INSR, and insulin receptor substrates 1 and 2, IRS1 and IRS2), glucose transporter 4 (GLUT4), and genes identified from GWAS of T2D quantitative traits with insulin resistance. We also evaluated gene environment interactions and insulin resistance. Methods: We investigated the independent and combined effect of variants in the study genes with exercise participation, waist to hip ratio (WHR) and body mass index (BMI) with both fasting insulin and the homeostasis model assessment of insulin resistance (HOMA-IR) as outcomes. We used data from 1952 non diabetic male participants of an ongoing, population-based prospective study conducted in Shanghai, China. Assessment of physical activity was obtained by using a validated physical activity questionnaire and total METs (metabolic equivalent task (MET)-hr/day) was calculated. Anthropometric variables were measured by trained interviewers. The association between single nucleotide polymorphisms (SNPs) and insulin and HOMA-IR was assessed by linear regression analyses. Results: A total of 87 SNPs were included in the analysis. We found one candidate gene, IGF1, had evidence for association with both insulin and HOMA-IR. We observed interactions with BMI and fasting insulin in 2 SNPs from INSR and 2 SNPs in the GLUT4 genes. Only one SNP, rs10417205 had an interaction factor with a P value <0.05 with HOMA-IR. A total of 4 SNPs in the INSR gene, 1 SNP in the INSR2 gene and 2 SNPs in the GLUT4 gene showed interactions with WHR. We also found that 4 SNPs in INSR had a P value for interaction less than 0.05 with total METs of physical activity with HOMA-IR, while we only observed 2 interactions with a P<0.05 with fasting insulin. Conclusion: Our data support the hypothesis that the IGF1 gene variation is associated with insulin resistance among middle age Chinese men. We also found interactions between some of the study genes with physical activity, BMI and WHR with insulin resistance.

1036S

EWAS to GxE: A robust strategy for detecting gene-environment interaction models for age-related cataract. *M.A. Hall[1], J.R. Wallace[1], S.A. Pendergrass[1], R. Berg[2], T. Kitchner[2], P. Peissig[2], M. Brilliant[2], C.A. McCarty[3], M.D. Ritchie[1]*. 1) Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA; 2) Marshfield Clinic, Marshfield WI; 3) Essentia Rural Health, Duluth, MN.

Gene-environment interactions (GxE) are essential to elucidating the underlying nature of complex traits. Despite the importance of GxE, extensive computational demands and adjusting for multiple tests make uncovering these interactions a challenge when explored with an exhaustive combinatorial search. We address this using an environment wide association study (EWAS) to identify predictive environmental factors, followed by a test for interaction with genome-wide SNPs for association with cataract. EWAS tests a diverse set of environmental variables in a high-throughput manner for trait association. Subsequently, putative predictors identified by EWAS can be tested for GxE to further explain the trait. Using PLATO software, we performed a dietary EWAS by evaluating 57 dietary exposures from the National Cancer Institute developed Dietary History Questionnaire using logistic regression, adjusting for age, gender, and type 2 diabetes (T2D) status in 2,629 samples (932 controls, 348 males (M); 1,697 cataract cases, 711 M) of European descent from the Marshfield Clinic Personalized Medicine Research Project, part of the Electronic Medical Records & Genomics (eMERGE) Network. Seven dietary measures were predictive of cataract ($p < 0.05$). A monounsaturated omega-9 fatty acid known as erucic acid (FA22:1) ($p=5.5\times10^{-4}$) met a Bonferroni corrected p-value threshold ($\alpha=0.05$). We then tested FA 22:1 for GxE using 498,829 SNPs in a subset of samples for whom genetic data was available (831 controls, 314 M; 1,511 cases, 633 M) using logistic regression adjusted for age, gender, and T2D status. Twenty SNP-FA22:1 models were statistically significant ($p<1.0\times10^{-4}$). The most significant GxE model was between FA22:1 and rs726712, an intronic SNP in *LPP* ($p=2.9\times10^{-5}$). The erucic acid-cataract association is novel; although significantly depleted levels of two polyunsaturated fatty acids have been found in cataractous human lenses when compared to mature typical lenses. *LPP* encodes a protein involved in cell-cell adhesion. A variety of studies have previously demonstrated the importance of cell-cell adhesion in cataract development, yet this is the first example of association with this specific factor. These findings indicate the role of GxE in susceptibility to cataract and demonstrate the utility of EWAS as a data-driven filtering method for investigating the genetic and environmental interplay of common, complex diseases.

1037M

Polygenic and Localized Genotype by Diabetes Duration Interaction Effects on Gene Expression. *J.W. Kent[1], J.M Proffitt[1], E.I. Drigalenko[1], H.H.H. Goring[1], J. Charlesworth[2], F. Thameem[3], J.E. Curran[1], M.P. Johnson[1], T.D. Dyer[1], S.A. Cole[1], M.C. Mahaney[1], L. Almasy[1], E.K. Moses[4], R. Duggirala[1], H.E. Abboud[3], J. Blangero[1], S. Williams-Blangero[1]*. 1) Dept Genetics, Tx Biomed Res Inst, San Antonio, TX; 2) Menzies Res Inst Tasmania, Hobart, TAS, Australia; 3) Univ of Tx Health Sci Center San Antonio, Div Nephrology, San Antonio, TX; 4) Centre for Genetic Origins of Health & Disease, Univ Western Australia, Perth, WA, Australia.

While prolonged hyperglycemia can lead to nephropathy, not all individuals with diabetes progress to kidney damage at the same rate. The causes of this differential risk are largely unknown. We used genome-wide RNA expression profiles of peripheral blood mononuclear cells (PBMCs) collected at initial recruitment from 1240 Mexican American participants in the San Antonio Family Heart Study, to identify genes whose expression showed evidence of polygenic genotype x duration of diabetes interaction (GxDoDI). We used as covariates sex, age, duration of diabetes, and the first 5 principal components derived from correlated gene expression, which appear to reflect (among other sources of correlation) sample-specific differences in relative abundance of PBMC species. Principal components correction substantially improved detection of GxDoDI: we found 74 transcripts with GxDoDI significant at a 1% false discovery rate, compared to 22 without inclusion of the principal components covariates. These transcripts represented genes involved in mitochondiral ribosome function, response to viral infection, and T-cell differentiation. The transcript of the pseudouridinylate synthase-like 1 gene *PUSL1* (polygenic GxDoDI p=3.3e-08) also showed evidence of cis-SNP GxDoDI, based on a panel of ~1M common variants (rs12410087, p= 1.9e-06; rs12402622, p=9.6e-06; rs6662635, p=1.2e-05). These results identify a core set of candidate genes for differential transcriptional response to duration of diabetes that may serve as a starting point for identifying molecular pathways associated with the development of diabetic complications, including kidney disease.

1038T

MS Risk Conferred by Obesity may be Independent of Predisposing Genetic Factors for Obesity: Results from the Kaiser Permanente MS Research Program. *M. Gianfrancesco[1], X. Shao[1], B. Rhead[1], L. Shen[2], H. Quach[1], A. Bernstein[3], C. Schaefer[2,4], L.F. Barcellos[1,2].* 1) School of Public Health, Dept of Epidemiology, UC Berkeley, Berkeley, CA; 2) Kaiser Permanente Division of Research, Oakland, CA; 3) Palm Drive Hospital, Sebastopol, CA; 4) Research Program on Genes, Environment and Health, Kaiser Permanente, Oakland, CA.

Multiple sclerosis (MS) [MIM 126200] is a demyelinating autoimmune disease affecting over 2.5 million people worldwide, thought to involve both genetic and environmental risk factors. Recently, childhood and adolescent obesity has been associated with a two-fold increased risk of MS. However, the relationship between obesity and disease may be confounded by lifestyle and socioeconomic characteristics that were not appropriately adjusted for in previous studies. We utilized Mendelian randomization to estimate the unconfounded causal relationship between obesity and MS using a weighted obesity genetic risk score (oGRS). The oGRS incorporates the cumulative effect of 32 SNPs associated with obesity identified from 16 published GWA studies of obesity phenotypes in European-descent samples (Belsky, 2013): *FTO, NEGR1, TNNI3K, PTBP2, SEC16B, FANCL, LRP1B, TMEM18, RBJ, CADM2, ETV5, GNPDA2, SLC39A8, POC5, ZNF608, TFAP2B, LING02, LMX1B, RPL27A, BDNF, MTCH2, FAIM2, MTIF3, PRKD1, NRXN3, MAP2K5, GPR5B, SH2B1, MC4R, KCTD15, QPCTL,* and *TMEM160*. A split-sample instrumental variable (IV) analysis was used by summing all minor alleles times the estimated effect of each minor allele on the phenotype as reported in the literature. We also examined the extent to which an obesity genetic risk score computed using supervised machine learning (Super Learner, R package) was associated with obesity in our sample, and whether instrumental variable results using this score differed from those generated by the published oGRS. Participants included non-Hispanic Caucasian female members of Kaiser Permanente (985 MS cases, 10,000 controls). Common and rare variant data for each candidate gene were obtained through genome-wide association profiling and imputation. Preliminary results show oGRS to be weakly correlated with childhood body size, body size in 20's, body mass index (BMI) in 20's and BMI in 30's ($r<0.10$). IV estimates demonstrated no association between oGRS and MS (causal odds ratio (OR) = 1.09, 95% CI 0.92, 1.29); however, the OR increased when restricting to individuals diagnosed with MS at age 30 or older (OR = 1.17, 95% CI 0.97, 1.42). MS risk conferred by obesity may be independent of predisposing genetic factors for obesity, suggesting that alternative mechanisms may mediate disease onset. Obesity acquired due to environmental factors rather than genetics may have greater influence on MS development, or interact with other risk factors to increase risk of disease.

1039S

Congenic analysis reveals a gene-stress interaction affecting obesity in mice. *S.M. Clee, C.L.K. Leung, J. Dong, S. Karunakaran, L. Innala, M. Wang, V. Viau.* Cellular and Physiological Sciences, University of British Columbia, Vancouver, BC, Canada.

The BTBR T+ *Itpr3[tf]*/J (BTBR) inbred mouse strain is predisposed to obesity and pre-diabetes symptoms compared to the commonly studied C57BL/6J (B6) inbred strain. We have previously mapped loci affecting these traits and identified a locus, *Modifier of obese 1* (*Moo1*), on mouse chromosome 2 that accounts for the largest portion of the strain difference in body weight. Subsequently we localized this quantitative trait locus (QTL) to a congenic strain in which a ~6 Mb region of B6 genome had been transferred into the BTBR strain background. These congenic mice have reduced body weight, body fat, food intake and pre-diabetes compared to their pure BTBR littermates. Analysis of this congenic strain has revealed a significant gene-environment interaction that modifies the effect of this QTL. Here we show that under standardized phenotyping procedures the genotypic effect on body weight was markedly influenced by their housing environment (from a 0 to 17% difference in body weight between homozygous B6 and BTBR congenic mice across 3 facilities; $P_{geno} < 0.0001$, $P_{env} < 0.0001$, $P_{interaction}=0.006$). One major difference between facilities was ongoing construction nearby. This is known to cause stress in rodents and to affect food intake. In the facility with the most adjacent construction and where the genetic effect was completely overcome, the mice were never observed to be sleeping. In light of this, we examined the orexin (*Hcrt*) pathway. Orexin is a neuropeptide that promotes wakefulness and food intake and is known to be affected by stress. Orexin expression was 5 to 10-fold higher ($P < 0.001$) in mice housed in the facility with the most adjacent construction. We discovered a significant genotype-environment interaction in orexin receptor 1 (*Hcrtr1*) expression in the brain ($P=0.02$), consistent with the changes in body weight between genotypes and facilities and suggesting that genotype at this locus modifies the influence of stress on orexin signaling. When the effects of stress were directly tested in the congenic mice, we demonstrated a significant genotype-stress interaction ($P=0.02$) on body fat. Thus these studies have identified stress as a significant environmental modifier of obesity risk that is subject to genetic modulation. Discovery of the gene underlying the QTL will reveal a mechanism by which the orexin-mediated effects of stress on obesity can be modulated, which may prove useful for the design of novel strategies to aid weight loss.

1040M

Genetic Interactions in Developing of Rheumatoid Arthritis. *L.M. Diaz-Gallo, K. Shchetynsky, L. Padyukov.* Rheumatology unit, Department of Medicine, Karolinska Institute, Stockholm, Sweden.

Rheumatoid arthritis (RA) is a prototype of complex disease that involves multiple genetic and environmental factors. The individual effects of the known genetic risk factors are most often very moderate with highest impact for seropositive RA from the shared epitope (SE) alleles of the HLA-DRB1 gene. We previously have described significant interactions between the SE and some non-HLA genetic variants (i.e. PTPN22) in RA that indicated relatively strong effects in the susceptibility to seropositive RA. To provide more general view on gene-gene interactions in RA risk we tested the hypothesis regarding enrichment of the interactions between the SE alleles and previously shown genetic variants in association with RA. The study cohort includes 1151 seropositive RA patients and 1079 healthy controls from the Swedish EIRA study. Genome wide genotyping was performed using the IlluminaHumanHap300 BeadChip and subsequent imputation was done with the 1000 Genomes Phase I(α) Europeans as reference panel. All the individuals were genotype by low resolution SSP for the HLA-DRB1. Finally, 49 SNPs that previously were shown in association with RA and 48 SNPs selected from the study of genetic basis for reading and writing ability were analysed for interaction with SE alleles. SNPs from the HLA and PTPN22 loci and SNPs with MAF<0.01 were excluded from the analysis. The GEIRA algorithm provides measures of interaction including the attributable proportion (AP) due to interaction. We compared the distribution of the AP p-values observed in two groups of SNPs, using the Kolmogorov-Smirnov test. Overall 92 tests for interaction were valid in our study due to statistical constrains and we observed an enrichment of p-values <0.05 for AP in group of RA associated SNPs in comparison to control group of SNPs (10 vs. 3). Distribution analysis of p-values reveals significant difference between two groups ($p<0.05$). Additionally we noticed that only 17 out of 49 tested SNPs for RA exhibited a significant association in our study, but 7 of these SNPs (41%) were also in interaction with SE alleles. Our data indicate that the SE risk effect in RA could be in part attributed to its interaction with other non-HLA genetic risk factors to seropositive RA. The interaction between SE and others genetic variants can explain a significant fraction of the genetic background of RA.

1041T

Smoking-dependent genetic effects on obesity traits: the GIANT (Genetic Investigation of Anthropomorphic Traits) Consortium. *V.A. Fisher[1], A.E. Justice[2], T.W. Winkler[3], N.L. Heard-Costa[4,5], K.L. Young[2], J. Czajkowski[6], M. Graff[2], X. Deng[1], T.S. Ahluwalia[7], Q. Qi[8], L. Qi[9,10], A. Mahajan[11], D. Hadley[12], A.Y. Chu[13], L. Barata[6], J.S. Ngwa[1], R.A. Scott[14], L. Xue[1], T.O. Kilpeläinen[7], C-T. Liu[1], D.I. Chasman[13], K. Mohlke[15], R.J.F. Loos[16], I.B. Borecki[6], K.E. North[2], L.A. Cupples[1,4], GIANT Consortium.* 1) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 2) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany; 4) NHLBI Framingham Heart Study, Framingham, MA; 5) Boston University School of Medicine, Boston MA, USA; 6) Department of Genetics, Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO; 7) The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 8) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY; 9) Department of Nutrition, Harvard School of Public Health, Boston, MA; 10) Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 11) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 12) Division of Population Health Sciences and Education, St. George's, University of London, London, UK; 13) Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 14) MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, UK; 15) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 16) The Genetics of Obesity and Related Metabolic Traits Program, The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

Obesity and cigarette smoking (SMK) are important risk factors for cardiovascular disease. They are related: average body weight is lower among smokers and smoking cessation leads to weight gain, but central obesity is often higher in smokers. To investigate obesity as body size and distribution of body fat, we studied body mass index (BMI), waist circumference adjusted for BMI (WC), and waist/hip ratio adjusted for BMI (WHR). This builds on previous work of the GIANT consortium that identified 63 loci associated with WC in 42 cohort studies. Since then 46 additional studies have joined the consortium, nearly doubling the total sample size to 210,153 adults of European ancestry with genome-wide or Metabochip data. We performed genome-wide meta-analysis for each trait, considering association models stratified by SMK (current smoker: yes/no) and also adjusted for SMK. This analysis allows genetic effects to differ between smokers and non-smokers, equivalent to gene-environment interaction with SMK. A joint 2df test of the SMK-adjusted genetic main effect and gene-SMK interaction summarizes the influence of each SNP on trait outcomes, and stratified results allow us to distinguish between loci reaching genome-wide significance (GWS) because of strong effect in a specific population, and those with more consistent moderate effect. With a GWS threshold of 5E-8 and discarded weaker signals within 500kB of each selected SNP to avoid spurious associations due to LD, we identify 81 significant loci for WC of which 14 are novel in the extended sample, 68 loci (10 novel) for BMI, and 50 (9 novel) for WHR. Among the novel GWS main effect loci, WC associated with variants on chromosome 16 near the *AGRP* gene known for involvement in metabolism and eating disorders, and WHR associated with several intragenic SNPs between *NOTCH4* and the HLA D-antigen region on chromosome 6. Two variants near *ADAMTS7* showed GWS interaction with SMK, with lower BMI in minor allele carriers only among smokers. In addition, four loci previously associated with tobacco use disorder were also associated with WC and WHR (*COBBL1, DNM3, RSPO3, TBX15/WARS2*). We find several new loci associated with obesity by accounting for the influence of SMK, and continue to investigate biological function of these variants. Our results highlight the importance of appropriately modeling genetic associations by considering known relationships between phenotypes and environment.

1042S

Tomato consumption, an anecdotal trigger of gout flares, interacts with three urate transporters (*ABCG2*, *SLC22A12* and *SLC22A7*) in a non-additive fashion to influence serum urate. *T. Flynn[1], M. Cadzow[1], N. Dalbeth[2], L. Stamp[3], R. Topless[1], T. Merriman[1].* 1) Department of Biochemistry, University of Otago, Dunedin, Otago, New Zealand; 2) Department of Medicine, University of Auckland, Auckland, New Zealand; 3) Department of Medicine, University of Otago, Christchurch, Canterbury, New Zealand.

Raised urate levels, the main cause of gout, arise through a complex combination of genetic and environmental factors. Thirty genetic loci and several dietary exposures have well-established effects on urate levels. Interactions between these genetic and environmental factors have been reported. Urate is ideal for investigating gene-environment interactions as causal environmental exposures can be measured concurrently with phenotype. This study aimed to test two hypotheses: that tomato consumption, an anecdotal trigger of gout flares, is associated with urate levels and whether any of 7 genes encoding uric acid transporters interact with tomato consumption. 12,720 European individuals from the Atherosclerosis Risk In Communities (n=7517), Cardiovascular Health Study (n=2151) and Framingham Heart Study (n=3052) cohorts were used. All analyses were adjusted for confounding variables. There was association between tomato intake and urate levels (β=0.66 $\mu molL^{-1}$; P=0.01), which was modified by *ABCG2*: *rs2231142*, *SLC22A12*: *rs3825018* and *SLC22A7*: *rs4149178* genotypes (Pinteraction=0.04, 0.02 (in men) and 0.049, respectively). In the presence of the urate-increasing genotypes of *rs2231142* (TT) and *rs3825018* (GG) consumption of tomatoes showed a urate-lowering effect (β=-4.81 $\mu molL^{-1}$, P=0.01; β=-3.34 $\mu molL^{-1}$, P=0.02 (in men), respectively) converse to the urate-increasing effect shown when no genotypic stratification was performed. In the presence of the *rs4149178* urate-increasing genotype (GG) consumption of tomatoes showed a significantly higher increase in urate levels than in the un-stratified analysis (β=2.92 $\mu molL^{-1}$, P= 0.03). Subject to replication these data provide insight into the mechanisms involved in urate control and may lead toward personalised dietary advice in strategies to reduce urate concentrations.

1043M

Preliminary evidence of an interaction between a polymorphism in the *BDKR2B* gene and the dietary potassium intake on the systolic blood pressure in a sample of healthy adults. *J. Giovanella, J.P. Genro, S.M. Dal Bosco, V. Contini.* Postgraduate Program in Biotechnology, Centro Universitário UNIVATES, Lajeado, RS, Brazil.

Background: Studies have suggested significant associations between variants in the bradykinin receptor B2 gene (*BDKR2B*) and the development of hypertension. The bradykinin is one of the most important kinins involved in the modulation of the blood pressure (BP) and many of its physiological functions are mediated by the B2 receptors. In addition, high sodium and potassium intake has also been related to changes in the BP. **Objective:** The objective of this study was to investigate the interaction between the polymorphism rs1799722 of the *BDKR2B* gene and the environmental factors - sodium and potassium intake - on the variation in the BP in a sample of healthy adults from a private university in the South of Brazil. **Methods:** The participants were investigated for food consumption and BP by professional in the Nutrition Ambulatory from the university. The measurements of sodium and potassium intake were assessed by the 24-hour recall method (24-H-RQ), using the DietWin Professional software. BP values were determined by the mean of three measurements with the digital apparatus Omron. The polymorphism was genotyped using the TaqMan SNP genotyping assay. The allele frequencies were estimated by direct counting and the Hardy-Weinberg Equilibrium (HWE) calculated by the chi-square test. The gene-nutrient interactions were tested using multiple linear regressions with manual backward stepwise modeling. **Results:** The sample was composed by 267 women and 71 men. The average age of the sample was 25.5 years. Compared to women, men showed significantly higher values in the systolic BP (SBP) (men, 124.4 ± 11.0 mmHg; women, 113.7± 10.2 mmHg) and in the sodium (men, 2403.8 ± 1725.6 mg 24-H-RQ; women 1756.2 ± 1017.1 mg 24-H-RQ) and potassium intake (men, 2.343.7 ± 1089.9 mg 24-H-RQ; women 1758.5 ± 684.9 mg 24-H-RQ) (all p<0.001). The allele frequencies for the polymorphism were 0.59 (C) and 0.41 (T) and the genotype frequencies did not reveal a significant deviation from the expected values for the HWE. Our results did not show any main effect of the polymorphism in the investigated outcomes. However, we detected an interaction between the TT-genotype and the dietary potassium intake on the SBP (p<0,001). The increased consumption of potassium, in carriers of the TT-genotype, was associated with an increasing on the SBP. **Conclusions:** This study provides a preliminary indication of a gene-nutrition interaction on the BP in a sample of healthy adults.

1044T

Genome-wide gene-physical activity interaction study of BMI and waist-hip ratio in 180,418 individuals. *T.O. Kilpeläinen[1], M. Graff[2], R.A. Scott[3], A. Justice[2], T.W. Winkler[4], L. Xue[5], A. Mahajan[6], D. Hadley[7], T.S. Ahluwalia[1], F. Renström[8], T. Workalemahu[9], M. den Hoed[10], A.Y. Chu[11], M.L. Grove[12], N. Heard-Costa[13], L. Quaye[14], J. Czajkowski[15], J. Ngwa[5], K. Young[2], Q. Qi[16], O. Pedersen[1], K. Monda[2], K.E. North[2], I.B. Borecki[17], R.J.F. Loos[18], L.A. Cupples[5], The GIANT Consortium.* 1) Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Denmark; 2) Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, NC; 3) MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, UK; 4) Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Germany; 5) Department of Biostatistics, Boston University School of Public Health, MA; 6) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 7) Division of Population Health Sciences and Education, St. George's, University of London, UK; 8) Department of Clinical Science, Genetic and Molecular Epidemiology Unit, Lund University, Sweden; 9) Department of Epidemiology, Harvard School of Public Health, MA; 10) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Sweden; 11) Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA; 12) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, TX; 13) Department of Neurology, Boston University School of Medicine, MA; 14) Department of Twin Research and Genetic Epidemiology, King's College London, UK; 15) Division of Statistical Genomics, Center for Genome Sciences & Systems Biology, Washington University School of Medicine, St. Louis, MO; 16) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY; 17) Department of Genetics, Washington University School of Medicine, MO; 18) The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, NY.

The global obesity epidemic underscores the importance of gaining a deeper understanding of the biology of weight regulation to develop better preventive strategies. Obesity has a strong genetic component, but lifestyle factors, such as physical activity (PA), may modify the impact of genetic susceptibility. Recently, a candidate-gene-based meta-analysis confirmed that the body mass index (BMI)-increasing effect of *FTO* is attenuated by PA. To identify novel loci that interact with PA on BMI or waist-hip ratio adjusted for BMI (WHRadjBMI), we performed a meta-analysis of 64 studies with genome-wide or metabochip data, including up to 180,418 individuals of European descent. We standardized PA by categorizing it into a dichotomous variable (inactive vs. active) in each study. Overall, 23% of individuals were categorized as inactive. We pooled the results from individual studies using fixed effects meta-analysis of the SNP main effect in inactive and active strata, separately, and screened for SNPxPA interactions by testing for difference between the strata. To identify loci showing both main and interaction effects, we performed a joint meta-analysis of the SNP main effect and SNPxPA interaction. In the genome-wide meta-analysis of SNPxPA interaction alone, we identified a novel locus near *CDH12* ($P=3\times10^{-8}$) which showed a BMI-increasing effect in the inactive group, but a BMI-decreasing effect in the active group. In the joint meta-analysis, 69 loci reached $P<5\times10^{-8}$ (43 loci for BMI, 26 for WHRadjBMI), four of which have not been previously identified (near *FLJ30838* or *MRAS* for BMI; near *HHAT* or *PMS2P5* for WHRadjBMI). For each of the 69 loci, the joint test association was driven by the SNP's main effect, rather than by its interaction with PA. When we examined the SNPxPA interaction effects of the 69 loci, separately from main effects, the strongest interaction occurred at the *FTO* locus ($P=5\times10^{-4}$); the BMI-increasing effect was 29% smaller in the active than in the inactive group. In this sample of 180,418 individuals, we identify a novel locus near *CDH12* interacting with PA on BMI and replicate the PA-interaction of *FTO*. Furthermore, while accounting for PA in the model, we identify four novel loci associated with BMI or WHRadjBMI. We are currently performing a detailed follow-up of the loci to elucidate their potential functional roles. Our study may yield important novel insights into the role of gene-lifestyle interactions in the etiology of obesity.

1045S

Association of physical activity with lower type-2 diabetes incidence is weaker in those with high genetic risk. *Y.C. Klimentidis, Z. Chen, A. Arora, C. Hsu.* Epidemiology and Biostatistics, University of Arizona, Tucson, AZ.

We examined whether the association of physical activity (PA) with type-2 diabetes incidence differs according to several types of genetic susceptibility. In a large prospective cohort with 838 incident cases, we examined interactions of physical activity with 1) each of 65 type-2 diabetes-associated single nucleotide polymorphisms (SNPs), 2) a genetic risk score (GRS) comprised of all 65 SNPs, 3) two GRSs comprised of SNPs implicated in insulin resistance and insulin secretion, and 4) GRSs for fasting insulin and glucose. We find a significant interaction of physical activity and the type-2 diabetes GRS (p=0.019), suggesting a weaker protective effect of physical activity in those at high genetic risk. Based on the interactions observed with the insulin resistance GRS (p=0.044) and the fasting insulin GRS (p=0.060), it appears that this overall type-2 diabetes GRS interaction most likely occurs through genetic susceptibility to insulin resistance, as opposed to insulin secretion. Furthermore, this interaction was more pronounced in women (p=0.0046) than in men (p=0.47). No single SNP stood out as displaying a strong interaction with physical activity. We conclude that although physical activity appears to have an overall protective effect on type-2 diabetes, this putative effect is weakest in women with high genetic risk for type-2 diabetes and insulin resistance.

1046M

Genome-wide scan for context-dependent marker SNP effects in coronary heart disease. *S.M. Raj[1], C.F. Sing[2], G. Dyson[3], A.G. Clark[1].* 1) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA; 3) Department of Oncology, Wayne State University, Detroit, MI, USA.

The existence of epistatic and genotype-environment interaction effects may limit the utility of genome-wide association studies (GWAS) to identify loci that are involved in determining the genetic architecture of a complex disease. We hypothesize that loci not detected by a GWAS may have allelic effects in the context of subgroups of the population who have the highest risk to disease. We have carried out the first genome-wide scan to detect such context-dependent locus effects.

We used a modified Patient Rule-Induction Method (PRIM) to assess which combinations of SNP marker and traditional risk factors identified subgroups with the highest cumulative incidence of coronary heart disease (CHD). The PRIM was applied to 840,000 autosomal SNPs and eight risk factors collected on 7,589 European-American adults from the Atherosclerosis Risk in Communities study. We also carried out a GWAS on this dataset to evaluate similarities and differences in the genetic signals captured by the PRIM.

The PRIM-based scan for context-dependent effects of marker SNPs identified approximately three times more hits than the GWAS, with females having a proportionately higher number of PRIM associations. These two approaches yielded very little overlap in associated genetic variants. We found that SNPs showing the strongest associations with CHD in a particular sub-group context defined by a traditional risk factor did not differ significantly between females and males, but did vary among contexts, with hypertension being the most common measure of context. These results suggest that consideration of context-based SNP associations with a complex disease may uncover new genetic pathways that contribute to disease pathogenesis. The role of context in determining the utility of genomic information has powerful implications for the clinical and population based management of the risk of a common disease having a complex multifactorial etiology.

1047T

Utility of the rhesus macaque (*Macaca mulatta*) as a novel genetic model for spontaneous human inflammatory bowel disease (IBD): sexual dimorphism and gene-by-sex effects on chronic diarrhea. *A. Vinson[1,3], M.J. Raboin[1], K.D. Prongay[2], E.R. Sunderhaus[3].* 1) Div. of Neuroscience, Oregon National Primate Research Center, Beaverton, OR; 2) Div. of Comparative Medicine, Oregon National Primate Research Center, Beaverton, OR; 3) Dept. of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR.

Inflammatory bowel disease (IBD) is one of the most prevalent gastrointestinal diseases in the U.S., and is known to have a strong genetic component. Moreover, recent studies support the role of female bias in genetic risk for IBD. However, no animal model of spontaneous IBD currently exists that will support genetic investigation of this disease, including sex-specific genetic effects. Based on our observation that the rhesus macaque displays chronic diarrhea at a high prevalence in captive colonies, our goal was to explore the utility of this species as a genetic model of spontaneous human IBD. Our aims were to assess the overall and sex-specific heritability of chronic diarrhea in a captive, pedigreed population of Indian-origin rhesus macaques. After clinical ascertainment of 233 animals affected with chronic diarrhea, we developed an extended pedigree of 5,328 relatives around these probands, using pedigree information available on ~23,000 macaques housed at the ONPRC since 1963. Using existing clinical veterinary records, we developed computational approaches to identify episodes of chronic diarrhea in all pedigree members. We then scored animals based on the number of days of diarrhea, per the total number of days present in the colony. After accounting for differences in housing, medication, and covariates of age and sex, we found low, but statistically significant heritability for this phenotype when both sexes were combined (h^2=0.12, P=1.61 × 10^{-8}). However, stratifying the pedigree by sex revealed substantial heritability among females (h^2=0.46, P=4.72 × 10^{-30}, N=1,676), and much lower, although statistically significant, heritability among males (h^2=0.16, P=5.40 × 10^{-4}, N=997). Consistent with the phenotypic sexual dimorphism suggested by covariate effects of sex (P-value=1.55 × 10^{-7}), initial tests of gene-by-sex interaction confirmed the significant effects of sex on genetic variance influencing chronic diarrhea in this pedigree (ρ_G = 0.02; test for ρ_G = 1.0 between males and females, P=5.37 × 10^{-4}). We conclude that genetic effects on chronic diarrhea in the rhesus macaque appear to play a larger role in females than in males, similar to observations in humans. These findings provide initial support for the utility of the rhesus macaque as a novel genetic model for sex-biased effects on spontaneous human IBD.

1048S

Genome wide gene-vitamin D interaction analysis suggests potential role for melanoma related genes in Parkinson disease. *L. Wang[1,2], L. Maldonado[1], J. Ritchie[3], M. Evatt[4], J.L. Haines[5], M. Pericak-Vance[1,2], G.W. Beecham[1,2], E.R. Martin[1,2], J.M. Vance[1,2], W.K. Scott[1,2].* 1) Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine; 3) Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA; 4) Department of Neurology, Emory University School of Medicine, Atlanta, GA; 5) Department of Epidemiology and Biostatistics, Case Western Reserve University, School of Medicine, Cleveland, OH.

Despite extensive genome-wide association studies (GWAS) in Parkinson disease (PD), much of the estimated genetic risk has yet to be detected. One source of the "missing heritability" may be unaccounted-for gene-environment interactions. Vitamin D (Vit D) deficiency has been associated with increased PD risk in multiple studies. As Vit D regulates expression of a wide array of genes, we sought to identify genes involved in modifying the association between Vit D and PD by analyzing genetic interactions with vit D deficiency. We imputed up to 7.2 M SNPs in 477 PD cases and 430 controls, whose Vit D metabolites were measured in stored plasma samples using mass spectrometry. Vit D deficiency was defined as having plasma 25(OH)D concentration <20 ng/ml and was strongly associated with PD (Odds Ratio (OR)=2.7, P<0.0001). Joint tests of gene-environment interaction were conducted by comparing a full model containing SNP dosage, vit D deficiency, an interaction term, and covariates age, sex and sampling season to a restricted model with only vit D deficiency and covariates. While no joint tests produced genome-wide significant results at P< 5 x 10-8, the most significant interaction terms were detected with SNPs in two genes associated with melanin production and risk of melanoma. This is intriguing given that PD is characterized by a loss of melanin-positive, dopaminergic neurons in the substantia nigra and an increased incidence of melanoma in PD patients has been reported in several studies. The strongest evidence for interaction was found at rs7312710 in FBRSL1 (P=2×10 6). The effect of this SNP on PD depended on vit D status: the minor allele is associated with increased risk in vit D deficient individuals (OR=2.3, P=0.0004) and with decreased risk in vit D non-deficient individuals (OR=0.7, P=0.0036). The second strongest interaction was found in C10orf11 with a similar pattern. We hypothesize that PD and melanoma share biological pathways (e.g. melanin production) that when perturbed modify risk to both diseases. This is supported by our previous pathway analysis using GWAS and microarray data suggesting that variation in the melanogenesis pathway contributes to PD risk. Our study demonstrates vit D-gene interactions and nominates new genes that were not implicated in previous GWAS in PD. Our findings illustrate that examining gene-environment interactions can identify novel genes underlying the "missing heritability" in PD.

1049M

Smoking Then and Now: What Can the Aggregate of Genome-wide SNPs Tell Us About the Correspondence of Genetic Factors Influencing Cigarette Smoking Initiation Between Birth Cohorts. *A.G. Wills[1,2], G. Carey[1,2], M.C. Keller[1,2].* 1) Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, USA; 2) Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA.

Twin studies have drawn attention to the importance of birth cohort when examining the genetic etiology of cigarette smoking behavior. For the smoking phenotype of initiation, we used whole-genome SNP data from a combined sample of 9189 European Americans that had participated in the Atherosclerosis Risk in Communities Study (ARIC) and the Multi-Ethnic Study of Atherosclerosis (MESA). The majority of these individuals were born between 1920 and 1950. Using Genome-wide Complex Trait Analysis[1], we estimated the SNP heritability for individuals born before (SNP h^2 = .16) and after (SNP h^2 = .23) the combined sample's median birth year of 1934. In the bivariate analysis, the correlation between the set of SNPs that explained variance in those born before 1934 and those born after was moderate (r_g = .40) and significantly different from 1 (p = .048). Further, in a model that allowed for a gene x cohort interaction parameter, we found a marginally significant interaction (.11, p = .06) between birth year and the conglomerate of SNP effects. To better understand the differing genetic contribution based on year of birth, we divided our sample into birth decades, and the largest genetic effect was found for those born in the late 1930s. Due to disparity in sample sizes and the potential for unexamined gender effects, we remain cautious in interpreting this result. However, these findings highlight the potential for etiological heterogeneity among smokers of different eras.

1.Yang J, Lee SH, Goddard ME and Visscher PM. GCTA: a tool for Genome-wide Complex Trait Analysis. Am J Hum Genet. 2011 Jan 88(1): 76-82.

1050T

Risk prediction and Type II Diabetes. *N. Furlotte, S. Dandekar, R. Smith, N. Eriksson, D. Hinds.* 23andMe, Mountain View, CA.

As the prevalence of Type II Diabetes (T2D) continues to increase worldwide, the ability to accurately assess the risk for developing this disease is becoming increasingly important in clinical practice. Accounts vary as to the clinical utility of genetic risk prediction models for T2D. Estimates of heritability vary widely, but it is generally accepted that environmental factors such as food consumption, exercise frequency and body mass index play a much larger role in the development of this disease than the genetic variations implicated through genome-wide association studies. Furthermore, the effects of genetic variations may only be apparent under particular environmental conditions - so called gene-by-environment interactions. We explore these issues in a cohort of over 300,000 23andMe customers. We evaluate the predictive power of the most recent and robust genetic variations implicated in the etiology of T2D through association analysis and compare their predictive ability with environmental predictors related to individual behavior. In addition, we search for genetic variations exhibiting environmentally specific genetic effects and quantify the proportion of the total trait variation attributed to these gene-by-environment interactions.

1051S

Investigate cytokine levels of cord blood samples in related to maternal allergic status. *H. Tsai[1,2,3], Y. Huang[2], Y. Tsai[2], Y. Ho[4], W. Hsiao[4], J. Wang[4,5].* 1) Dept Pediatrics, Northwestern Univ Sch Med, Chicago, IL; 2) Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Taiwan; 3) Department of Medical Genetics, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan; 4) Department of Pediatrics, National Cheng Kung University Hospital, Tainan, Taiwan; 5) Institute of Basic Medical Sciences, College of Medicine, National Cheng Kung University Hospital, Tainan, Taiwan.

Introduction: The increasing prevalence of early childhood allergic traits such as eczema, wheezing, asthma and allergic rhinitis has been a substantial public health burden in Taiwan and worldwide. Previous studies have reported certain kinds of cytokines and chemokines are involved in the pathophysiology of asthma and/or atopy. Based on our previous work, we hypothesize that maternal allergic status during pregnancy may play an important role influencing cytokine levels of their fetus, that is, may have an effect of immune dysregulation in-utero. Methods: A total of 55 maternal participants was recruited in this study. In addition to maternal biospecimen, we also collected their cord blood samples. We stimulated cord blood mononuclear cells (CBMCs) with lipopolysaccharide (LPS) and specific allergen from Der. Pteronyssinus (Derp), separately, and measured cytokine levels, specifically, IL-6, IL-8, IL-10, IL-23 and Tumor necrosis factor alpha (TNF-alpha), from the stimulated cord blood samples. We then applied Student's t test and Wilcoxon rank sum test, separately, to examine whether examined cytokine levels differed related to maternal allergic status. Results: We examined 5 different cytokine levels in medium (not stimulated), LPS and Derp stimulated CBMCs in related to maternal allergic status. We only observed that there was borderline significance in TNF-alpha between maternal subjects with and without allergy (p=0.06). Beyond that, the results indicated that there was no difference in the examined cytokine levels between maternal subjects with and without allergy, no matter medium, LPS or Derp stimulated cord blood samples. Conclusions: The present study indicated cytokine levels of IL-6, IL-8, IL-10 and IL-23 in cord blood were not different in related to maternal allergic status. Further investigation will be warranted to better understand maternal influence on TNF-alpha levels in -utero.

1052M

Estrogen-dependent upregulation of *IRF5* in human immune cells. *S.E. Lofgren[1], J. de Azevedo Silva[2], M.E. Alarcon-Riquelme[3], S.V. Kozyrev[4].* 1) Department of Cell Biology, Embriology and Genetics, Federal University of Santa Catarina, Florianopolis, Brazil; 2) Department of Genetics, Federal University of Pernambuco, Brazil; 3) GENYO; Centre for Genomics and Oncological Research; Pfizer; University of Granada; Andalusian Regional Government; Granada, Spain and Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, USA; 4) Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden.

Background: Systemic lupus erythematosus (SLE) as well as other autoimmune diseases show a strong gender bias, where the great majority of patients are women. Although controversial, the effect of estrogen has been implicated with SLE in several ways. However, the mechanism on how estrogen contributes to the development of SLE is still largely unknown. Objective: To contribute to the knowledge of the mechanisms behind the strong sex-biased prevalence of SLE, we investigate the role of estrogen on the expression of one of the strongest associated gene with SLE, the interferon regulatory factor 5 (*IRF5*), in human immune cells. Material and methods: *IRF5*, as well as *IRF3*, *IRF4*, *IRF7* and *IRF9* expression was measured in PBMCs, LCLs, monocytes and monocyte-derived macrophages from both male and female origin. Cells were treated with different concentrations of estrogen and gene expression was measured by quantitative PCR. Results: We found that the initial levels of *IRF5* in PBMCs were almost 2-fold higher in women that in men, although not reaching statistical significance. After 12 h in culture the *IRF5* levels became roughly equal in both sexes, and further stimulation with estrogen lead to up-regulation of *IRF5* levels in both PBMCs and monocytes in both women and men. No difference was seen for *IRF3*, *IRF4*, *IRF7* and *IRF9* expression, and none of the genes analyzed was up-regulated in LCLs, upon estrogen treatment, regardless of the gender. The later could be explained by the fact that we could not detect any expression of estrogen receptor *ESR1* on those cell lines. Conclusions: We could show that in human PBMCs and monocytes from healthy individuals *IRF5* expression can be regulated by exogenous estrogen. The effect in PBMCs was seen in both gender but was more pronounced in women. This feature might be specific to *IRF5* since four other *IRF* genes tested did not show any up-regulation in these cells. This data could to some extent explain the sex bias in the development of SLE as well as other autoimmune diseases.

1053T

Genetic and early life environmental influences on body mass index. *A. Smith[1], K. Jensen[1], C. Sartor[2], J. Kaufman[2], H. Kranzler[3,4], J. Gelernter[1,5].* 1) Div Human Genetics, Dept Psychiatry, Yale School Medicine, New Haven, CT; 2) Dept Psychiatry, Yale School of Medicine, New Haven, CT; 3) Dept Psychiatry, University of Pennsylvania, Philadelphia, PA; 4) Philadelphia VAMC; 5) West Haven VAMC.

Introduction: The public health consequences of overweight and obesity include risk for a range of metabolic and cardiovascular disorders. Early life adversity has been shown to exert lasting effects on body mass index (BMI) and related health outcomes throughout adulthood. Improved understanding of the pathogenesis of obesity may therefore require examination of the interplay between genetic and early life environmental risk factors. Methods: We recruited a heavily substance-dependent (SD) adult sample of >9,000 European- and African-American (EA and AA) subjects with BMI data and careful diagnosis of SD and other psychiatric traits via the semi-structured assessment of drug dependence and alcoholism (SSADDA). Severe childhood adversity (CA) (e.g., child abuse, violent crime, or death of a parent in early childhood) was also assessed; 30.2% of men and 31.2% of women reported at least one such event. We conducted in this sample a study of how both genetic variation and early life stress influence BMI in a single group of subjects. Results: In our combined sample (AA and EA), after adjusting for SD diagnoses, age, race/ ethnicity, and household income in a linear regression model, we found a consistent relationship between a CA sum score and log transformed BMI (Pmales = 1.9 x 10-4, beta = 0.01, n = 5,095; Pfemales = 2.0×10-10, beta = 0.03, n = 4,175). Large consortia have previously carried out GWASs of BMI that were restricted to European-ancestry populations. In the EA sub-sample of our dataset, we generated genetic risk scores (GRSs) for each individual using effect size estimates for 32 loci with known genome-wide significant BMI associations. GRSs fell on a continuous scale from 3.3 to 10.0. After incorporation of GRSs into the regression model, in both male and female EAs significant associations were observed between CA and BMI (Pmales = 1.5 x 10-2, beta = 0.02, n = 1,482; Pfemales = 6.8×10-5, beta = 0.04, n = 931), and between GRS and BMI (Pmales = 2.3 x 10-10, beta = 0.03; Pfemales = 3.9×10-7, beta = 0.04). In males, but not females, we observed a significant interaction between GRS and CA (P = 7.6 x 10-3), which reflected an increase in the influence of GRS on adult BMI as the CA sum score increased. Conclusion: These sex-specific findings represent an initial step in understanding how genetic variation and the developmental environment individually and jointly may exert lasting effects on metabolism and health.

1054S

A new GATK framework for RNA-seq variant discovery identifies differential A-to-I RNA editing in autistic brains. *A. Eran[1,2], A. Levy-Moonshine[2], R. Anner[3], N. Milo[3], E. Bachmat[3], R. Zhang[4], J.B. LI[4], D.M. Margulies[1], L.M. Kunkel[5], I.S. Kohane[1].* 1) Center for Biomedical Informatics, Harvard Medical School, Boston, MA; 2) Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Department of Computer Science, Ben Gurion University, Beer Sheva, Israel; 4) Department of Genetics, Stanford University, Stanford, CA; 5) Department of Genetics, Boston Children's Hospital and Harvard Medical School, Boston, MA.

Adenosine-to-inosine (A-to-I) RNA editing is an epigenetic mechanism that fine-tunes synaptic function in response to environmental stimuli. Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental disorder commonly associated with synaptic dysfunction. Targeted RNA sequencing studies have identified consistent A-to-I editing alterations of candidate genes in ASD, suggesting that A-to-I RNA editing may play a role in the disorder. However, transcriptome-wide editing profiles of individuals with ASD are still unknown, and the overall contribution of A-to-I editing alterations to the ASD phenotype remains to be discovered. Here we present a comprehensive framework for accurate SNP and indel detection in RNA-seq data, and apply it to conduct a first transcriptome-scale survey of A-to-I RNA editing in postmortem prefrontal cortex and cerebellum from individuals with ASD. The Genome Analysis Toolkit's RNA-seq variant analysis framework is based on local de-novo assembly of RNA haplotypes. It begins with RNA-seq data pre-processing, including spliced alignments, duplicate marking, splicing and indel realignment, and the recalibration of base and mapping scores. It continues with de-Bruijn assemblies of variable regions, their likelihood calculations, and identification of variants on the most likely haplotypes. It ends with variant filtering and the production of a SNP and indel call set in a VCF file. We evaluated the performance of this approach using whole exome sequences from the same tissues, and applied it for a first genome-wide comparative survey of A-to-I RNA editing in autistic brains. We compared cerebellar to prefrontal cortex editing levels, in synaptic vs. non-synaptic transcripts, across isoforms, genes, and gene families. We identified differential editing of bona fide autism genes and pathways, and characterized their relations to expression and splicing patterns. We then validated these findings using an independent deep targeted sequencing approach, mmPCR-seq. These results improve our understanding of the role of A-to-I editing in ASD and demonstrate the effectiveness of the GATK's RNA-seq variant analysis framework.

1055M

Risk Prediction for Age-Related Macular Degeneration Using Genetic and Environmental Factors. *W. Wu[1,2], F. Wu[1,2], L. Zhao[1,2], H. Luo[1,2], C. Wen[1,2], N. Sfeir[1,2], S. Patel[1,2], M. Goldbaum[1], H. Ferreyra[1], K. Zhang[1,2].* 1) Department of Ophthalmology and Shiley Eye Center, University of California, San Diego, La Jolla, CA; 2) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA.

Purpose: To determine the individual and composite effects of genetic factors and smoking on the development of age-related macular degeneration (AMD), including the early form of AMD, which is characterized by the presence of confluent drusen in the retina, as well as the advanced forms of AMD, which include geographic atrophy (GA) and choroidal neovascularization (CNV). Methods: DNA samples from 522 patients without AMD, 239 patients with confluent drusen, 271 patients with GA, and 1159 patients with CNV were analyzed for four genetic variants known to be associated with advanced AMD. Patients were also asked about their smoking status. Logistic regression was performed to generate a predictive model for AMD risk, and odds ratios were calculated based on individual and composite factors. Results: Odds ratios (ORs) for CNV were 5.26 (95% CI: 3.51-8.13) for HTRA1/ARMS2 rs10490924 TT, 2.15 (95% CI: 1.41-3.28) for CFH rs1061170 CC, 3.71 (95% CI: 2.31-6.00) for CFH rs2274700 CC, and 1.33 (95% CI: 1.05-1.68) for C3 rs2330199 CG when compared to individual reference genotypes. Smoking status alone was associated with an OR for CNV of 1.55 (95% CI: 1.24-1.95). ORs for GA were 3.85 (95% CI: 2.21-6.79) for HTRA1/ARMS2 rs10490924 TT, 1.89 (95% CI: 1.07-3.38) for CFH rs1061170 CC, 6.83 (95% CI: 3.10-16.21) for CFH rs2274700 CC, and 2.97 (95% CI: 1.61-5.52) for C3 rs2330199 GG when compared to individual reference genotypes. Smoking status alone was not associated with an increased risk of GA. For smokers carrying a minimum of six out of eight possible risk alleles, the OR for CNV was 15.83 (95% CI: 6.06-45.19), and the OR for GA was 15.39 (95% CI: 3.47-110.84). For smokers carrying five risk alleles, the OR for drusen was 6.97 (95% CI: 2.13-27.86). The composite effect of all four genotypes combined with smoking was significant for CNV ($p < 2*10^{-16}$), GA ($p < 2*10^{-16}$) and drusen ($p = 8.1*10^{-12}$). Conclusions: Combined genetic and environmental factors influence the risk of AMD significantly. CFH rs2274700 and C3 rs2330199 conferred greater risk for GA than for CNV. Because the onset of AMD can be insidious, it is important to develop predictive models to target high-risk individuals. Monitoring and early interventions may improve clinical outcomes in these patients.

1056T

Study of genetic polymorphism and oxidative stress markers analyses in psoriasis patients from Indian. *S. Chettiar[1], S. Desai[1], D. Gimhre[1], D. Umrigar[6], J. Dattani[4], A. Patel[3], D. Jhala[2], R. Uppala[5].* 1) Department of Biotechnology, SRICEAS, surat, Gujarat, India; 2) Department of Zoology, Gujarat University, Ahmedabad, India; 3) Genexplore Diagnostic and Research Centre, Ahmedabad, India; 4) Regional Office for Health and Family Welfare, GOI, Ahmedabad, India; 5) Green cross Pathology and Molecular Biology laboratory, Paldi, Ahmedabad, India,; 6) Department of Dermatology, New Civil Hospital, Surat.

Psoriasis (PSORS) is a common, immune-mediated, genetically determined long-lasting (chronic) skin disease that usually presents after puberty. It is characterized by red, scaly patches, papules, plaques usually inflamed skin lesions and patches can show up in most parts of the body. The prevalence varies widely, depending on the ethnicity and geographic location of the population, ranging from 1 to 11%. In India, it affects ~ 3% of adults and is the most common among skin diseases. Various genetic studies of different populations provided evidence of several loci / genes on various chromosomal regions contributing to PSORS, including HLA-Cw6 gene, which has strong association with PSORS. We have studied 53 clinically proven PSORS patients from Gujarat, India and an equal number of age matched controls from the same geographic origin. After informed consent, detailed dermatological evaluation was conducted for each patient included in the study and genetic polymorphism analysis for all the subjects. It was found that 88% of affected subjects were with plaque type psoriasis. Serum oxidative stress markers for example TBARS assay GSH quantification and catalase activity revealed significant difference as compared to normal controls. The ACE genotype with deletion allele frequencies were significantly higher in PSORS patients as compared to normal controls. The present study suggests that the ACE gene polymorphism confers susceptibility to PSORS in the population of Gujarat.

1057S

Multiscale Analysis of Chronic Fatigue Syndrome. *N.D. Beckmann[1,2], G.E. Hoffman[1,2], C. Becker[3], E.E. Schadt[1,2].* 1) Genetic and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

Chronic Fatigue Syndrome (CFS, or Myalgic Encephalomyelitis) is a disease that afflicts between 1 and 4 million people in the United States. Yet the etiology is poorly understood. CFS affects several body systems and is characterized by a severe fatigue or exhaustion that does not improve with rest. A wide variety of symptoms are often exacerbated by physical or mental activity. Here we perform a multiscale analysis of RNAseq and genotype data from 15 patients and 15 controls matched for age and gender to characterize the molecular signature of CFS. The RNAseq data is comprised of 6 cell types (whole blood, monocytes, natural killer cells, granulocytes, B-cells and T-cells) over 4 time points in order to assess the differential response to moderate exercise between patients and controls. The first time point characterizes the baseline level, before exercise; the following time points represent, respectively, one, two and three days after exercising. We consider differential expression of RNAseq data between patients and controls in order to identify genes playing an important role in CFS, in a cell-type or time-specific manner. We construct coexpression networks to identify interpretable subsets of genes that could be perturbed by the disease status and interpret differentially expressed genes in terms of these coexpressed gene modules. Integrating these RNAseq data with genotyping data, we identify expression quantitative trait loci (eQTLs) specific to the disease for each cell type and time point. Our multiscale approach to disease biology elucidates relevant genes, pathways, and cell types, and is a first step toward better understanding CFS.

1058M

Genetic influences common to bronchial asthma and pulmonary tuberculosis present targets for intervention. *R.C. McEachin[1], M.B. Freidin[2], E. Bragina[2], V. Puzyrev[2], L.A. Koneva[1,2].* 1) DCM&B, Univ Michigan, Ann Arbor, MI; 2) Research Institute of Medical Genetics, SB RAMS, Tomsk, Russia.

Purpose: Bronchial asthma and pulmonary tuberculosis (TB) are both serious health issues worldwide. Though asthma is an allergic disease and TB is an infectious disease, both show evidence of genetic influences on susceptibility. Interestingly, epidemiological data show their co-occurrence is rare, relative to the expected. Historically, TB caused more deaths in industrialized countries than any other disease during the 18th to early 20th centuries. While TB is now rare in the progeny of survivors of this epidemic, asthma is prevalent. Equally, in populations that currently have high rates of TB, asthma is rare. One explanation for these observations is that differences in hygiene may influence differential TB and asthma rates. We hypothesize that there are also genetic influences shared across asthma and TB, including risk variants for one disease that are protective for the other. In this work, we test this hypothesis and develop a model that a) places the results into biological context and b) presents gene/drug targets for intervention. Methods: Gene expression is a primary cellular response to the environment, so we looked for genes differentially expressed (DE) in both asthma and TB, but in opposite directions. We assessed DE in T-cells from asthma patients vs controls, and from TB patients vs controls, as well as in lungs of non-infected TB-resistant vs TB-susceptible mice (fold change $\geq \pm 2.0$ AND FDR ≤ 0.05). We used GeneGo to model interactions among DE genes that responded in opposite directions in human TB and asthma, emphasizing transcriptional regulation as the mechanism of action. We used DAVID to assess enrichment of annotation for DE genes in the mouse TB model. Results: Consistent with our hypothesis, 14 human genes were DE in opposite directions in asthma and TB. The network shows how these 14 genes interact, and how a given variant could be a risk factor for one disease and protective for the other. Two genes in the network are known targets for 35 drugs currently in use for a wide range of disorders, including lung disorders. The mouse TB model also yields results consistent with the hypothesis, with “asthma” as the most significantly enriched annotation for the set of DE genes (FDR 1.17E-08). Conclusions: The proposed model presents a novel hypothesis on how TB and asthma responsive genes could have opposite effects in predisposing the two diseases and presents immediate targets for follow on testing in asthma and/or TB.

1059T

Identification of a Common Pathogenesis for Chronic Kidney Disease: Perspectives from Gene Ontology analysis. *W. WU[1], J. Guo[3], R. Yenchek[2], J. Abraham[2], A. Cheung[2], L. Jorde[2].* 1) Department of Human Genetics, University of Utah, Salt Lake City, UT; 2) Division of Nephrology, Department of Internal Medicine, University of Utah, Salt Lake City, UT, United States; 3) Department of Oncological Sciences, University of Utah, Salt Lake City, UT, United States.

Background: Chronic kidney disease (CKD) affects more than 13% of people in the United States, and is one of the leading causes of death. Genetic studies have provided molecular insights into CKD, but there is substantial locus heterogeneity. Yet the clinical manifestations of CKD are often shared across various etiologies. Is there a common pathogenesis that may explain these heterogeneous diseases? Here we employed a bioinformatics approach to identify one or a few common groups of genes for CKD pathogenesis. Methods: We first identified known causative genes associated with CKD curated by the Online Mendelian Inheritance in Man (OMIM). These genes were mapped to the Gene Onotology (GO) networks, and we traversed these trees to trace back to the root nodes. The GO terms hit by these genes and their parent-child relationships were parsed. A statistical hypothesis test with theoretical hypergeometric distribution was carried out to test if the genes are significantly over-represented in each GO node. Simulation was performed to correct for multiple comparisons. These GO terms and their relationships were structured and further visualized for grouping for possible common pathogenesis. Results: From OMIM, we identified 84 causative genes for CKD. When mapping back to the GO networks, 2728 GO terms were hit, among which 2159 were in the biological process (BP) domain, 263 in the cellular component (CC) domain, and 306 in the molecular function (MF) domain. However, after testing for over-representation and correction for multiple comparisons for all terms, only 141 GO terms were statistically significant, among which 117 were in the BP domain, 23 in the CC domain, and 1 in the MF domain. Structure and visualization of these GO nodes and edges showed a diverse pattern across the network. Conclusions: We have derived both positive and negative lists of ontologies to be ruled in or ruled out for future studies of CKD. We also provided a quantitative prior probability for each GO term that can be integrated into a computational model for pathogenesis prediction.

1060S

Pathway Burden Analysis Identifies Genes Underlying Complex Human Limb Disorders. *D. Alvarado[1], G. Haller[1], P. Yang[1], C. Cruchaga[2], M. Harms[3], M. Dobbs[1,4], C. Gurnett[5].* 1) Orthopaedic Surgery, Washington Univ, St Louis, MO, USA; 2) Psychiatry, Washington Univ, St. Louis, MO, USA; 3) Neurology, Washington Univ, St. Louis, MO, USA; 4) Shriners Hospital for Children, St. Louis, MO, USA; 5) Pediatrics, Washington Univ, St. Louis, MO, USA.

Clubfoot, vertical talus and hip dysplasia are congenital birth defects with complex inheritance patterns. The genetic and morphological basis of all three lower limb disorders is poorly understood. Although each disorder describes a different clinical phenotype, their common occurrence within individuals and families suggests a common genetic developmental etiology. To identify the genetic risk factors associated with complex human limb disorders, we performed exome sequencing and genome-wide rare variant pathway burden analysis on 119 patients (88 families) with human lower limb malformations and 298 (249 families) unaffected controls. Using famSKAT, we calculated a collapsed gene burden analysis for all novel coding variants not previously identified in dbSNP137. An enrichment of muscle development, extracellular matrix, hindlimb morphogenesis and chondrocyte differentiation GO term genes was found using the top 100 rank ordered genes. Genome-wide collapsed pathway burden analysis of novel coding variants for all non-redundant GO term gene pathways revealed an association of lower limb malformation cases with collagen catabolism/metabolism and hindlimb morphogenesis GO term genes ($p=1.22\times10^{-5}$ and $p=1.39\times10^{-5}$). Based on these results, candidate genes associated with collagen or hindlimb morphogenesis GO terms with collapsed famSKAT p-value < 0.05 were selected for re-sequencing in a larger cohort. Using Multiplexed Direct Genomic Selection (MDiGS), a newly described BAC capture approach for CNV and SNP/INDEL detection, we sequenced candidate genes in an additional 400 cases. Overall, these results suggest that clubfoot, vertical talus and hip dysplasia are highly heterogeneous disorders related to each other through an increased frequency of novel variants in genes responsible for hindlimb morphogenesis and extracellular matrix organization.

1061M

Integrative Analysis of Transcriptomic and Epigenomic Data to Reveal Regulation Patterns for Osteoporosis. *J.G. Zhang[1, 2], L.J. Tan[3], C. Xu[1, 2], Q. Tian[1, 2], H.W. Deng[1, 2, 4].* 1) Center of Genomics and Bioinformatics, Tulane University, New Orleans, LA, 70112, USA; 2) Department of Biostatistics, Tulane Univ Sch Public Health & Tropical Medicine, New Orleans, LA. 70112, USA; 3) Laboratory of Molecular and Statistical Genetics and Key Laboratory of Protein Chemistry and Developmental Biology of the Ministry of Education, Hunan Normal University, Changsha, Hunan, 410081, China; 4) Center for Systems Medicine, Shanghai University for Science and Technology, Shanghai, 200093, China.

Multiple profiling studies have uncovered a number of risk factors/genes for human complex diseases but are less informative of protein-protein interactions, post-transcriptional modifications and regulation of functional networks. Integration of genomic data and construction of functional gene networks may provide additional insights into the molecular mechanisms of complex diseases. Osteoporosis is a worldwide public health problem characterized by low BMD (bone mineral density) and has strong genetic determination. With the advent of high-throughput technologies, the increasing multi-omics data has created an opportunity for a comprehensive understanding of osteoporosis. In this study, we analyzed gene expression microarray, miRNA microarray and methylation sequencing data of peripheral blood monocytes (PBM) from 5 high hip BMD subjects and 5 low hip BMD subjects. Through integrating the transcriptomic and epigenomic data, we firstly identified BMD-related genetic factors, including 9 protein-coding genes and 2 miRNAs, of which 3 genes (FAM50A, ZNF473 and TMEM55B) and one miRNA (hsa-mir-4291) showed the consistent association evidence from both expression and methylation data, and 3 genes (TMEM55B, RNF40 and ALDOA) were confirmed in the meta-analysis of 7 GWAS samples and GEnetic Factors for OSteoporosis consortium (GEFOS) summary results. Secondly in network analysis we identified an interaction network module with 12 genes and 11 miRNA including AKT1, STAT3, STAT5A, FLT3, hsa-mir-141, hsa-mir-34a and hsa-mir-302b which were proved to be associated with BMD in previous studies. This module revealed the crosstalk among miRNAs, mRNAs and DNA methylation and showed four possible regulatory patterns of gene expression to influence the BMD status, including regulation by gene methylation, by miRNA and its methylation, by transcription factors and co-regulation by miRNA and gene methylation. In conclusion, the integration of multiple layers of omics can bring more evidence than analysis of single layers alone. Integrative analysis from transcriptomics and epigenomic data improves the ability to identify the causal genetic factors, and more importantly uncover the regulation pattern of multi-omics for osteoporosis.

1062T

The Brainstorm project; a cross-phenotype analysis of 14 brain disorders by heritability-, constraint- and pathway-based methods, using genome-wide association data from 500,000 samples. *V. Anttila[1,2,3], S. Ripke[1,2], R. Malik[4,5], T.H. Pers[2,6], K.-H. Farh[2,7], L. Duncan[1,2,3], A. Biffi[2,8,9,10], P. Lee[7], K. Kendler[11,12], J. Scharf[7], J. Smoller[7], A. Palotie[2,7,13], M. Daly[1,2], J. Rosand[8,9,10], B. Neale[1,2,3].* 1) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Harvard Medical School, Boston, MA; 4) Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-University, Munich, Germany; 5) Munich Cluster for Systems Neurology (Synergy), Munich, Germany; 6) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA; 7) Psychiatric and Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA; 8) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 9) Division of Neurocritical Care and Emergency Neurology, Department of Neurology, Massachusetts General Hospital, Boston, MA; 10) J. Philip Kistler Stroke Research Center, Massachusetts General Hospital, Boston, MA; 11) Departments of Psychiatry and Human Genetics, Virginia Commonwealth University, Richmond, VA; 12) Virginia Institute for Psychiatric and Behavior Genetics, Richmond, VA; 13) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland.

Many neurological and psychiatric diseases have considerable co-morbidity, and the strength of the etiological boundaries is a topic of active debate. However, studying these co-morbidities in the real world has proven difficult, and new approaches are needed. To this end, hypothesis-free methods using large genetic datasets may provide a way to study these co-morbidities by examining the genetic correlation through orthogonal approaches. Our aim is to provide insight into the molecular basis of these phenotypic co-morbidities and leverage large-scale genome-wide association data to reveal more of the underlying pathophysiology of an array of brain disorders in a statistically robust fashion. To this end, we have set up collaborations with a number of disease-specific genetics consortia, to obtain summary statistics from roughly 150,000 cases from 14 different diseases and corresponding controls. Meticulous efforts have been applied to identify overlapping studies and samples, using algorithmically generated identifiers where possible, as well as using a novel statistical approach to account for overlapping summary statistics where needed. In addition to successfully replicating and expanding on the previous cross-disorder results from the Psychiatric Genetics Consortium, we report the first comprehensive look into the genetic co-morbidity of common neurological diseases, as well as cross-over analyses between each. Using a novel uni- and bivariate heritability estimation method, we report the total observed heritability for each disease and genetic correlations between each disease pair. In addition, we report some tissue-of-effect analyses using ENCODE and GTEX data, as well as analysing pairwise correlations through pathway membership of the associated loci in existing in silico data. In this study, we set out to integrate the best available neurological and psychiatric genome-wide association results, and to use them to explore cross-disorder analyses of common genetic risk factors using large-scale genetic data.

1063S

RAI1 is a multi-hit regulator of obesity gene expression networks. *J.T. Alaimo[1], S.V. Mullegama[1], R. Tahir[2], J. Beach[2], J. Cruz[2], E. McMullen[2], B. Burns[2], S.H. Elsea[1,2].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA.

Smith-Magenis syndrome (SMS, OMIM 182290) is a complex neurobehavioral disorder characterized by a deletion of chromosome 17p11.2 and due to haploinsufficiency of *retinoic acid induced 1* (*RAI1*). Affected individuals have an array of phenotypic features that include sleep disturbance, varying degrees of intellectual disability, skeletal and craniofacial abnormalities, neurological and behavioral issues, and obesity. A majority of individuals with SMS are obese by early adolescence, and SMS mouse model work has further implicated *Rai1* in both obesity and diet induced obesity. We used a three-pronged approach to identify obesity-related gene networks and tested whether these networks were dysregulated in SMS. First, we utilized Human Phenotype Ontology and identified 156 genes that are known to cause a significant increase in body weight. Second, we took a global genomics approach and identified 723 RAI1-regulated genes by performing gene expression microarrays using HEK293 cells haploinsufficient for *RAI1* generated by siRNA knockdown. Third, we narrowed our focus to genes that are neuronally regulated by RAI1 and performed an additional gene expression microarray using neuroblastoma cell lines haploinsufficient for *RAI1* generated by siRNA knockdown and found 549 differentially expressed genes. Using Ingenuity Pathway Analysis, we analyzed and integrated all three gene lists to generate obesity-related co-expression gene networks. Expression analysis of genes within each network in both SMS patient fibroblasts and across multiple SMS mouse model tissues (brain, liver, abdominal fat, subcutaneous fat) found that *RAI1* haploinsufficiency causes the dysregulation of multiple nodes, including *MC4R*, *POMC*, *CMKLR1*, *BBS4*, and *BMP2K*. Bioinformatics analyses revealed multiple RAI1 consensus binding sites amongst the dysregulated genes, suggesting possible regions of direct regulation. To test one of the genes for responsiveness to RAI1 transcriptional regulation, we employed a luciferase assay using the putative binding region in *CMKLR1*. Reporter assays revealed that RAI1 acts as an enhancer of *CMKLR1* expression, confirming expression data in both SMS cells lines and mouse tissues and the physical interaction of RAI1 at the consensus site. These data support RAI1 as a regulator of multiple obesity-related gene networks and implicate RAI1 as a contributor to early onset obesity.

1064M

Uncovering the pairs of tissue-epigenetic mark which are relevant for a trait. *T. Bigdeli, D. Lee, B.P. Riley, A.H. Fanous, K.S. Kendler, S-A. Bacanu.* Virginia Commonwealth University, Richmond, VA.

Background: For many phenotypes there is a strong working assumption regarding which tissues are biologically relevant to the underlying pathophysiology (e.g. brain tissues for psychiatric disorders). While network and gene-set methods often support these a priori expectations, they largely analyze each tissues univariately. However, only a multivariate analysis of all tissues can empirically assess the likelihood of these assumptions being true. Moreover, such a joint analysis might provide much stronger evidence for other-perhaps even unexpected-tissues as being relevant for the investigated trait. **Method:** We develop a method based only on summary statistics from genome-wide association studies (GWAS), which evaluates the significance of the multivariate (partial) effect on a given trait for all tissue-mark pairs provided by the Roadmap Epigenomics and ENCODE projects. The statistical procedure consists of a regression approach based on the estimated correlation matrix between signals from GWAS and tissue-mark pairs. Given the high levels of correlation (i.e. linkage disequilibrium) between adjacent GWAS SNPs, we divide these into 1,000 disjoint sets, thus ensuring an average spacing of at least 300Kb between variants in the same set. Upon the analysis of each such set of quasi-independent SNPs, for each tissue-mark pair, the ordered series of 1,000 correlated coefficients is statistically tested for a non-zero mean assuming a circular ARMA (1,1) model. **Results:** Application of the proposed method to more than 25 publicly available datasets suggests new and surprising etiological pathways, while maintaining false discovery rates at nominal levels. Among those psychiatric disorders we considered, only for bipolar disorder were brain tissues relevant at statistically significant levels. For schizophrenia, epigenetic marks for certain lymphocyte types yielded much larger signals compared to those for brain tissues. Perhaps unexpectedly, the exact lymphocyte-mark pairs are also the most prominent in rheumatoid arthritis. However, maybe even more unexpectedly, cortical epigenetic marks seem to be the most important tissue-mark pair for some seemingly unrelated phenotypes, such as body-mass index. **Discussion:** The identification of novel, putatively relevant tissues has the potential to inform our as yet incomplete understanding of these disorders, and, in some instances, could even suggest possible treatments.

1065T

Network analysis to identify human genes influencing susceptibility to Mycobacterium tuberculosis and Nontuberculous mycobacteria infection. *E.M. Lipner[1,2], M. Strong[2].* 1) University of Colorado, School of Medicine, Department of Pharmacology, Aurora, CO, 80045, US; 2) National Jewish Health, Integrated Center for Genes, Environment, and Health, Denver, CO 80206, US.

Mycobacterial diseases, including tuberculosis (TB) and nontuberculous mycobacterial (NTM) infections, result in over 1.4 million deaths per year. In order to investigate and identify human genes that may influence control or susceptibility to mycobacterial disease, we compiled gene-disease relational information from several databases. We then performed network and enrichment analysis to identify relevant and enriched human genes and pathways. We hypothesized that such an approach would provide a more comprehensive understanding of the biological pathways involved in mycobacterial infection, and facilitate the identification of novel, candidate genes that have not been previously identified via conventional genetic analytic methods. In order to identify genes associated with TB and NTM phenotypes, we searched three publically available databases; OMIM, HuGE Navigator, and CTD. Our search identified 50 TB-associated and 15 NTM-associated genes, with literature support. Based on these genes, we employed network and functional enrichment analysis using bioinformatic tools including Ingenuity Pathway Analysis (IPA) and DAVID in order to build and examine networks of TB-associated genes, NTM-associated genes, and combined TB and NTM gene sets. For TB, the highest ranking network included 14 query genes in a 35 gene network, including *IL12B* (MIM 161561), *TLR1* (MIM 601194), *TLR4* (MIM 603030), *TLR6* (MIM 605403), *TLR9* (MIM 605474), *TIRAP* (MIM 606252), as well as the putatively associated genes *IRAK* (MIM 300283) and *Akt* (164730). For NTM, the highest ranking network included 14 query genes in a 35 gene network, including *IL12B* (MIM 161561), *TLR2* (MIM 603028) and *STAT1* (MIM 600555), and putatively associated genes *NFkB* complex (MIM 1644011), *ERK1/2* (MIM 176948) and *MAPK14* (MIM 600289). We also investigated the intersection and union among these gene sets, which identified *NFkB* complex (MIM 1644011), *Vegf* (192240), *CSF2* (MIM138960) and *Fcer1* (MIM 147140) as potential candidate genes for further investigation in the role pertinent to mycobacterial disease. Our approach places previously identified and new TB and NTM associated genes in a systems biology framework, where we can investigate the overlap of genes and pathways that may imply similar susceptibility and control mechanisms for TB and NTM. Additionally, these networks suggest candidate genes that may have biological relevance to TB and NTM infection susceptibility for further study.

1066S

Reevaluating the clinical delineation of inflammatory bowel disease using genetic and subphenotype data. *L. Jostins[1,2], International IBD Genetics Consortium.* 1) Statistical Genetics, Wellcome Trust Centre Human Genetics, Oxford, United Kingdom; 2) Medical Genomics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Inflammatory bowel disease (IBD) is generally considered to consist of two major diseases, Crohn's disease (CD) and ulcerative colitis (UC). However, both show significant clinical heterogeneity, and the delineation between the diseases remains the subject of significant controversy. Recent developments in IBD genetics (with >160 loci described to date) may help shed light on the clinical characteristics of this heterogeneous disorder.

Clinicians across 54 centers collated IBD subphenotypes on 34,819 patients (19,713 CD, 14,683 UC) according to the Montreal Classification. All subjects had genotypes generated by Immunochip and data was QC'd and collated at a central data coordinating center. The clinical data recapitulated known epidemiological results, such as the static nature of CD location (colonic vs ileal) over time, and the dependence of disease progression on location at diagnosis.

We confirmed known subphenotype loci (at *NOD2* for CD location, $p<10^{-34}$, and MHC for multiple traits, $p<10^{-22}$), and identified a novel signal (at *MST1* for age of onset, $p<10^{-11}$). Aggregate risk scores of other IBD risk loci showed association with all subphenotypes (e.g. CD vs UC score with CD location, $p<10^{-17}$), suggesting a role for most or all IBD risk loci in disease subphenotype. Reverse regression suggested that the majority of the outcome subphenotypes associations in CD were explained by factors that are present at diagnosis (the location of the disease and the age of onset).

Different risk variants and scores predict different aspects of disease subphenotype (including age of onset and disease outcome), but the dominant axis appears to separate progressive, restricted ileal disease from more stable colonic disease. UC and ileal CD are on opposite ends of this axis, and colonic CD has a genetic profile that is distinct from, and intermediate between, the two. Blind rephenotyping by clinicians of outlying patients (e.g. CD patients with UC-like genetic profiles) demonstrated diagnostic blur at the extremes of this axis. 27% of outlying patients had a doubtful or incorrect diagnosis, compared to 8% of randomly selected patients ($p<0.001$). These results, combined with epidemiological data, suggest that IBD should be considered a spectrum of disorders, with Crohn's colitis as a distinct disease intermediate between UC and ileal CD.

1067M

Integration of diverse genomic datasets identifies novel pathways and key regulatory networks for type 2 diabetes and related traits. *L. Shu, V. Mäkinen, X. Yang.* Department of Integrative Biology and Physiology, University of California Los Angeles, Los Angeles, CA.

Type 2 diabetes (T2D) is a highly prevalent metabolic disease and is among the top leading cause of death in developed countries. Recent genome-wide association studies revealed a number of susceptible loci associated with T2D and its relevant traits. However, the molecular mechanisms underlying many of these loci remain unclear and together these loci only explain ~10% of the genetic heritability. In this study, we utilized an integrative genomics approach to leverage tens of existing, diverse genetic and genomic datasets from humans and mouse including 1) GWAS of T2D and related traits such as HbA1c and beta-cell function indices from DIAGRAM and MAGIC consortium, 2) knowledge-driven biological pathways, 3) data-driven, tissue-specific gene co-expression networks, and 4) data-driven, tissue-specific Bayesian networks. The integration of these diverse datasets allowed us to identify tissue-specific gene subnetworks (i.e., specific parts of gene regulatory networks) that are enriched for risk variants of T2D and relevant traits, as well as the key regulators of these subnetworks. Our results confirmed that subnetworks for beta-cell regulation and glucose transport are shared among the different T2D traits. More interestingly, we captured multiple trait-specific subnetworks such as incretin, adipocytokine and vesicle budding subnetworks for T2D, amyloid, nitric oxide metabolism and protein folding subnetworks for HbA1c trait, and bile acids and iron transport subnetworks for beta-cell function traits. Finally, by utilizing the gene network topology of the subnetworks, we pinpointed various tissue-specific key regulators, both known (e.g. ACSL1 in adipose and FASN in liver for T2D; ATF3 in islet for beta-cell function) and novel (e.g. BIRC5 and GLO1 in adipose and RGS12 in muscle for T2D; ZBTB10 in islet for beta-cell function). Our integrative analysis provides deeper insights into the mechanisms underlying T2D development and reveals novel key regulators, which could serve as potential targets for T2D intervention when experimentally validated.

1068T

Using random forests to identify genetic links between Alzheimer's disease and type 2 diabetes. *B. Darst[1], C. Yao[2], R. Koscik[3,4], B. Bendlin[3,5], B. Hermann[3,4], A. La Rue[3,4], S. Johnson[3,4,5], M. Sager[3,4], C. Engelman[1].* 1) Department of Population Health Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; 2) Department of Dairy Science, University of Wisconsin, Madison, WI, USA; 3) Alzheimer's Diseases Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; 4) Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; 5) Geriatric Research Education and Clinical Center, Wm. S. Middleton Memorial VA Hospital, Madison, WI, USA.

Background: Increasing evidence suggests that type 2 diabetes (T2D) is a risk factor for Alzheimer's disease (AD), but the genetic mechanism linking these conditions is unknown. Purpose: The purpose of this study was to use random forest to investigate whether interactions between single nucleotide polymorphisms (SNPs) in the vacuolar protein-sorting pathway, linked to both AD and T2D, and risk factors for T2D, influence cognition in a cohort of middle-aged adults enriched for a parental history of AD. Methods: Participants from the Wisconsin Registry for Alzheimer's Prevention were included with data on age, gender, homeostasis model assessment-estimated insulin resistance index (HOMA-IR), waist-hip ratio (WHR), physical activity, *APOE* status, 30 SNPs in the *SORL1* and *SORCS1* genes, and 4 cognitive factors (Immediate Memory, Verbal Learning & Memory, Speed & Flexibility, and Visual Learning). These variables were input into random forest, a machine-learning algorithm that calculates importance scores based on the variance explained by each variable in a model while allowing for interactions. Because random forest does not specifically identify interacting variables, we used a novel approach that explicitly identifies potential interactions by determining how often a pair of variables descends together in a random forest. Results: Utilizing a sample of 836 individuals, our findings strongly suggest that rs7907690 within *SORCS1*, and rs2282649 and rs1010159 within *SORL1*, interact with known predictors of insulin resistance to influence cognition. Rs7907690 appeared in the top 25 descendant pairs for all 4 cognitive factors, consistently paired with WHR, HOMA-IR, age, and physical activity. Rs2282649 and rs1010159 also appeared in the top 25 for all cognitive factors paired with the same traits, although not quite as consistently. Many of the interactions identified with the descendant pair method consisted of discordantly ranked pairs, with one variable having a high importance score and the other having a low importance score. Conclusion: These results suggest that interactions between SNPs associated with AD and T2D and strong risk factors for T2D may contribute to the relationship between AD and T2D. Further, the observed discordant descendant pairs indicate that this method captures interactions that the standard random forest method does not.

1069S

Clusters of urate transporter genes as genetic biomarkers in the early detection, diagnosis and prediction of gout. *C. Chung[1,2], H. Tu[3], A. Ko[4], S. Lee[5], H. Lai[6], C. Lee[7], C. Lee[8], C. Huang[9], Y. Ko[1,2].* 1) Graduate Institute of Clinical Medical Science, China Medical University, , Taichung, Taiwan; 2) Environment-Omics-Disease Research Center, China Medical University Hospital, Taichung, Taiwan; 3) Department of Public Health and Environmental Medicine, School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan; 4) Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig D-04103, Germany; 5) Division of Plastic Surgery, Department of Surgery, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan; 6) Division of Rheumatology, Allergy and Immunology, Department of Internal Medicine, Chang Gung Memorial Hospital-Kaohsiung Medical Center, Kaohsiung, Taiwan; 7) Department of Public Health, Kaohsiung Medical University, Kaohsiung, Taiwan; 8) Graduate Institute of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan; 9) Division of Immunology and Rheumatology, China Medical University Hospital, Taichung, Taiwan.

Objective: Gout is the most prevalent inflammatory arthritis in men. We aimed to evaluate the contribution of clusters of urate transporter genes and environmental factors to provide predictive biomarkers for gouty arthritis progression. Methods: We took a pathway based candidate gene approach to investigate the association between genetic variants in clusters of urate transporter genes and gout. Four single nucleotide polymorphisms (SNPs) in urate transporter pathway (rs2231142 in ABCG2 gene, rs3733591 and rs1014290 in SLC2A9 gene, and rs475688 in SLC22A12 gene) were genotyped in 558 male Taiwanese Hans. Results: Per copy increment in the Q141K [T], rs1014290 [A] and rs475688 [C] independently predicted risk of tophaceous gout (OR=3.33, P=9.76×10-7; OR=2.18, P=1.08×10-2; OR=2.50, P=3.87×10-4, respectively). Multifactor dimensionality reduction identified interaction of Q141K, rs1014290 and rs475688 locus had OR=14.01 (95% confidence interval [CI] 6.52-30.11) the risk of developing tophaceous gout. Higher genetic scores of high risk alleles at three genes were associated with hyperuricemia (OR=1.83, P=1.63×10-2), non-tophaceous gout (OR=5.12, P=7.18×10-10) and tophaceous gout (OR=14.03, P=5.86×10-11) and decreased with earlier onset age in tophaceous gout patients (p-trend=2.88×10-3). The calculated attributable fraction indicates that 7.07% of hyperuricemia patients, 20.5% of non-tophaceous gout and 39.3% of tophaceous gout progression can attribute their disease to the detrimental effects of higher genetic risk scores Conclusion: Our results provide additional insight into genetic mechanisms underlying gout arthritis.

1070M

Identification of novel predisposing genes in neural tube defects by whole exome sequencing in multiplex families. *P. Lemay[1], A. Dionne Laporte[2], D. Spiegelman[2], E. Henrion[2], O. Diallo[2], P. De Marco[3], E. Merello[3], G.A. Rouleau[2], V. Capra[3], Z. Kibar[1].* 1) CHU Ste-Justine, Montreal, Canada; 2) Montreal Neurological Institute, Mcgill University, Montreal; 3) Gaslini hospital, Genova, Italy.

Neural tube defects (NTD), including anencephaly and spina bifida, are among the most common and severe birth defects affecting nearly 1 in 1000 births. They are caused by partial or complete failure of neural tube closure in the rostro-caudal axis during embryogenesis. The etiology of NTD is complex including genetic and environmental factors that remain largely unknown. While preconceptional intake of folic acid protects against 50-70%; of NTDs, they still affect thousands of families urging the need for novel preventive and counseling strategies. Here we suggest and approach prioritizing highly pathogenic variants and biological processes burdened by a high amount of rare coding mutations to better understand the etiology of NTDs. The main goal of our study is to identify new candidate pathways or genes implicated in NTD development. Our hypotheses are: (1) rare coding mutations are an important part of the NTD etiology and (2) whole exome sequencing in multiplex affected families and prioritization based on highly pathogenic mutations and biological process enrichment will identify new NTD related genes. To date, 11 multiplex families with at least 2 affected individuals were recruited. Exome sequencing data obtained on 32 individuals from 8 families was treated by a bioinformatics pipeline using BWA, GATK v2.7 and ANNOVAR. The output was filtered for rare (>5% in public databases) and high quality mutations resulting in a list of 200 to 300 mutations per family. The list was then screened for enrichment of biological processes and highly pathogenic mutations. Candidate processes or genes are then validated in a cohort of 87 control families. Out of the 8 families already sequenced, three families were enriched for mutations in certain biological processes using both the DAVID 6.7 program and our control cohort. Our results suggest that whole exome sequencing combined to a hypothesis free gene enrichment approach may offer new insights in the complex etiology of NTDs. Identification of new genetic factors will lead to better clinical management of affected families and may identify novel protective supplements.

1071T

The Genetic Architecture of Age-Related Macular Degeneration in the Amish. *J.D. Hoffman[1], J.N. Cooke Bailey[2], R.J. Sardell[4], A.C. Cummings[1], L.N. D'Aoust[1], W.K. Scott[4], M.A. Pericak-Vance[4], J.L. Haines[2,3].* 1) Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN; 2) Department of Epidemiology & Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH; 3) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH; 4) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL.

The Amish are a genetically and culturally isolated founder population descended from Swiss and German Anabaptists who emigrated from Western Europe to North America in the 1700 and 1800s. The Amish also live a more homogeneous lifestyle, thus reducing the variability of environmental factors typically seen in complex diseases. Previous work demonstrated that the cumulative genetic risk across known age-related macular degeneration (AMD) loci is significantly lower in Amish AMD cases as compared to non-Amish European American cases, although the prevalence of AMD is similar. In this study we perform a pathway analysis using data from an Affymetrix 6.0 SNP chip to explore the potential role of functional pathways in AMD in the Amish population. Single-variant association analysis corrected for relatedness was performed on 92 Amish cases and 640 Amish controls using the MQLS program. Pathway analysis was performed on the results of the single variant tests using the Pathway Analysis by Randomization Incorporating Structure (PARIS) algorithm, and restricted to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database, which contains 199 defined pathways. Results of the AMD pathway analysis showed nominal associations with the heparan sulfate biosynthesis (HSBS; $p=8.0*10^{-4}$), MAPK signaling (p=0.0002) and FC∈ R1 Signaling (p=0.003) pathways. Somewhat surprisingly, the complement and coagulation signaling cascade (CCSC) pathway was not significant (p=0.47). We compared these results to a non-Amish European-American dataset of 534 cases and 379 controls. In the non-Amish dataset, the p-values for the HSBS, MAPK, and FCER1 pathways were 0.995, 0.10, and 0.91 respectively while the CCSC pathway was highly significant ($p<5.0*10^{-5}$). The results for the HSBS pathway are interesting since heparin sulfate plays a role as an inactivator in the complement alternative pathway through cleavage of C3b, also a target of complement factor H (CFH). From this preliminary analysis we observe signals in pathways previously shown to be associated with AMD. Interestingly, the complement pathway is not significant in this analysis; a major genetic risk factor for AMD. These data further support the hypothesis that the genetic architecture in the Amish is different than in the general European-descent population.

1072S

Genome-wide profiling of gene expression and DNA methylation changes in Alzheimer's disease brains. *M. Allen[1], D. Serie[4], Z. sun[3], S. Baheti[3], M. Walsh[4], F. Zou[1], H.S Chai[3], C.S Younkin[1], J. Crook[4], V.S Pankratz[3], M.M Carrasquillo[1], A. Nair[3], S. Middha[3], S. Maharjan[3], T. Nguyen[1], L. Ma[1], K. Malphrus[1], S. Lincoln[1], G. Bisceglio[1], R.C Petersen[6], N.R Graff-Radford[2], D.W Dickson[1], S.G Younkin[1], T. Ordog[5], Y.W Asmann[4], N. Ertekin-Taner[1,2].* 1) Dept Neuroscience, Mayo Clinic Florida, Jacksonville, FL; 2) Department of Neurology, Mayo Clinic Florida, Jacksonville, FL; 3) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 4) Department of Health Sciences Research, Mayo Clinic Florida, Jacksonville, FL; 5) Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, MN; 6) Department of Neurology, Mayo Clinic, Rochester, MN.

Late-Onset Alzheimer's disease (LOAD) is a complex neurodegenerative disease with a substantial genetic component. Despite the success of GWAS studies, much of the heritability of LOAD remains unexplained. We hypothesize that evaluation of the genetic and epigenetic mechanisms that influence gene expression in the brain will lead to identification of novel LOAD genes and improved understanding of vulnerable molecular pathways. We previously performed a gene expression GWAS that assessed genome-wide genotypes and mRNA levels of ~24,000 transcripts in two brain regions (temporal cortex and cerebellum) for ~ 200 AD subjects and ~200 subjects with other pathologies (Non-AD). Gene expression levels were measured using Illumina's WG-DASL array and genotypes were generated using the Illumina Hap300 chip. We have now performed transcript profiling analysis comparing gene expression levels in AD vs. non-AD subjects in both brain regions, using linear regression carried out in R, and conducted pathway enrichment analysis on these data using the functionalities in Metacore. In addition, we have collected CpG methylation data (methylome), using reduced representation bisulfite sequencing, on a subset of the temporal cortex samples (55AD, 55 NonAD) to evaluate the influence of DNA methylation on gene expression in the brain. DNA methylation levels at promoter regions were tested for association with gene expression levels using linear regression carried out in R. Transcript profiling analysis identified 743 targets in the cerebellum and 2839 targets in the temporal cortex (un-corrected pdiff <0.01) that were selected for pathway analysis. In the temporal cortex the most significant pathway identified was oxidative phosphorylation. Many of the top differentially expressed genes replicated across both brain regions and represent the most robust observed changes. DNA methylation-gene expression analysis identified more significant associations than expected by chance alone. Many of the top genes identified in this analysis were also found to be differentially expressed in the transcript profiling study indicating that epigenetic changes that influence gene expression may be associated with disease status. Further integrative analysis of genome-wide genotype, methylation and expression data in this cohort provides an opportunity to enhance our understanding of the role of genetic and epigenetic changes on brain gene expression and risk for LOAD.

1073M

Heterogeneous Network Link Prediction Prioritizes Disease-Associated Genes. *D. Himmelstein[1], S. Baranzini[1,2].* 1) Biological and Medical Informatics, University of California, San Francisco, San Francisco, CA; 2) Department of Neurology, University of California, San Francisco, San Francisco, CA.

The first decade of Genome Wide Association Studies (GWAS) has uncovered a wealth of disease-associated variants. Two important tasks will be translating this information into a multiscale understanding of pathogenic variants, and increasing the power of existing and future studies through prioritization. We show that link prediction on heterogeneous networks—graphs with multiple node and edge types—accomplishes both tasks. Given a heterogeneous network comprised of diverse information sources, our technique elegantly learns the importance of each connection type. We demonstrate the method's distinctive performance and apply it to infer influential genes in a complex disease phenotype, multiple sclerosis (MS). First we constructed a network with 18 node types—genes, diseases, tissues, pathophysiologies, and 14 MSigDB collections—and 19 edge types from high-throughput publicly-available resources. From this network we extracted features describing the topology between specific genes and diseases. Using a machine learning approach that relied on GWAS-discovered associations as positives during model training, we predicted the probability of association between each protein-coding gene and each of 23 diseases. These predictions achieved a testing AUROC of 0.814 and a 285-fold enrichment in precision at 10% recall. We quantified the performance of each network component and identified pleiotropy, transcriptional signatures of perturbations, pathways, and protein interactions as the most predictive domains. The full model greatly outperformed any individual domain, highlighting the benefit of integrative approaches. We created an interactive prediction browser, available at http://het.io, that decomposes a prediction to show the contribution of each component. Since the entirety of GWAS knowledge is incorporated into every prediction, our method performed well for diseases with few seed genes—where most existing methods fail. We masked the largest MS GWAS from our network leaving only 15 MS-associated genes. Nonetheless, our method predicted the 36 genes discovered by the masked study with an AUROC of 0.78. Finally, we predicted 5 novel susceptibility genes for MS, 4 of which (JAK2, REL, TNFAIP3, RUNX3) achieved Bonferroni validation on the 9,772-case masked GWAS. Regions containing two of these genes were uncovered in a recent MS ImmunoChip-based study, highlighting our ability to identify the causal gene within a locus.

1074T

Expression pathway analysis for genes associated with rheumatoid arthritis. *K. Shchetynsky, L. Diaz, L. Klareskog, L. Padyukov.* Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden.

Background: Whole-transcriptome RNA expression analysis provides new insight into disease mechanisms. Interpretation of data from studies of differential gene expression in complex human diseases based on RNA-seq data is sensitive to study design and require rigorous verification. Integration of genetic association data with transcriptome analysis facilitates cross validation of expression and GWAS data and may provide additional clue for identification of new candidate genes and gene-networks. We performed this analysis in a common autoimmune disease, rheumatoid arthritis (RA). **Materials and methods:** RNA-seq was performed on whole-blood RNA samples from 12 healthy controls, 6 untreated RA patients, and 6 patients receiving standard treatment with Methotrexate. RNA-seq was done on the Illumina HiSeq 2000 platform with TruSeq library construction and selection for polyadenylated mRNA, with 28 M paired-end reads per sample. TopHat and Cufflinks software packages were used for RNA-seq data analysis. Pathway analysis was carried out with Ingenuity Pathway Analysis software. **Results:** We address the expression of genes containing markers previously associated with RA in GWAS. Out of 109 genes, 39 genes were differentially expressed (DE) between any two of the three groups (Cuffdiff $p<0.05$) - naïve RA patients, treated RA patients and healthy controls. These DE genes were grouped by pathway analysis into 4 functional protein networks, involving 75 additional "connector" molecules. Out of these 75 corresponding genes suggested by pathway analysis, the information for 37 were present in our RNA-seq data, and 15 were differentially expressed (Cuffdiff $p<0.05$). The majority of these new candidate genes (9/15) coded for proteins from a single network, where tumor protein p53 (TP53) was a hub, directly connected with other 8 genes. Clustering samples based only on the expression of the 9 members of the TP53 network produced reasonable grouping of healthy controls and patients in our data. **Conclusion:** An integration of RNA-seq data with prior information from association studies allowed us to infer new candidate genes for rheumatoid arthritis using pathway analysis. Our data implies, that TP53 could be a hub in the network that could contribute to the molecular pathological mechanisms behind rheumatoid arthritis.

1075S

Stratified enrichment test for dissecting colocalized genomic annotations to fine-map complex trait variants. *G. Trynka[1,2], B. Han[1,2], K. Slowikowski[1,2], HJ. Westra[1,2], H. Xu[3,4], M. Daly[5,2], XS. Liu[3,4], R. Klein[6], S. Raychaudhuri[1,2,7].* 1) BWH, Harvard Medical School, Boston, MA; 2) Broad Institute of Harvard and MIT, Boston, MA, USA; 3) Dana-Farber Cancer Institute, Boston, MA, USA; 4) Harvard School of Public Health, Boston, MA, USA; 5) Massachusetts General Hospital, Boston, MA, USA; 6) Icahan School of Medicine at Mt. Sinai, New York NY USA; 7) University of Manchester, Manchester, UK.

Motivation: Enrichment tests offer the opportunity to translate genetic associations to biological mechanisms by integrating trait-associated variants with genomic annotations. However, genomic annotations highly colocalize with each other: regulatory elements cluster in the proximity of genes, DNase hypersensitive sites (DHS) colocalize with exons and transcription factors co-occur with enhancers. This is a critical issue because, if the colocalization of annotations is not accounted for, the observed enrichment might reflect inappropriate annotations rather than the most biologically relevant annotations. This might obscure fine-mapping efforts. Method: We developed a stratified enrichment test (*GoShifter*) that uses a local annotation shifting method to robustly quantify the enrichment of a given annotation while taking into account co-localized background annotations. Results: (1) We tested *GoShifter* using simulated GWAS SNPs, where associated SNPs tag loci at exons or DHS, as these are known to colocalize. Standard test that does not correct for colocalization between these annotations resulted in 48% higher enrichment, while *GoShifter* reported enrichment in only 10% of time, reflecting appropriate false positive rate. (2) We applied *GoShifter* to SNPs influencing gene expression (eQTL) in whole blood. The unstratified test resulted in significant enrichment for DHS, genes, proximity to transcription start (TSS) ($p<10^{-4}$) and histone marks (H3K4me3, H3K9ac and H3K4me1, $p<0.01$), many of which colocalize. Stratified analysis with *GoShifter* revealed that the DHS enrichment best explains eQTL associations. (3) Finally, we applied *GoShifter* to NHGRI GWAS Catalog variants and observed that DHS were the most strongly enriched annotation ($p<10^{-4}$) above and beyond TSS proximity or genes. H3K4me3 and H3K4me1 showed less significant, but independent signals, from DHS enrichment ($p<0.05$). We estimate that 20-40% of GWAS Catalog loci capture causal variants in DHS, a much lower proportion than the widely reported 80%. Conclusions: *GoShifter* can dissect colocalized annotations to distinguish the true driving annotations from spuriously enriched annotations due to annotation correlations. We expect that it will be widely used to understand the molecular mechanisms underlying common traits associations.

1076M

Transcriptional Analysis of Sepsis Patients Reveals Differential Expression Patterns. *D.L. Dinwiddie[1], E.L. Tsalik[2,3], J.C. van Velkinburgh[4], C.W. Woods[3,5], N.A. Miller[6], S.F. Kingsmore[6], R.J. Langley[7].* 1) Department of Pediatrics and Clinical Translational Science Center, University of New Mexico Health Sciences Center, Albuquerque, NM; 2) Research Service and Emergency Medicine Service, Durham Veterans Affairs Medical Center, Durham, NC; 3) Duke Institute for Genome Sciences & Policy and Department of Medicine, Duke University School of Medicine, Durham, NC; 4) National Center for Genome Resources, Santa Fe, NM; 5) Medicine Service, Durham Veterans Affairs Medical Center, Durham, NC; 6) Center for Pediatric Genomic Medicine, Children's Mercy Hospitals and Clinic, Kansas City, MO; 7) Department of Respiratory Immunology, Lovelace Respiratory Research Institute, Albuquerque, NM.

The Community-Acquired Pneumonia and Sepsis Outcomes Diagnostics (CAPSOD) project is a multidisciplinary observational collaboration to understand sepsis pathophysiology and identify biomarkers. This study was designed to identify biomarkers that differentiate sepsis from systemic inflammatory response syndrome (SIRS) of other causes, define the biology of sepsis progression, and understand the molecular predictors of sepsis mortality. To this end, we conducted transcriptome sequencing of 132 patients with SIRS suggestive of possible sepsis at emergency department presentation, t0. We conducted expression analysis with TopHat and Cufflinks pipeline using HTseq-Tools. A distinct transcriptional response was identified in sepsis survivors, sepsis nonsurvivors, and controls. Differentially expressed genes were analyzed to determine related biological pathways and functions. Expression patterns were also analyzed based on pathogenic organism (gram negative vs gram positive). Finally, we investigated the ability to use variant calls from RNA-seq to examine genomic variation that might impact RNA expression using STAR alignment and variant calling with VQSR by the GATK v3.1. In summary, we provide an in depth transcriptional analysis of patients with SIRS, sepsis survivors, and sepsis deaths and provide evidence of differential gene expression patterns capable of delineating the three groups.

1077T

Analysis of Endosomal Trafficking and Protein Recycling Genes in Parkinsonism. *E.K. Gustavsson[1,2,3], J. Trinh[1], I. Guella[1], H.E. Sherman[1], C. Szu-Tu[1], S. Appel-Cresswell[4], A.J. Stoessl[4], A. Rajput[5], A.H. Rajput[5], J.O. Aasly[2,3], M.J. Farrer[1].* 1) Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Department of Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway; 3) Department of Neurology, St. Olav's Hospital, Trondheim, Norway; 4) Pacific Parkinson's Research Centre, Department of Medicine (Neurology), University of British Columbia Vancouver, BC, Canada; 5) Division of Neurology, University of Saskatchewan and Saskatoon Health Region, Saskatoon, SK, Canada.

Pathogenic mutations in genes involved in endosomal protein recycling, trafficking and lysosomal autophagy have been implicated in autosomal dominant late-onset Parkinson's disease (PD). One key component of protein recycling and trafficking is the retromer complex, which consists of SNX-BAR dimers (SNX1, SNX2, SNX5 and SNX6), and a cargo recognition trimer (VPS26, VPS29 and VPS35). DNAJC13 also regulates endosomal trafficking and protein recycling by coating and uncoating clathrin-coated vesicles. In this study, we aim to assess rare missense variants (MAF<0.01) in these genes. Next Generation Sequencing (NGS) of *VPS26A/B*, *VPS29*, *VPS35*, *SNX1*, *SNX2*, *SNX5* and *SNX6* and *DNAJC13* were performed in 411 patients with PD, 224 atypical parkinsonism patients and in 371 controls. Patient and control subjects are unrelated and of multi-ethnic background. Identified rare missense variants (MAF<0.01) were Sanger sequenced and subsequently genotyped in two PD cohorts (Canada, 970 cases and 470 controls; Norway, 950 cases and 660 controls). For carriers with family history segregation analysis was performed when DNA was available. Seven variants were observed across *VPS26A*, *VPS29* and *VPS35*, five of which were not seen in the control population. Furthermore, 22 variants were observed across the sorting nexin genes (*SNX1*, *SNX2*, *SNX5* and *SNX6*), twelve of which only within patients. Eighteen mutations were identified in *DNAJC13*, eleven of which were only seen in patients. Segregation analyses in families were unequivocal for these variants. Subsequent genotyping of *DNAJC13* mutations revealed p.Glu1740Gln and p.Leu2170Trp to be more frequent in patients (OR=7.65, p=0.041 and OR=4.12, p=0.007, respectively), and p.Pro336Ala, p.Val722Leu, p.Nsn855Ser, p.Arg1266Gln and p.Thr1895Met were not seen in any control subjects. Deficits in endosomal protein recycling, trafficking and lysosomal autophagy are highlighted by the discovery of pathogenic mutations in *VPS35* and *DNAJC13* (*VPS35* p.Asp620Asn and *DNAJC13* p.Nsn855Ser). The missense mutations within *DNAJC13* and the retromer subunits, in PD and atypical parkinsonsm cases, converge with membrane-protein recycling and endosome-trans-Golgi trafficking as a central component of PD pathogenesis. Further evaluation of components within membrane-protein recycling and endosome-trans-Golgi trafficking in parkinsonism is warranted.

1078S

Transcriptome analyses of patient-specific induced pluripotent stem cell (iPSC) derived neural stem cells implicate neurodevelopmental pathways involved in Tourette Syndrome. *N. Sun[1,2], J. Xing[1,2], L. Deng[1,2], M. Moreau[3], M. Sheldon[1,2,3], JC. Moore[1,2,3], RP. Hart[2,3,4], GA. Heiman[1,2], JA. Tischfield[1,2,3].* 1) Genetics, Rutgers University, Piscataway, NJ; 2) Human Genetics Institute of New Jersey, Piscataway, NJ; 3) RUCDR Infinite Biologicals NIMH Stem Cell Resource, Piscataway, NJ; 4) Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ.

Tourette Syndrome (TS) is a childhood onset neurodevelopmental disorder of unknown etiology characterized by the presence of waxing and waning motor and vocal tics. Family and twin studies showed that TS segregates within families and across multiple generations. However, the transmission pattern of TS is unclear and no common genetic factors and biological pathways are reported to be responsible for all TS patients. In order to look at the manifestation of TS at the gene expression level and to identify critical biological pathways which could be universally affected in all TS patients during early neurodevelopment, we conducted an expression study (RNA-seq) using iPSCs from TS patients and controls. The study included 1) TS patient-specific iPSCs (TS-iPSCs), 2) control iPSCs, 3) TS-iPSC derived neural stem cells (TS-NSCs), and 4) control NSCs. Based on the RNA-seq data, we will determine which genes are differentially expressed in the TS cells compared with control cells. By clustering these differentially expressed genes under co-expression and/or interaction criteria, we aim to unveil biological pathways affected during the neurodevelopmental process of TS patients and further to understand the molecular mechanisms of the etiology of TS.

1079M

Liver-specific long non-coding RNAs and its association with liver disease. *J. Fu[1], B. Atanasovska[1,2], M. van der Sijder[1], S. Rensen[3], P. Deelen[1], D. Zhernakova[1], L. Franke[1], V. Kumar[1], S. Withoff[1], C. Wijmenga[1], M. Hofker[2].* 1) Dept Genetics, UMCG, University of Groningen, Groningen, Netherlands; 2) Department of Paediatrics, Molecular Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 3) Department of Surgery, University Hospital Maastricht and Nutrition and Toxicology Research Institute (NUTRIM), Maastricht University, Maastricht, The Netherlands.

Introduction: Long noncoding RNAs (lncRNAs) have been shown to cover a significant portion of the noncoding transcriptome in mammalian genome. Recently, an increasing number of studies have shown that lncRNAs can be involved in various biological processes and their dysregulation can underlie a number of complex human diseases. However, the role of lincRNAs in liver pathology and overall mechanism remain largely unknown. Here, we investigated the role of lncRNA in the nonalcoholic steatohepatisis (NASH) using expression data and the cis-regulation of lncRNA expression by genetic variants (cis-eQTL). Methods: Whole-transcriptome expression in four different tissue types was simultaneously determined in 93 severely obesity individuals with BMI> 35 kg/m2 and with varying degrees of NASH. We focused on the expression of 2,359 microarray probes that co-localize with lncRNA and assessed their correlation with NASH phenotypes. We further validated the correlation using quantitative RT-PCR (qRT-PCR) in subset of 34 samples and investigated the association between lncRNA expression and genetic variation. Results: The correlation analysis based on microarray data identified 15 significant correlations for 12 lncRNAs at P<2.1×10-5, corresponding to Bonferroni-corrected 0.05 level. We observed that one liver-specifically expressed lncRNA probe was correlated with four different NASH phenotypes: ALAT (r=0.50, P=3.00×10-6); Kleiner score steatose (r=0.47, P=7.02×10-6); Lobular inflammation (r=0.56, P=1.39×10-7) and Large lipoglanulomas (r=0.52, P=7.22×10-7). The further qPCR analysis validated its association in NASH and a large-scale public RNA-seq data in over 3,000 samples confirmed its liver specificity. Cis-eQTL analyses identified liver-specific eQTL for 3 out of 4 liver-specific lncRNA. Conclusion: Our study identified 15 significant correlations between lncRNA expression in multiple human tissues and NASH. Further validation analysis confirmed the association of some liver specific lncRNAs. Our finding indicates the potential role of lncRNAs in liver metabolism and development of NASH and they can be subjected to genetic regulation. Further functional studies are warranted.

1080T

The Effects of Genetic Perturbation on Networks of Phenotypes in Complex Diseases. *J. Chu[1,2], P.J. Castaldi[1,2], M.H. Cho[1,2,3], B.A. Raby[1,2,3], E.K. Silverman[1,2,3].* 1) Channing Division of Network Medicine , Brigham & Women's Hospital, Boston, MA; 2) Department of Medicine, Brigham & Women's Hospital, Boston, MA; 3) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA.

Complex diseases are often assessed using an array of disease-related phenotypic variables, which may have subtle, hidden relationships not captured by standard epidemiological analyses. Investigation of the potentially nonlinear relationships among phenotypes and between phenotypes and specific genetic variants is a challenging problem. To examine the effects of genetic perturbations on the relationships between disease-related phenotypes, we developed a novel network-based approach using the Gaussian graphical model (GGM) to infer the connections between each pair of phenotypic variables and create a phenotypic network based on partial correlations. First, for each SNP we constructed GGM phenotypic networks for groups of subjects defined by their genotypes. Then, for each pair of phenotypic variables, we test for differences in the network connectivity between different genotypes. We applied this method to COPDGene, a large, well-characterized set of smokers from a study of chronic obstructive pulmonary disease (COPD). We selected 10 key quantitative COPD-related phenotypes to build the phenotypic networks, and the effects of perturbations at 467 top loci (P < 0.001) from previous case-control studies of COPD populations were tested on 5,471 non-Hispanic White subjects. Although most SNPs have little or no effect on the phenotypic networks, there are many SNPs with significant effects on multiple interactions, disrupting a large part of network (15-20 out of 45 possible edges), and these SNPs may provide clues regarding underlying mechanisms by which genetic variant influence disease susceptibility and outcome. For example, in the networks characterized by rs7671167 (near *FAM13A*), which was previously associated with COPD in genome-wide association studies, we found that measures of lung function and emphysema were negatively correlated in the group with COPD-protected alleles but not correlated in the group with COPD-associated alleles, which suggests that *FAM13A* may lead to reduced lung function through mechanisms other than increased emphysema. Our results are based on the genetic effects on networks of multiple quantitative phenotypes within one complex disease and would not have been observed using traditional multivariate approaches. More importantly, the results can help understand the relationships between these disease-related phenotypes and provide novel insight into disease susceptibility, disease severity, and genetic mechanisms.

1081S

Mediation effect of eQTLs reveals trans-regulation of gene expression in complex disease traits. *C. Yao[1,2], R. Joehanes[1,2,3], T. Huan[1,2], B.H Chen[1,2], X. Zhang[1,2], C. Liu[1,2], A.D. Johnson[1,2], P. Courchesne[1,2], P. Munson[1,2,3], D. Levy[1,2].* 1) The National Heart, Lung and Blood Institute; the NHLBI's Framingham Heart Study, National Institutes of Health, Bethesda, MD, USA; 2) Population Sciences Branch, National Institutes of Health, National Heart, Lung, and Blood Institute, Bethesda, MD, USA; 3) Mathematical and Statistical Computing Laboratory, Center for Information Technology, National Institutes of Health, Bethesda, MD, USA.

Expression quantitative trait locus (eQTL) analyses have shed light on transcriptional regulation in numerous complex diseases. An eQTL can be classified as cis or trans acting, based on its association with nearby or remote gene expression, respectively. Compared to cis-eQTLs, the underlying regulatory mechanisms responsible for remote effects of trans-eQTLs are still largely unknown. We hypothesize that expressed genes of cis-eQTLs can act as trans regulators. Using 11,672 published trait- or disease-associated SNPs for 427 complex traits, we explored the gene expression associations of trait-associated SNPs from GWAS that are eQTLs in 5,257 individuals from the Framingham Heart Study. At a false discovery rate (FDR) <0.05 (p<10-6), we identified 177 trans-eQTLs that are associated with expression of 546 genes representing 1162 eQTL-gene pairs. Among these eQTLs, 131 (associated with 231 genes, and representing 589 eQTL-gene pairs) also have cis associations with genes residing within 1 megabase. Mediation testing suggested that for 116 of these 131 eQTLs, the disease-associated SNPs may alter expression of remote (i.e. trans) genes via their cis associations with genes near the eQTL (p<0.005 for 1,000 permutations). Examples of cis mediated effects on remote genes include 1) a trans-effect of rs174546 (a lipid-associated SNP from GWAS) on LDLR and SREBF2 that appears to be mediated by cis effects of the eQTL on FADS1 and FADS2 expression, and 2) trans-effect of rs3184504 (associated in GWAS with type 1 diabetes and blood pressure) on STAT1 that is mediated by expression of nearby TCTN1. We conducted functional enrichment analysis of the eQTLs and the trans-genes potentially mediated by them. For the 90 eQTLs that are annotated in RegulomeDB, we found that 56 (62%) affect transcription factor binding. For the 196 trans-regulated genes annotated in the Database for Annotation, Visualization and Integrated Discovery (DAVID), we found that 20% to 26% were previously reported to be associated with disease phenotypes in OMIM or GWAS studies, providing supportive evidence of a mediation mechanism linking trans-eQTLs to complex diseases.

1082M

Beyond GWAS: Probing the landscape between pathway associations, genome-wide associations and protein-protein interaction networks in Chronic Obstructive Pulmonary Disease. *M. McDonald[1], M. Mattheisen[1,2], M. Cho[1,3], Y. Liu[1], B. Harshfield[1], C. Hersh[1,3], P. Bakke[4], A. Gulsvik[4], C. Lange[5], T. Beaty[6], E. Silverman[1,3] on behalf of the GenKOLS, COPDGene and ECLIPSE study investigators.* 1) Channing Division of Network Medicine, Brigham & Women's Hosp/Harvard Sch Med, Boston, MA,USA; 2) Department of Biomedicine and Centre for integrative Sequencing (iSEQ), Aarhus University, Aarhus, Denmark; 3) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 4) Department of Clinical Science, University of Bergen, Bergen, Norway; 5) Harvard School of Public Health, Boston, MA, USA; 6) Johns Hopkins School of Public Health, Baltimore, MD, USA.

In a traditional genome-wide association study (GWAS), thousands to millions of markers are tested; however, only a small fraction of markers reaches a level of genome-wide significance (GWS). As such, we sought to apply a traditional pathway approach and an integrative systems biology approach to maximize information from GWAS results to gain insight into Chronic Obstructive Pulmonary Disease (COPD). Specifically, we performed gene-based and gene-set associations, defined from gene-sets in the molecular signatures database (MSigDB), with FORGE software. For discovery, we analyzed GWAS results from two well-characterized COPD cohorts, COPDGene and GenKOLS. As replication, we used a third well-characterized COPD case-control cohort (ECLIPSE). Next, we used an integrative systems biology approach that integrates GWAS results with protein-protein interactions (PPI) to identify COPD disease modules with the R library dmGWAS. Using the traditional pathway approach, no gene-sets reached experiment-wide significance in either discovery population. Using the systems biology approach, we identified a consensus network of 10 genes identified in modules by integrating GWAS results with PPI that replicated in COPDGene, GenKOLS, and ECLIPSE. We found enrichment of members of four gene-sets among these 10 genes: (i) lung adenocarcinoma tumor sequencing genes, (ii) IL7 pathway genes, (iii) kidney cell response to arsenic, and (iv) CD4 T cell responses. Further, several genes, including KCNK3 and NEDD4L, have also been associated with pathophysiology relevant to COPD. In particular, KCNK3 has been associated with pulmonary arterial hypertension, a common complication in advanced COPD. Thus, we report a set of new genes that may influence the etiology of COPD that would not have been identified using traditional GWAS and pathway analyses alone. This project was supported by NIH R01 HL089856, R01 HL089897, P01 HL105339, and R01 HL111759. GenKOLS and ECLIPSE (SCO104960, NCT00292552) were funded by GlaxoSmithKline.

1083T

Systemic effects of genetic variation on gene expression. *D. Plichta[1], T. Esko[2, 3, 4, 5], H.B. Nielsen[1, 6], T.H. Pers[1, 2, 3], J.N. Hirschhorn[2, 3, 4].* 1) Center for Biological Sequence Analysis, Technical University of Denmark, Kgs. Lyngby, Denmark; 2) Division of Endocrinology and Center for Basic and Translational Obesity Research, Children's Hospital, Boston, MA, USA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) Department of Genetics, Harvard Medical School, Boston, MA, USA; 5) Estonian Genome Center, University of Tartu, Tartu, Estonia; 6) Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark.

Expression quantitative trait locus (eQTL) analysis is a well-established approach to discover novel biological mechanisms underlying genome-wide association study (GWAS) associations. Recent studies have shown that eQTLs may affect expression levels of multiple nearby and distal genes indicating that eQTLs may impact the transcriptional regulation of biological pathways. To search for common variants with systemic effects on expression levels, we developed an approach to identify eQTLs that regulate modules of co-expressed genes (transcriptional components). We used whole blood expression data from 814 genotyped, Caucasian individuals to construct 775 transcriptional components using independent component analysis (ICA). We annotated each transcriptional component by assessing enrichment of highly loading genes in a diverse set of pre-defined gene sets, including canonical pathways from the REACTOME database, Gene Ontology terms, and sets of genes that upon knock-out in mice result in the same phenotype. We then used individuals' scores on each of the transcriptional components as phenotypes in an eQTL association analysis based on 2,933 independent SNPs that have previously been associated with diseases and complex traits at genome-wide significance. Finally, we applied our approach on Framingham Heart Study expression and genotype data from 2,324 unrelated individuals to replicate our findings. We identified 152 significant eQTLs, which associated with 106 independent transcriptional components (at false discovery rate <10%). Majority of them exerted systemic effects on multiple genes. The strongest eQTL was between a variant previously associated with inflammatory bowel disease (rs1363907; Spearman r=0.8, $P<10^{-100}$) and a transcriptional component enriched for telomere maintenance and respiratory electron transfer genes ($P<0.05$). The variant was in high linkage disequilibrium with a splice site variant (rs2248374), which previously has been reported to result in non-sense mediated decay of the immune system related ERAP2 gene transcripts. The ERAP2 gene was the highest loading gene on the associated transcriptional component but more than 250 other genes loaded high as well (removal of the ERAP2 gene from the transcriptional component did not abolish the association). We replicated this association along with 33 other eQTLs in the Framingham data. Our work supports the idea that genetic variants can have broad biological effects.

1084S

Pathway analyses of extreme age-related macular degeneration phenotypes using whole exome sequencing data. *R.J. Sardell[1], W.K. Scott[1], G. Wang[1], W. Cade[1], J.N. Cooke Bailey[2], M.D. Courtenay[1], S.G. Schwartz[3], J.L. Kovach[3], A. Agarwal[4], M.A. Brantley[4], J.L. Haines[2], M.A. Pericak-Vance[1].* 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 3) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 4) Department of Ophthalmology and Visual Sciences, Vanderbilt University School of Medicine, Nashville, TN.

Age-related macular degeneration (AMD) is a leading cause of vision loss among older adults in the US. Although multiple common genetic variants as well as several environmental risk factors for AMD have been identified, a substantial portion of the variance in disease risk and heritability remains unexplained and may reflect rare variants, gene-gene or gene-environment interactions. We used whole exome sequencing of 74 phenotypically extreme individuals to identify rare AMD related risk variants. We hypothesized that these extreme affected AMD individuals are more likely to carry variants of large effect. A genetic risk score was calculated based on 16 common (≥1% minor allele frequency) single nucleotide variants (SNVs) known to be associated with risk of advanced AMD. We then selected 38 individuals with bilateral neovascular AMD, the lowest genetic risk score, and youngest age at examination, and 36 unaffected controls with no or very few small drusen, the highest genetic risk score, and oldest age at examination for sequencing. Variants were annotated for predicted function using SeattleSeq. We used a pathway based analysis approach to combine information on multiple genes within biological pathways using WebGestalt (WEB-based GEne SeT AnaLysis Toolkit). P-values from gene-based tests run using SKAT were used to identify genes of interest for input (those with $p<0.05$). The Kyoto Encyclopedia of Genes and Genomes (KEGG) was used to identify pathways. Genes that contained SNVs used to define the risk score were excluded from analyses. Considering all common and rare SNVs, we did not identify any pathways significantly associated with the extreme phenotype. Considering rare (minor allele frequency <5%), damaging variants only, we identified one enriched pathway that remained significant after Bonferroni correction ($p=0.02$); the Glycosphingolipid biosynthesis lacto and neolacto series. This pathway contained four genes with $p<0.05$ in our original SKAT analysis: *FUT2*, *FUT3*, *FUT7* and *ABO*. Sphingolipid metabolism has been linked to retinal pigment epithelium cell inflammation and death, suggesting this pathway may be relevant to the AMD disease process. Although *FUT3* is nearby the previously identified AMD-associated gene, *C3*, variants at these two loci were not in strong linkage disequilibrium. These initial results highlight a pathway that may be associated with extreme AMD phenotypes and support the validity of this approach to identify new risk genes.

1085M

Exome variants underneath linkage peaks in multiplex Sardinian multiple sclerosis pedigrees implicate genes with roles in autoimmunity and neuroinflammation. *A. Hadjixenofontos[1], L. Foco[2], A. Ticca[4], P. Bitti[5], R. Pastorino[2], L. Bernardinelli[2,3], J.L. McCauley[1].* 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) Department of Public Health, Neurosciences, Experimental and Forensic Medicine, Medical Statistics and Epidemiology Unit, University of Pavia, 27100 Pavia, Italy; 3) Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK; 4) Divisione di Neurologia, Ospedale S. Francesco, Nuoro, Italy; 5) Immunohaematology and Blood Transfusion Department, Ospedale S. Francesco, Nuoro, Italy.

Multiple sclerosis (MS) is an autoimmune, demyelinating disease of the central nervous system that affects 2 million people worldwide. The prevalence of MS in the isolated population of Sardinia, Italy is unusually high (143 to 262 per 100,000). We sequenced the exomes of 31 affected individuals from 5 Sardinian multiplex families and examined the genes implicated by the currently established list of 110 MS-associated SNPs. Two genes (*MANBA* and *IKZF1* at 7p12) in two regions shared identical-by-descent in one family carry missense variants that are predicted to be probably damaging, highly conserved, and that are rare in outbred Europeans, but have a 5 to 10-fold higher frequency in Sardinia. Both of these are present in at least 6 out of 7 affecteds in the same family. In further examining these exomes for rare, missense variants we performed non-parametric linkage analysis in a dataset of 19 multiplex, extended pedigrees with 75 affected individuals, using a linkage map created from the variation tiled on the custom Illumina beadchip dubbed the “Immunochip”. Five regions reached a LOD* > 1, on chromosomes 4, 6, 7, 12 and 18. Multiple genes that were previously implicated in autoimmune diseases and some that were explored as drug targets in experimental autoimmune encephalomyelitis are located in these regions; one example is the serpin cluster on 18q21-22. Interestingly, the 12p13 peak spans the natural killer cell gene complex at that locus. The 7p15-12 peak replicates previously reported linkage peaks from Sardinian datasets. One of the currently established MS-associated loci also overlaps this region. We then performed fine-mapping family-based tests of association using CAPL, in the regions defined by the support intervals of the five linkage peaks. Four regions produced nominally significant ($p\leq0.05$) results. The implicated genes are *JAF1* (7p15), *CD226* (18q22), *EPHA7* (6p21) and *KLRC2* (12p13). Interestingly, the literature reports upregulated expression of both *EPHA7* and *KLRC2* in active MS lesions, while variation in *JAZF1* and *CD226* has previously been implicated in MS. Of particular interest, our exome sequencing data shows that multiple affecteds from all families carry the alternative alleles for many, primarily rare variants in the *CD226* gene. The overall functional themes in our results in this population highlight genes important in natural killer cell activity and transmigration of lymphocytes across the blood-brain barrier.

1086T

Family-Based Linkage Analysis of Coding Variants with Cardiometabolic Traits in the Diabetes Heart Study. *L.M. Raffield[1,2,3], J.N. Hellwege[2,3], A.J. Cox[2,3,4], C.D. Langefeld[5,6], J.J. Carr[7], B.I. Freedman[8], N.D. Palmer[2,3,4,6], D.W. Bowden[2,3,4].* 1) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 5) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 6) Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC; 7) Department of Radiology, Vanderbilt University Medical Center, Nashville, TN; 8) Department of Internal Medicine - Nephrology, Wake Forest School of Medicine, Winston-Salem, NC.

In contrast to Mendelian disorders, family-based linkage analysis has had limited success in identifying genetic variants relevant to complex traits such as cardiovascular disease and type 2 diabetes (T2D). Current efforts to identify high effect, low frequency coding variants have renewed interest in linkage methods. Data from an Illumina Infinium HumanExome BeadChip array, designed to assay less common coding variants, was available in the Diabetes Heart Study (DHS), a study of subclinical cardiovascular disease in T2D-enriched families (T2D prevalence 83.7%). Two-point linkage analysis was performed in 1190 European-American individuals from 468 families (1318 sibpairs). We hypothesized that linkage analysis might highlight variants and regions of interest for cardiometabolic traits, including measures of vascular calcified plaque, dyslipidemia, inflammation, renal function, and adiposity, and could be aligned with single variant association analysis results. Both two-point linkage and association analyses were performed using the Sequential Oligogenic Linkage Analysis Routines (SOLAR) program adjusting for age, sex, and T2D status. Across 26 traits, 174 LOD scores over 3.0 were observed; the highest was located intergenically on chromosome 4 (rs10939444, LOD=5.69, minor allele frequency (MAF)= 0.35) with plasminogen activator inhibitor-1 levels. Notable evidence of linkage was observed for several traits on chromosome 6, including C-reactive protein (max LOD=3.97, rs2395730, MAF=0.43), high-density lipoprotein cholesterol (max LOD=3.53, rs4413654, MAF=0.23), albumin/creatinine ratio (max LOD=3.14, rs927332, MAF=0.48), adiponectin (max LOD=2.94, rs9370418, Ser387Pro in GFRAL, MAF=0.24), and waist circumference (max LOD=3.72, rs1264344, MAF=0.47). These variants were distributed over a 9 to 49 Mb region, depending on the traits included, suggesting a possible pleiotropic influence on related cardiometabolic traits. Linkage and association analyses also highlighted a variant in cadherin 13 (rs72807847, Asn39Ser, MAF= 0.006) with evidence for both association ($p= 6.23 \times 10^{-6}$) and linkage (LOD=2.62) with abdominal aortic calcification; cadherin 13 has previously been associated with blood pressure, adiponectin levels, and myocardial infarction risk. A combined linkage and association analysis approach may be effective for identifying loci, including loci for late onset diseases such as cardiometabolic traits, in sibpair studies without parental data.

1087S

Identification of Juvenile Myoclonic Epilepsy Loci in Chromosomes 16p13.3, 13q31.1 and 4q35.2 in Honduran Families: Linkage to Epilepsy and Encephalography Traits. *Y. Lin[1, 2], J.E. Wight[1, 2], V.H. Nguyen[1, 2], M.T. Medina[2, 4], R.M. Duron[2], C. Patterson[1, 2], M. Tanaka[1, 2, 3], D.S. Bai[1, 2, 3], S. Aftab[1, 2, 3], Y. Molina[2, 4], J.N. Bailey[1, 2, 5], A.V. Delgado-Escueta[1, 2, 3].* 1) Epilepsy Genetics/Genomics Laboratories, VA GLAHS - West Los Angeles, Los Angeles, CA; 2) GENESS International Consortium; 3) Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, CA; 4) School of Medical Sciences, National Autonomous University of Honduras, Tegucigalpa, Honduras; 5) Fielding School of Public Health at UCLA, Los Angeles, CA.

Juvenile myoclonic epilepsy (JME), the most common genetic generalized epilepsy, accounts for 3% to 11% of all epilepsies. Because JME is the most common cause of a life-long burden of devastating grand mal convulsions, researchers search for JME causing genes. We ascertained three multigenerational families from Honduras through a proband diagnosed with JME, and validated the clinical and EEG phenotypes of all family members. After genotyping with 6500 SNPs by CIDR, we performed two-point parametric and multipoint linkage analyses, as well as haplotype analysis. Analyses used four different diagnostic models based on epilepsy diagnoses of family members and presence of clinically asymptomatic members with polyspike wave EEG traits. Analyses were performed for each family and for all three Honduran families pooled together. Two-point parametric linkage analysis assumed autosomal dominant inheritance with 0.7 penetrance, gene frequency of 0.01 and penetrance of the homozygote non-disease allele of 0.001. With three families pooled together, and all persons with clinical epilepsy as well as all asymptomatic persons with EEG spike waves considered affected, Zmax passed threshold for significance with LOD of 4.29 for rs1203974 and 3.77 for rs1211375 ($\theta m=f$ 0.00) in chromosome 16p13.3. A third SNP, rs3127540, in 13q31.1 also reached significance (Zmax=3.72). With asymptomatic family members with EEG polyspike waves not considered affected and only members with epilepsy considered affected, Zmax was 3.20 at rs6553022 ($\theta m=f$ 0.00) in chromosome 4q35.2 Analyses for individual families did not produce Zmax >3.3. Haplotype analyses confirmed segregation in all clinically affected and asymptomatics with EEG polyspike waves in 4q35.2 and 13q31.1 in all three Honduran families, and in two families linked to 16p13.3. Linkage and haplotype analyses identified two linkage areas where epilepsy genes have not been previously found (13q31.1 and 4q35.2) and confirmed a previously determined linkage region in 16p13.3. Whole exome sequencing (WES) narrowed our search to a handful of genes within linkage areas 13q31.1 and 4q35.2 in two of these families. Combining linkage analyses with WES can identify genes that share disease pathways and lead to finding cures.

1088M

Candidate high-penetrance locus for myopia identified on chromosome 7 using linkage and family-based association analyses of exome chip data in 3 U.S. populations. *J.E. Bailey-Wilson[1], C.L. Simpson[1], L. Portas[1,2], F. Murgia[1,2], S.S. Yee[3], E.B. Ciner[4], D.D. Stambolian[3].* 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; 2) Institute of Population Genetics, National Research Council of Italy, Sassari, Italy; 3) Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA; 4) Pennsylvania College of Optometry, Salus University, Elkins Park, PA.

Myopia, the most common cause of visual impairment worldwide, is one of several related refractive error traits and is thought to have both environmental and genetic causes. Current understanding of the genetic basis of myopia, based on genetic linkage and genome-wide association studies suggest that there are many genetic loci that influence variation of mean spherical equivalent (MSE) and risk for myopia (MSE ≤ -1Diopter). Some of these loci appear to have common risk alleles with smaller effects on trait heritability while other loci may have rarer variants with higher penetrance and larger effects on trait heritability. The current study analyzed Exome Chip (Illumina WGHum-ExomePlus) data in extended pedigrees from Ashkenazi Jewish (64 families), Old Order Amish (40 families), and African American (103 families) ethnic groups where multiple individuals were diagnosed with myopia, in a search for higher penetrance susceptibility alleles. This genome-wide chipset includes a set of 4761 GWAS tag SNPs with common minor allele frequencies (MAF), 3468 ancestry informative SNPs, > 230,000 SNVs (some with more rare MAFs) that occur in exons and 30,000 custom SNVs that were chosen to provide denser coverage of chromosomal regions that had been previously reported to show linkage or association to myopia and/or MSE. After extensive quality control procedures, allele frequencies of all variants were estimated from pedigree founders in each ethnic group separately. Parametric linkage and family-based association analyses were performed in each ethnic group separately using the PARAMLINK and FBAT programs, respectively. All three sets of families showed evidence of linkage of myopia to the same region of chromosome 7p (maximum two point LOD scores were 3.38 at 54.9cM, 2.45 at 56.9cM and 4.35 at 44.1cM in the Ashkenazi, Amish and African-American samples). Both the Ashkenazi and Amish families showed evidence of association in this linkage region to variants in the gene *Homo sapiens oxysterol binding protein-like 3*, *OSBPL3*, (different variants in each population, p=2.51E-6 at 42.32cM and 2.76E-6 at 42.18cM respectively). *OSBPL3* encodes a protein that is involved in the regulation of cell adhesion and organization of the actin cytoskeleton. These results give strong evidence of a myopia susceptibility locus on chromosome 7p15. Follow-up of variants in the linkage regions is ongoing to determine which variants best explain the linkage signal in each linked family.

1089T

Linkage-based Analytical Approaches with GWAS Data to Localize Variants Underlying Complex Traits. *N.D. Palmer[1,2], J.N. Hellwege[2], K.D. Taylor[5], J.M. Norris[3], C. Lorenzo[4], J.I. Rotter[5], C.D. Langefeld[6], D.W. Bowden[1,2].* 1) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, CO; 4) Department of Medicine, University of Texas Health Science Center, San Antonio, TX; 5) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA; 6) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC.

In the era of Next Generation Sequencing, novel approaches will be needed to assess the contribution of low frequency, functional variants postulated to explain the missing heritability in common complex disease. We have computed two-point linkage analysis of genome-wide association study (GWAS) and exome chip data in the IRAS Family Study Hispanic American cohort in an effort to map low frequency, high impact variants contributing to novel, diabetes-related phenotypes. This approach reduces the stringent multiple comparison penalty in association testing by highlighting linked genomic regions for focused association and burden testing, with the advantage of being easily aligned. Analysis of 687,095 SNPs across six diabetes-related quantitative traits in 88 families (1034 individuals) revealed 1170 nominal (LOD>3) and 19 significant (LOD>5) signals. Consistent evidence of linkage was observed for acute insulin response (AIR; beta-cell function) on 1q24.3-25.2 (LOD=5.1-6.4). Among the genes implicated in association of common and rare variants was *RALGPS2*/*ANGPTL1* (rs6693283, $P=4.2\times10^{-4}$) although family-based burden testing did not enhance the evidence of association (P=0.20). A more defined interval on 2p13.1 was also linked to AIR (LOD=5.3) encompassing *ALMS1* which underlies Alstrom Syndrome; an autosomal recessive syndrome including insulin resistance, hyperinsulinemia, and type 2 diabetes. Analysis of common and rare variants revealed association with a missense SNP (S2027P, P=0.0042). Family-based burden testing weighted for rare variants more significantly supported these results ($P=1.5\times10^{-7}$). Among other phenotypes, insulin sensitivity (S_I) was linked to 7q11.22 (LOD=5.1). Regional evaluation of candidates in the interval using association analysis revealed signals in *CALN1*, a member of the calmodulin family. Association analysis revealed common variants associated with S_I (rs844762, P=0.0091) while only a single rare variant was captured by the exome chip. Family-based burden testing at this locus revealed nominal association (P=0.048) likely due to down weighting of common variants. Family-based linkage is a powerful approach in contemporary array-based genetic datasets to prioritize genomic intervals for association analysis and burden testing. Whole genome sequence data should add to the power of this approach. Thus, a combined linkage and association testing framework using high density genetic data has the potential to reveal novel loci.

1090S

Meta-analysis of birth weight genome-wide association studies identifies five novel loci extending links between early growth and adult metabolic diseases. *M. Horikoshi[1,22], M.N. Kooijman[2], J.P. Bradfield[3], D.P. Strachan[4], N. Grarup[5], N. Pitkänen[6], L-P. Lyytikäinen[6], L. Carstensen[7], N.V. Tejedor[8], R. Rueedi[9], E. Kreiner-Møller[10], E. Thiering[11], W. Ang[12], P.K. Joshi[13], E.T.H. Boh[14], I. Ntalla[15], E.M. van Leeuwen[16], V. Lindi[17], R. Joro[17], D.L. Cousminer[18], Y. Wu[19], T.S. Ahluwalia[5], V. Lagou[1], A. Mahajan[1], S.F.A. Grant[3], I. Prokopenko[20], R.M. Freathy[21], M.I. McCarthy[1,22], A.P. Morris[1,23], EGG Consortium.* 1) The Wellcome Trust Centre for Human Genetics, Oxford, UK; 2) Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; 3) Center for Applied Genomics, The Children's Hospital of Philadelphia, PA; 4) Population Health Research Institute, St George's, University of London, London, UK; 5) The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 6) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 7) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 8) Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain; 9) Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois (CHUV) University Hospital, Lausanne, Switzerland; 10) COPSAC and the Danish Pediatric Asthma Center, Copenhagen University Hospital, Copenhagen, Denmark; 11) Helmholtz Zentrum Muenchen, Munich, Germany; 12) School of Women's and Infants' Health, The University of Western Australia, Perth, Australia; 13) Centre for Population Health Sciences, University of Edinburgh, Edingburgh, UK; 14) Life Science Institutes, National University of Singapore, Singapore, Singapore; 15) University of Leicester, Leicester, UK; 16) Dept of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands; 17) Department of Physiology, Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland; 18) Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland; 19) Department of Genetics, University of North Carolina, Chapel Hill, NC; 20) Department of Genomics of Common Disease, School of Public Health, Imperial College London, London, UK; 21) University of Exeter Medical School, Royal Devon and Exeter Hospital, Exeter, UK; 22) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, UK; 23) Department of Biostatistics, University of Liverpool, Liverpool, UK.

Lower birth weight (BW) is associated with increased risk of future type 2 diabetes (T2D) and cardiovascular disease. Based on HapMap 2 imputed European genome-wide association (GWA) studies, we previously reported 7 loci associated with BW, of which two (*ADCY5*, *CDKAL1*) have been implicated in T2D and one (*ADRB1*) in hypertension. Here we report analyses based on an increased sample size (55,729 full term singletons) from multiethnic groups (22 European and 6 non-European of African-Americans (AA), Philippines, Moroccan, Surinamese, Turkish and Chinese descent) and imputation up to 20.8M SNPs from the more dense 1000 Genomes Project reference panel. Using inverse-variance fixed-effects meta-analysis, GWA summary statistics between standardized sex-specific Z-scores of BW (gestational week adjusted) and each SNP were combined across studies in sex-combined and sex-stratified manner. We detected five novel loci at genome-wide significance: near *EPAS1* ($p=2.4\times10^{-9}$), *YKT6* ($p=1.2\times10^{-8}$), *SREBF2/CENPM* ($p=3.5\times10^{-8}$) and *MAFB* ($p=4.1\times10^{-8}$) from sex-combined meta-analysis and *KIAA0907* (pmale=0.22, pfemale=2.8×10^{-8}, pheterogeneity=0.0023) from sex-stratified meta-analysis. There was no heterogeneity between ethnicities at any of the novel loci (Cochran's Q $p > 0.05$). All loci were common variant loci except for *YKT6* (MAF EUR:1%, AA:0.2%). Other variants in the *MAFB* locus have been implicated in hyperlipidemia. We then fine-mapped established and novel loci by constructing credible sets of variants with 99% overall posterior probability of being causal. The 99% credible sets included fewer than 20 SNPs at 7 loci. At *ADRB1*, the credible set consisted of just 5 variants, including missense G389R (posterior probability 0.13, r^2 with index SNP in EUR:0.85, AA:0.47 and East Asian:0.95). Approximate conditional analysis showed that *ADRB1* signal could be explained by G389R (*ADRB1* index SNP $p_{unconditional}=6.1\times10^{-10}$, $p_{conditional\ on\ G389R}=0.22$). Collectively, we have extended the number of BW associated signals from 7 to 12 at both common and low-frequency loci, of which 4 loci provide genetic links between BW and T2D, hypertension and hyperlipidemia, highlighting complex non-linear relationships between genetic variation, early growth and later metabolic disease.

1091M

Runs of homozygosity reveal inbreeding depression on cognitive function and stature. *P.K. Joshi[1], T. Esko[2], H. Mattsson[3,4], N. Eklund[3], I. Gandin[5], A.U. Jackson[6], T. Nutile[7], C. Schurmann[8], O. Polasek[1], J.F. Wilson[1,9], ROHgen consortium.* 1) Centre for Population Health Sciences, University of Edinburgh, Edinburgh, United Kingdom; 2) Estonian Genome Centre, University of Tartu, Tartu, Estonia; Boston Children's Hospital and Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki, Finland; 4) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 5) Department of Medical Sciences, University of Trieste, Italy; 6) Department of Medical Sciences, University of Trieste, Italy; 6) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA; 7) Institute of Genetics and Biophysics 'Adriano-Buzzati Traverso', CNR, Naples, Italy; 8) The Genetics of Obesity and Related Metabolic Traits Program, The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 9) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, EH4 2XU, Scotland.

The effect of inbreeding on fitness traits in plants and animal populations has been studied since the time of Darwin. Close inbreeding is also associated with increased risk of Mendelian diseases in humans, however the effect of the distant inbreeding found in present day populations on complex traits is less well understood. Our study analysed 16 health-related quantitative traits in up to 350,000 individuals from multiple continental ancestries in 102 cohorts. Within each ethnically homogeneous sub-cohort, we modelled the effect on each trait of increased inbreeding as measured by the sum of lengths of runs of homozygosity in base pairs (SROH), with covariates of age, sex, and at least 3 principal components of genetic variation. We then used fixed effect inverse-variance meta-analysis to combine results. We found that SROH is associated with reduced stature (as measured by height and reflected in a measure of lung function), and reduced cognitive function (as measured by educational attainment and the general cognitive factor, g). In all four cases the effect sizes were small (-0.00075, -0.00112, -0.00126, -0.00160 phenotypic standard deviations per megabase of SROH, respectively), but statistically very significant (Nominal p <1e-16, 5.2e-6, 2.5e-12, 4.7e-9 respectively, with 16 traits and only one genetic effect tested). These effect sizes correspond to a decrease in height of 0.0068 cm and decrease of 0.0043 years in education for each 1Mb increase in SROH. Offspring of first cousins are thus predicted to be 1.1 cm shorter and complete 8 months less education. Contrary to earlier reports in much smaller samples, no evidence of an association was seen for the burden of homozygosity on blood pressure traits or LDL cholesterol, or on other cardiovascular risk factors tested (classical lipid, glycaemic and adiposity traits). Even for the traits where we observed an effect, we estimate the proportion of phenotypic variance explained in an outbred European population to be less than 0.001%. We therefore consider it unlikely that population cardiovascular health or the Flynn effect can be materially explained by background inbreeding or secular changes in the level of outbreeding. Inbreeding depression is predicted to arise in evolutionary fitness-related traits. Thus, our study suggests that increased stature and cognitive function may have been positively selected in human evolution, while many risk factors for late onset complex diseases have not.

1092T

Population-wide linkage screen for successful aging in the Amish. *J.E. Hicks[1], J.R. Gilbert[1], L. Caywood[1], L. Reinhart-Mercer[1], D. Fuzzell[2], R. Laux[2], M.L. Cuccaro[1], M.A. Pericak-Vance[1], J.L. Haines[2], W.K. Scott[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, FL; 2) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland OH.

As the population of industrialized nations ages, the health problems associated with aging will become a major public health concern. The Amish of Indiana and Ohio provide an ideal population to study aging because of their geographic stability and cultural isolation that minimizes environmental variability. The pedigree of the Amish is too large and complex for conventional linkage (e.g. lod score) analysis. For linkage analysis in the Amish, the pedigree must be cut into smaller pedigrees, which may result in false negative results. An alternative approach is to use long range chromosomal phasing (LRP), which allows the detection of segments shared identical-by-descent (IBD) derived from dense SNP genotypes; this avoids the computational complexity of conventional linkage analysis approaches. Population-based linkage using LRP was assessed for successful aging (SA), a composite phenotype consisting of lack of cognitive impairment, lack of depression (as determined by the Geriatric Depression Scale), high performance on the EPESE lower extremity performance battery, and unimpaired performance on four scales: Activities of Daily Living, Instrumental Activities of Daily Living, Nagi (a test of musculoskeletal function), and Rosow-Breslau (a test of lower-extremity function). 74 individuals met all of these criteria and formed primarily cousin-pair relationships. Amish participants (n=813) were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. Missing genotypes were imputed in BEAGLE, and IBD segments of at least 250kb were identified using the GERMLINE unphased algorithm. The configuration of these IBD states was quantified for each as the proportion of affected pairs IBD for the marker. Statistical significance for each marker was determined by repeated sampling without replacement of an equivalent number of individuals from the population and assessing their IBD configuration. The p-value was the proportion of samples where the IBD configuration was as or more extreme than that of the successfully aged individuals. Suggestive evidence of linkage ($p<1.7\times10\text{-}3$; Lander and Kruglyak 1995) was detected on chromosomes 4, 6, 9, and 16. The interval providing the most extreme evidence of linkage ($p=1.2\times10\text{-}4$, hg18 coordinates: chr6:109,186,957-109,761,180) was shared by 12.8% of SA pairs and contained the single gene SESN1, which involved in the cellular response to oxidative stress, making it an excellent candidate gene in the aging process.

1093S

Simple Linkage-Based Methods to Identify Cardiometabolic Risk in Families. *J.N. Hellwege[1,2], L.M. Raffield[1,2,3], N.D. Palmer[1,2,4,5], A.J. Cox[1,2,4], J.M. Norris[6], C. Lorenzo[7], Y.-D.I. Chen[8], J.I. Rotter[8], C.D. Langefeld[5,9], B.I. Freedman[5,10], D.W. Bowden[1,2,4].* 1) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston Salem, NC; 2) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 5) Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC; 6) Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Denver, CO; 7) Department of Medicine, University of Texas Health Science Center, San Antonio, TX; 8) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA, Torrance, CA; 9) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 10) Department of Internal Medicine - Nephrology, Wake Forest School of Medicine, Winston-Salem, NC.

Linkage-based approaches in family studies were abandoned by investigators with the advent of genome-wide association studies (GWAS) due to diminished enthusiasm or lost funding. With contemporary genetic datasets of high density arrays coupled with imputations, it seems relevant to revisit linkage methods. Family-based cohorts vary dramatically in size and complexity from very large families through nuclear families and sib pairs. We evaluated cardiometabolic traits using conventional two-point linkage analysis from exome chip data in combination with genetic association data with potentially interesting implications. Two-point analysis has the advantage of being easily aligned with simple association, and meta-analysis is equally straightforward by simple addition of LODs (assuming $\theta=0$). Two-point linkage analysis was performed separately in three diverse cohorts (African Americans (n=596) and Hispanics (n=1414) from the Insulin Resistance Atherosclerosis Family Study (large pedigrees, average family size 14-15 individuals) and European Americans (n=1190) from sibships from the Diabetes Heart Study) for 24 traits using SOLAR. SNPs present in all three cohorts were collated (n=46,420), and linkage meta-analysis performed for each trait by adding the two-point LOD scores across all sample groups to identify variants with substantial contributions to evidence of linkage by all three cohorts. LOD_{cohort} scores up to 5.69 were observed. Combining all three LOD scores for overlapping SNPs across all traits resulted in a total of 502 LOD_{sum} scores greater than 3, including 59 which were greater than 4. Of these, 282 were contributed to by more than one cohort ($LOD_{cohort}<3$). Just one SNP, rs2398162, had a LOD_{cohort} score greater than 1 with adiponectin levels in all three cohorts ($LOD_{sum}=4.27$). The maximal LOD_{sum} score with no LOD_{cohort} >3 was 5.20: rs1073525 (*MARCH8*) with plasminogen activator inhibitor-1 levels. This LOD score was contributed to by LOD_{cohort} scores of up to 2.63. Further evidence of multi-cohort contribution to linkage was seen with rs11708014 and HDL levels (*MCF2L2*; $LOD_{sum}=4.99$), rs250853 with C-reactive protein levels *(FSTL4*; $LOD_{sum}=4.72$), and rs8068318 with waist circumference (*TBX2*; $LOD_{sum}=4.66$), among 9 additional LOD_{sum} scores >4 with no LOD_{cohort} >3. Overall, these SNPs and regions are relevant for functional follow-up due to consistent support for linkage across multiple ethnic groups and with a variety of cardiometabolic phenotypes.

1094M

Linkage and association mapping for osteoarthritis progression in the Genetics of Generalized Osteoarthritis Study. *M.S. Yau[1], L.M. Yerges-Armstrong[1], J.R. O'Connell[1], R.D. Jackson[2], M.C. Hochberg[1], B.D. Mitchell[1], V.B. Kraus[3].* 1) University of Maryland, Baltimore, MD; 2) The Ohio State University, Columbus, OH; 3) Duke University, Durham, NC.

Osteoarthritis (OA) is the most common form of arthritis and is a major cause of morbidity. To identify genetic risk factors associated with structural knee OA progression, we undertook a genome-wide linkage scan using data from the longitudinal component of the Genetics of Generalized Osteoarthritis (GOGO) study (N=1,245, 79% women). We defined knee OA progression as worsening of Kellgren-Lawrence (KL) grade, osteophyte grade, or joint space narrowing grade in individuals with radiographic evidence of OA (KL grade=2 or 3) at baseline (N=703). Individuals who progressed to total joint replacement were also considered cases. We have previously reported linkage of OA progression to two chromosomal regions, Xq21-25 (LOD=4.3, P<0.01) and 18q21 (LOD=2.0, P<0.01). To fine map these regions, we conducted association analyses using markers genotyped on the Illumina 550K platform and replicated our most significant associations in the Osteoarthritis Initiative (OAI), a multi-center natural history study of knee OA (N=1,682, 58% women). Within the X chromosome linkage region, the most significantly associated SNP was rs437033 (OR=1.6, P=9.6E-5), which is located in *SLC25A43*. This did not replicate in the OAI, although the most significant SNP across the chromosome (rs6528082; OR=1.8, P= 9.5E-5) located on Xp22 between *SMS* and *PHEX* showed nominal association with OA progression in the OAI (OR=1.5, P=0.03). Within the chromosome 18 linkage region, the most significantly associated SNP was rs470497 (OR=1.6, P=6.1E-4), located near *CDH7*. This SNP did not replicate in the OAI, although the most significant SNP across the chromosome on 18q12 was rs660959 (OR=1.7, P=5.8E-5), which is closest to *GAREM*. Interestingly, though not in LD with rs660959, another SNP in this gene, rs76800900, was one of the most highly associated SNPs with OA progression in the OAI (OR=0.2, P=3.2E-7). *GAREM* encodes a GRB2 associated regulator of MAPK1 that is involved in the epidermal growth factor receptor-mediated signaling pathway. In summary, while our association studies did not identify any replicated SNPs that could potentially explain the linkage to OA progression, they did provide an interesting lead for a novel OA progression gene on chromosome 18, i.e., *GAREM*. Further bioinformatics analyses will be needed to determine the contribution of these association SNPs to the linkage regions and the specific role that these genetic markers may play in the pathogenesis of OA progression.

1095T

Using monogenic phenotypes to identify mechanisms of GWAS variants associated with insulin resistance. *H. Yaghootkar[1], R.A. Scott[2], C.C. White[3,4], W. Zhang[5], E. Speliotes[6], P.B. Munroe[7], G.B. Ehret[8,9], J.C. Bis[10], C.S. Fox[3,11,12], M. Walker[13], I.B. Borecki[14], J.W. Knowles[15], L. Yerges-Armstrong[16], C. Ohlsson[17], J.R.B. Perry[2], J.C. Chambers[5], J.S. Kooner[18], N. Franceschini[19], C. Langenberg[20,21], M.F. Hivert[22,23], Z. Dastani[24], J.B. Richards[25,26], R.K. Semple[27,28], T.M. Frayling[1].* 1) University of Exeter Medical School, Exeter, Devon, UK; 2) MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, UK; 3) Framingham Heart Study, National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health, Framingham, Massachusetts, USA; 4) Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA; 5) Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK; 6) Department of Internal Medicine, Division of Gastro enterology and Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 7) Clinical Pharmacology and Barts and The London Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 8) Cardiology Center, Geneva University Hospital, Geneva, Switzerland; 9) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 10) Cardiovascular Health Research Unit and Department of Medicine, University of Washington, Seattle, Washington, USA; 11) Center for Population Studies, National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health, Framingham, Massachusetts, USA; 12) Division of Endocrinology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; 13) Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK; 14) Department of Genetics, Washington University School of Medicine, St Louis, MO; 15) Department of Medicine and Cardiovascular Institute, Stanford University School of Medicine, Stanford, California, USA; 16) University of Maryland School of Medicine, Division of Endocrinology, Baltimore, Maryland, USA; 17) Center for Bone and Arthritis Research, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 18) Cardiovascular Science, National Heart & Lung Institute, Imperial College London, London, UK; 19) Department of Epidemiology, University of North Carolina Chapel Hill, Chapel Hill, North Carolina, USA; 20) Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; 21) Department of Epidemiology and Public Health, University College London, London, UK; 22) Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical School, Boston, MA, USA; 23) General Medicine Division, Massachusetts General Hospital, Boston, MA, USA; 24) Departments of Human Genetics and Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada; 25) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 26) Department of Medicine, Human Genetics, Epidemiology and Biostatistics, McGill University, Montreal, Canada; 27) The National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge, UK; 28) The University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Cambridge, UK.

Insulin resistance is linked to high risk of type 2 diabetes, hypertension and coronary artery disease but the underlying mechanisms are poorly understood. Studies of monogenic forms of insulin resistance, however, established mechanistic paradigms of potential relevance to the common form of the disease. Genome wide association studies (GWAS) have found 19 common genetic variants associated with fasting insulin based measures of insulin resistance but the variants have subtle effects and most have unclear mechanisms. We aimed to combine the distinct characteristics of monogenic forms of insulin resistance and findings of GWAS to get mechanistic insights. We used results from GWAS of 8 metabolic traits that can distinguish between different forms of monogenic insulin resistance and used hierarchical clustering to group the 19 insulin resistance variants. We analysed the effect of each group against type 2 diabetes (12,171 cases), coronary artery disease (40,365 cases) and blood pressure (69,828 individuals). Hierarchical clustering identified a group of 11 variants among 19 insulin resistance variants associated with a metabolic profile resembling rare form of insulin resistance secondary to lack of adipose tissue under the skin - lipodystrophy. Collectively the 11 insulin resistance risk alleles were associated with higher triglycerides (β=0.018; p=4×10^{-29}), lower HDL-C (β=-0.020; p=7×10^{-37}), greater hepatic steatosis (β=0.021; p=3×10^{-4}), higher alanine transaminase (β=0.002; p=3×10^{-5}), lower SHBG (β=-0.010; p=9×10^{-13}) and lower adiponectin (β=-0.015; p=2×10^{-26}). The same risk alleles were associated with lower BMI (per-allele β=-0.008; p=7×10^{-8}), and increased visceral-to-subcutaneous adipose tissue ratio (β=-0.015; p=6×10^{-7}). Individuals carrying >= 17 fasting insulin raising alleles were slimmer (0.30 kgm-2) but at increased risk of type 2 diabetes (odds ratio [OR] 1.46, per-allele p=5×10^{-13}), coronary artery disease (OR 1.12, per-allele p=1×10^{-5}), and increased blood pressure (systolic and diastolic blood pressure of 1.21 mmHg (per-allele p=2×10^{-5}), and 0.67 mmHg (per-allele p=2×10-4), respectively, compared to individuals carrying <=9 risk alleles. Our results provide genetic evidence for an association between reduced fat under the skin and the risk of insulin resistance and metabolic traits. Our study gives an example of how monogenic studies can complement GWAS to understand mechanisms underlying complex traits/diseases.

1096S

Copy number of the salivary amylase gene *AMY1* modulates serum amylase levels and is associated with the metabolic profile. *J.S. El-Sayed Moustafa[1], T. Martin[2], M. Beaumont[2], L. Yengo[3,4,5], P. Maboudou[6], C. Menni[2], T. Brousseau[6], T.D. Spector[2], P. Froguel[1,3,4,5], M. Falchi[1,2].* 1) Department of Genomics of Common Disease, Imperial College London, London, United Kingdom; 2) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 3) CNRS UMR 8199, Lille Pasteur Institute, Lille, France; 4) Lille 2 University, Lille, France; 5) European Genomics Institute for Diabetes (EGID), Lille, France; 6) CHRU de Lille, Pôle de Biologie-Pathologie-Génétique, UF Biochimie automatisée, Lille, France.

We have recently shown low *AMY1* copy number to be strongly associated with increased body mass index (BMI) and risk of obesity (Falchi *et al*, *Nature Genetics*, 2014), suggesting that salivary amylase copy number may play an important role in energy balance and metabolism. We have measured both *AMY1* genomic copy number and salivary serum amylase levels in 219 female twins from the UK TwinsUK cohort. Salivary amylase levels were highly heritable (h^2 = 0.81) and were strongly correlated with *AMY1* copy number (R^2 = 0.32; $P < 2.2 \times 10^{-16}$). BMI was negatively associated with salivary amylase levels (β = -0.062; $P = 4.15 \times 10^{-3}$) and also inversely associated with fasting insulin levels (β = -0.08; P = 0.02). Association analysis with blood metabolomic profiles generated using mass spectrometry in the TwinsUK sample suggest that genetically-determined low salivary amylase levels are accompanied by dysregulation of lipid and glucose metabolism. Alterations observed in lipid metabolism, bile acids and amino acid catabolism may also reflect changes in the gut microbiome. Salivary amylase is a digestive enzyme involved in the metabolism of dietary starch. Our analysis of serum amylase levels suggests low salivary amylase levels to be a genetic trait associated with increased risk of metabolic abnormalities. In order to further investigate its role in human metabolism, we are currently expanding our sample size by measuring serum amylase levels in a further 1,500 subjects from the TwinsUK cohort as well as over 2,000 subjects from the French DESIR cohort for whom *AMY1* copy number measurements and metabolomics data are also available.

1097M

Accurate molecular prediction in inflammatory bowel disease. *H. Huang[1,2], J.B. Essers[3], B.M. Neale[1,2], S. Ripke[1,2], M. Vu[4], R. Xavier[5], A. Ananthakrishnan[5], P. Fleshner[6], M. Dubinsky[7], T. Harituniains[6], S.R. Targan[6], D.P.B. McGovern[6], M.J. Daly[1,2].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA; 3) Inflammatory Bowel Disease Center, Children's Hospital Boston, Boston, MA; 4) Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA; 5) Gastrointestinal Unit, Massachusetts General Hospital, Boston, MA; 6) F. Widjaja Family Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 7) Department of Pediatrics, Cedars-Sinai Medical Center, Los Angeles, CA.

Inflammatory bowel diseases (IBD) are chronic inflammatory diseases of the gastrointestinal system. Ulcerative colitis (UC) and Crohn's disease (CD) are two major types of IBD. While some IBD therapies are universally effective, others show greater clinical efficacy in either CD or UC. Accurate classification of IBD can lead to optimal treatment choices and reduce disease complications and relapses. Due to the heterogeneous presentation of IBD, diagnosis can be challenging in up to 20% of patients using the conventional classification. Here we propose a molecular based risk model as a novel clinical tool to discriminate CD from UC by aggregating genetics, serology and tobacco smoking information.

The genetics component includes polygenic risk scores for CD and UC using independently associated SNPs from 163 IBD loci [Jostins et al]. IBD-associated serum immune responses to ASCA, ANCA, anti-Cbir1, anti-OMPC and anti-I2 were generated by ELISA. Smoking status at diagnosis was uniformly collected and categorized as never, ex-smoker and current. The study cohort, including 1363 CD and 501 UC patients, was split evenly into training and validation groups. The overall risk model achieved an area under the curve (AUC) of 0.80 for discriminating CD from UC. Individually: genetics AUC=0.72; serology AUC=0.79; and smoking AUC=0.55. This genetics model outperformed the genetics component of the commercially available IBD sgi diagnostic by PROMETHEUS (AUC=0.55). Serology alone showed the best AUC but is uninformative in over half of patients. Importantly genetics performs similarly well when serology is uninformative (AUC=0.69), establishing that genetics and serology provide independent information. We identified 42 patients who developed 'denovo' CD after surgery with a UC diagnosis. We assigned our CD risk scores to these patients and correctly classified 26 (62%) of them. We achieved the same success in other cohorts using only the genetics component. Out of 31 patients whose diagnoses changed from UC to CD, our model gave correct diagnoses to 24 (77%). In addition to diagnosis, the risk score from our model was also associated with a younger age of onset in CD (p=3E-12).

In conclusion, we have shown that our molecular-based model combining genetics, serology and smoking information could complement current diagnostic strategies and is particularly useful when conventional diagnosis of IBD is challenging.

1098T

Improving the Power of Genetic Association Tests with Imperfect Phenotype Derived from Electronic Medical Records. *J.A. Sinnott[1], W. Dai[1], K.P. Liao[2], S.Y. Shaw[3], A.N. Ananthakrishnan[4], V.S. Gainer[5], E.W. Karlson[2], S. Churchill[6], P. Szolovits[7], S. Murphy[5,8], I. Kohane[6,9], R. Plenge[10], T. Cai[1].* 1) Biostatistics, Harvard School of Public Health, Boston, MA; 2) Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, MA; 3) Center for Systems Biology, Massachusetts General Hospital, Boston, MA; 4) Division of Gastroenterology, Massachusetts General Hospital, Boston, MA; 5) Research Computing, Partners Healthcare, Charlestown, MA; 6) i2b2 National Center for Biomedical Computing, Boston, MA; 7) Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA; 8) Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA; 9) Center for Biomedical Informatics, Harvard Medical School, Boston, MA; 10) Merck Research Laboratories, Boston, MA.

To reduce costs and improve clinical relevance of genetic studies, there has been increasing interest in performing genetic studies in hospital-based cohorts by linking phenotypes extracted from electronic medical records (EMRs) to genotypes assessed in routinely collected medical samples. A fundamental difficulty in implementing such studies is extracting accurate information about disease outcomes from large numbers of EMRs. Recently, many algorithms have been developed to infer phenotypes from EMRs, and although these algorithms are quite accurate, they typically do not provide perfect classification due to the difficulty in inferring meaning from text. Some algorithms produce for each patient a probability that the patient is a disease case. This probability can be thresholded to define case-control status, and this estimated case-control status has been used to replicate known genetic associations in EMR-based studies. However, using this dichotomized outcome (denoted DO) in place of the true disease status results in outcome misclassification, which can diminish test power and bias odds ratio (OR) estimates. We propose to instead directly model the algorithm-derived probability of being a case. We demonstrate how our approach (denoted p_D) improves test power and OR estimation in simulation studies: improvements in test power are small but consistent, but improvements in odds ratio estimation are dramatic (e.g., a bias of 34% using DO vs. 3% using p_D). We also demonstrate our method's performance in an EMR-based study of rheumatoid arthritis (RA). For some SNPs, our approach recovers the OR observed in a recent meta-analysis: e.g., for rs6457620 in *HLA*, the meta-analysis OR is 2.35 and the p_D OR is 2.28, while the DO OR is 2.03. For other SNPs (e.g., rs6679677 in *PTPN22*), both EMR-based estimates differ from the meta-analysis estimate. We discuss possible causes for these discrepancies, which may impact other EMR studies: our cases are likely to have more severe disease than a random sample of RA cases, while our controls are likely to have more comorbidities than typical "healthy controls." We also provide power and sample size calculations in terms of algorithm accuracy measures for use in planning future studies. Our work provides an easily implemented solution to a major practical challenge that arises when using EMR data, which can facilitate the use of EMR infrastructure for more powerful, cost-effective, and diverse genetic studies.

1099S

Use of diverse electronic medical record systems for a genomewide association study of colonic diverticular disease in European-ancestry populations. *M.G. Hayes[1], L.L. Armstrong[1], W.K. Thompson[1], J.A. Pacheco[1], M. Brilliant[2], P.L. Peissig[2], M. de Andrade[3], S.J. Bielinski[3], J. Pathak[3], I.J. Kullo[3], D.C. Crawford[4], J.C. Denny[4], G. Tromp[5], K.M. Borthwick[5], M.J. Shellenberger[5], H. Kuivaniemi[5], D.S. Carrell[6], G.P. Jarvik[7], D.R. Crosslin[7], M.D. Ritchie[8], M.E. Smith[1], R.L. Chisholm[1], A.N. Kho[1].* 1) Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Marshfield Clinic, Marshfield, WI; 3) Mayo Clinic, Rochester, MN; 4) Vanderbilt University, Nashville, TN; 5) Geisinger Health System, Danville, PA; 6) Group Health Research Institute, Center for Health Studies, Seattle, WA; 7) University of Washington, Seattle, WA; 8) Penn State University, State College, PA.

Diverticulosis is defined by small outpouchings formed through weakening of the colonic mucosal wall muscles that when inflamed (diverticulitis) can cause significant abdominal pain and morbidity. As part of the electronic Medical Records and Genomics (eMERGE) network, we identified diverticulitis and diverticulosis cases and controls for GWAS using data captured through routine clinical care across six institutions with different electronic medical record systems. All participants had to have an abdominal CT scan or a colonoscopy to be included. Using natural language processing of these procedure reports, cases had a positive mention of diverticula, diverticulosis or diverticulitis; while controls had ≥1 colonoscopy and no mention of any type of diverticula. This approach yielded an 88-100% positive predictive value and an 89-100% negative predictive value for identification of diverticular cases and controls compared against clinician review when validated across three eMERGE sites. We identified 5,285 cases and 3,761 controls for diverticulosis and 569 cases and 8,477 controls for diverticulitis, all of European ancestry. Previously genotyped data associated with these participants were collaboratively quality-controlled, merged, and imputed to 1000Genomes reference across all six institutions. Logistic regression analyses were performed for case-control status under an additive model, adjusting for age, sex, BMI, smoking status, study site, and the first two principal components for ancestry from the genotype data. For diverticulosis, we observed a robust association signal spanning *CAV1*/*CAV2* on 7q31.2 lead by rs76633992 that nearly reached genomewide significance ($p=1.2\times10^{-7}$). Another association signal approaching genomewide significance was observed (chr3:168780737:d; $p=1.7\times10^{-6}$) near *EVI1*. For diverticulitis we observed a genome-wide significant association with rs148547269 ($p=4.4\times10^{-8}$) on 12q22. This variant occurs ~20 kb upstream of *CLLU1*, in a highly conserved region containing a CTCF zinc finger transcription factor binding site. Multiple association signals approaching genome-wide significance were also observed including rs28646272 ($p=1.1\times10^{-6}$) in *PALLD* on 4q32.3, rs4584511 ($p=1.4\times10^{-6}$) in *HPS1* on 10q24.2, and rs12347739 ($p=1.7\times10^{-6}$) in *DENND4C* on 9p22.1. Our results identify a number of novel associations for follow-up fine-mapping in order to elucidate the genetic architecture of diverticular disease in European ancestry populations.

1100M

Comparative Analysis of Electronic Health Record (EHR)-driven and Conventional Cohort-driven Genomic Research. *V. Thaker[1], T. Lingren[2], B. Namjou[2], C. Perry[1], N. Crimmins[2], C. Brady[2], I. Solti[2], G. Savova[1], I. Kohane[1], J. Harley[2].* 1) Boston Childrens Hospital, Boston, MA; 2) Cincinnati Childrens Hospital and Medical Center, Cincinnati, OH.

BACKGROUND: Cost-effective and time-efficient systems to perform genetic research require phenotypic data extraction from EHRs. Within the eMERGE network (electronic MEdical Records and GEnomics, an NHGRI funded initiative), Boston Children's Hospital (BCH) and Cincinnati Children's Hospital Medical Center (CCHMC) collaborated to develop an algorithm for Early Childhood Obesity. We describe the process of case selection, gaps identified, logistical and cost analysis of potential solutions. METHODS: The phenotypic identification of cases and controls was performed using a validated algorithm with structured and unstructured data using Natural Language Processing (NLP). Biorepository databases were searched for identified cases and controls. New systems targeted for biological sample collection were defined to bridge gaps in availability. BCH established the conventional method of cohort-driven biorepository. CCHMC used EHRs for subject recruitment. RESULTS: Both institutions are recruiting subjects at weight management and primary care clinics. BCH is recruiting a cohort for family study including probands and first-degree relatives. Using a study algorithm, 194 unique study subjects have been approached with an enrollment of 35 (18%). The IRB approval required 7 months. The cost of recruitment of each subject is ~$400.00 including permission from physician and patients, enrollment, sample collection at Clinical Study Unit. After IRB approval (12 months), CCHMC is recruiting new cases using EHRs. Clinical research coordinators evaluate patient data with the algorithm and flag cases. Flagged patients are approached during the visit for consent. Consented patients provide saliva for DNA extraction. The cost of recruitment is ~$53 per subject including screening, approach and enrollment. Current consent rates (73.9%, 17/23) indicate support of previous claims made about the advantages of EHR-based recruitment.1 CONCLUSION: The selection of a common phenotype does not guarantee availability of DNA samples for genetic studies. For genomic studies using EHR, the burden of development of a phenotype algorithm, identifying cases, collecting samples, satisfying regulatory considerations can be substantial. Augmentation of institutional biorepositories to enhance cost effective genomic research using EHR is recommended. 1 Kohane, Isaac S. "Using electronic health records to drive discovery in disease genomics." Nature Reviews Genetics 12.6 (2011): 417-428.

1101T

UK BiLEVE, the first genetic study in UK Biobank, identifies novel regions associated with airway obstruction phenotypes using a custom genome-wide array in 50,000 individuals. *L.V. Wain[1], N. Shrine[1], I. Ntalla[1], J.P. Cook[1], A.P. Morris[2], E. Zeggini[3], J. Marchini[4,5], P. Deloukas[6], A. Hansell[7], R. Hubbard[8], I. Pavord[9], N.C. Thomson[10], D.P. Strachan[11], I.P. Hall[12], M.D. Tobin[1,13].* 1) Department of Health Sciences, University of Leicester, Leicester, United Kingdom; 2) Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom; 3) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom; 4) Department of Statistics, University of Oxford, Oxford, United Kingdom; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 6) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University London, London, United Kingdom; 7) Faculty of Medicine, School of Public Health, Imperial College London, London, United Kingdom; 8) Faculty of Medicine and Health Sciences, School of Medicine, University of Nottingham, Nottingham, United Kingdom; 9) Respiratory Medicine, University of Oxford, Oxford, United Kingdom; 10) Institute of Infection, Immunity and Inflammation, University of Glasgow, United Kingdom; 11) Population Health Research Institute, St George's University of London, London, United Kingdom; 12) Division of Therapeutics and Molecular Medicine, University of Nottingham, Nottingham, United Kingdom; 13) National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, United Kingdom.

UK Biobank recruited 500,000 UK individuals aged 40-69 years (95% European ancestry), collecting detailed health and lifestyle information, DNA, and physical measures, including spirometry. In the UK Biobank Lung Exome Variant Evaluation (UK BiLEVE) we undertook nested case-control studies in European ancestry UK Biobank subjects with high-quality spirometry measures. Among heavy smokers, we selected groups defined by low percent predicted forced expiratory volume in 1 second (%FEV_1, n=10,000), normal %FEV_1 (n=10,000) and high %FEV_1 (n=5,000). Similarly, among never smokers, we sampled low %FEV_1 (n=10,000), normal %FEV_1 (n=10,000) and high %FEV_1 (n=5,000) groups. Genome-wide genotyping was undertaken using a custom designed Affymetrix array that included 130K rare missense and loss of function variants (selected to be polymorphic in UK populations based on currently available "exome chip" data), 642K variants selected for optimal imputation of common variation and improved imputation of low frequency variation (MAF 1-5%), and 9000 variants selected for improved coverage of known and candidate respiratory regions. Nested case-control analyses were undertaken, using genotyped and imputed variants within the heavy-smoking and never-smoking strata. We reconfirmed associations at previously identified loci and identified additional novel signals of association, notably a large ~1.5Mb region of association on chromosome 17 centred on a directly genotyped SNP, rs2696671, upstream of *KANSL1* (in never smokers, airway obstruction phenotype (low %FEV_1) versus high %FEV_1, $P=7.20\times10^{-10}$). This signal spans a number of genes in a well characterised inversion (17q21.31) and overlaps previously identified signals of association for multiple traits including idiopathic pulmonary fibrosis and Parkinson's disease. We will compare the genetic architecture of extreme lung function phenotypes in smokers to that in non-smokers, and report other emerging novel findings from UK BiLEVE. The genotype data for the UK BiLEVE study will be made available to the scientific community via UK Biobank from 2014, as will subsequent genome-wide genotyping from a similar array in the remaining UK Biobank participants as this becomes available from 2015. This research has been conducted using the UK Biobank Resource.

1102S

***APOL1*-associated kidney disease risk and hypertension management in primary care - A project of the IGNITE Network (Implementing GeNomic medicine In pracTicE).** *N.S. Abul-Husn[1,2], R. Negron[3], K. Ferryman[4], G. Nadkarni[1], S.B. Ellis[1], S. Sanderson[1,2], E. Madden[5], H. Junkins[5], S.E. Kimmel[6], E. Bagiella[3], D. Hauser[7,8], N. Calman[7,8], C.R. Horowitz[3], E.P. Bottinger[1].* 1) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Health Evidence and Policy, Icahn School of Medicine at Mount Sinai, New York, NY; 4) The New School, New York, NY; 5) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 6) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 7) The Institute for Family Health, New York, NY; 8) Department of Family Medicine and Community Health, Icahn School of Medicine at Mount Sinai, New York, NY.

African Americans are more likely to develop kidney disease than European Americans and are at higher risk for progression to kidney failure. Two alleles in the last exon of the *APOL1* gene confer a five-fold increased risk for kidney disease in hypertensive, non-diabetic individuals of African ancestry: the missense variants S342G and I384M (in almost complete linkage disequilibrium; referred to as the G1 allele) and the 6 base pair deletion N388/Y389 (referred to as the G2 allele). Approximately 14% of African Americans carry two *APOL1* risk alleles (vs. < 0.5% of European Americans), contributing to the high burden of kidney disease in this population. Management of hypertension in African Americans with high risk *APOL1* genotypes constitutes an important opportunity for genome-guided medicine. Here, we describe a randomized controlled trial studying the effects of incorporating *APOL1* genotype information into the primary care management of African ancestry patients with hypertension. We will recruit hypertensive, non-diabetic adults with no kidney disease and self-reported African ancestry from ten primary care practices in New York City and randomize them to undergo genetic testing for the *APOL1* G1 and G2 variants at baseline (intervention) or at one year (control). We will educate providers about genome-guided medicine and *APOL1* risk alleles in kidney disease through training sessions, electronic medical records-enabled decision support, and educational materials. Guided by genetic counselors, trained staff will return results to tested patients, and provide them with educational materials suitable for a range of literacy levels. During patient encounters, providers will receive point-of-care clinical decision support to alert them of their patients' *APOL1* genotype and risk status, and to provide recommendations. At three and twelve months, we will assess primary outcomes (blood pressure reduction and appropriate renal surveillance) and secondary, psycho-behavioral outcomes, comparing three patient groups: increased genetic risk, no increased genetic risk, and genetic risk not yet assessed. This study will help to establish the effective implementation of *APOL1* risk-informed management of hypertension in African Americans who are at high risk of kidney disease, and will provide a framework for future endeavors to implement genome-guided medicine in clinical practice.

1103M

Mendelian Randomization study of body mass index/waist hip ratio-associated SNPs and five cancer types. *C. Gao[1], C. Patel[2], S. Lindstrom[1], B. Pierce[3], P. Kraft[1] on behalf of the Genetic Mechanisms in Oncology (GAME-ON) initiative.* 1) Program in Genetic Epidemiology and Statistical Genetics, Harvard School of Public Health, Boston, MA; 2) Center for Biomedical Informatics, Harvard Medical School, Boston, MA; 3) Department of Health Studies, The University of Chicago, Chicago, IL.

Body mass index (BMI) has been shown to be associated with multiple cancer types, but whether the observed associations are causal is unclear due to possible confounding and reverse causality. Mendelian Randomization is a technique that uses genetic predictors of a putative causal factor to eliminate reverse causality and reduce confounding bias. Here, we carried out Mendelian Randomization analyses to assess the causal relationship between BMI and waist-hip-ratio (WHR) and risk of breast, ovarian, prostate, colorectal and lung cancers. BMI and WHR genetic risk scores were derived from 32 GWAS-significant BMI-associated variants and 14 WHR-associated variants. We also derived sex-specific scores based on twelve GWAS significant WHR-associated SNPs for women and four GWAS significant WHR-associated SNPs for men. We tested the association between these genetic scores and risk for each of the five cancers using summary statistics from the Genetic Associations and Mechanisms in Oncology (GAME-ON) Consortium, which conducted GWAS meta-analyses involving between 7,272 and 62,528 cases and controls (depending on cancer type). The test was performed by calculating a weighted average across SNPs of the allelic log odds ratio for cancer, where the weights were defined using the allelic effect on BMI (or WHR) and the standard error of the estimated log odds ratio.

We found that the genetic risk score of BMI was significantly associated with decreased risk of breast cancer (OR=0.89; 95%CI: 0.85, 0.93; pvalue= 3.37×10-7); conversely, the genetic risk score for BMI was associated with increased risk of ovarian cancer (OR=1.08, 95%CI: 1.00, 1.15; pvalue= 0.047). Positive associations, though not significant, were observed between the genetic score for WHR and risk of all cancer types except breast cancer. We observed a significant association between the male-specific genetic score for WHR and decreased risk of breast cancer (OR=0.41; 95%CI: 0.24, 0.72; pvalue=0.002). Our findings implicating opposite risk profiles between adiposity and breast/ovarian cancers may reflect the complex relationship between menopausal status, adiposity, and cancer. By presenting BMI-associated risks across five different cancer types, findings from this study provide additional insights to help understand the relationships between adiposity and cancer risk.

1104T

Distinct differences in HLA genotypes for latent autoimmune diabetes in adults (LADA) and type 1 diabetes within the same extended pedigree. *K.J. Basile[1], V. Guy[1], S. Schwartz[2], D.S. Monos[3], S.F.A. Grant[1,4].* 1) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Main Line Health System, Wynnewood, PA; 3) Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, PA; 4) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Latent autoimmune diabetes in adults (LADA) shares characteristics of both type 1 diabetes (T1D) and type 2 diabetes (T2D), thus often referred to as "type 1.5 diabetes". According to the World Health Organization, LADA is a slowly progressing form of T1D resulting from autoimmune destruction of pancreatic beta-cells. However, LADA also shares characteristics with T2D, such as adult age onset, non-insulin dependence and strong association with variants in the T2D-implicated locus, TCF7L2. Symptoms of LADA are often indistinguishable from early stage T2D, and it has been estimated that 8-10% of patients diagnosed with T2D are in fact positive for specific circulating auto-antibodies. In order to further clarify differences between these forms of diabetes, we performed HLA genotyping of members of a large family consisting of four LADA cases (age of onset >25 yrs), four T1D cases (age of onset <15 yrs) and eight unaffected members, all derived from the National Disease Research Interchange. Of particular note was the generational split for the two diseases, with LADA being all in one generation and T1D all occurring in the next generation. We elected to perform HLA class I and II genotyping (HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DQB1, HLA-DQA1) for all DNA samples obtained from this family using next-generation sequencing tools. Interestingly, the HLA genotypes of the family members with T1D were distinctly different, while the HLA genotypes of the family members with LADA were more similar to unaffected family members. All the LADA cases had HLA genotypes that were identical to at least one unaffected family member. However, none of the family members diagnosed with T1D had any genotypes in common with an unaffected family member and exhibited more classical HLA genotypes (e.g. one T1D case exhibited DRB1*0401 homozygosity, DQB1*0302 homozygosity and DQA1*0301 heterozygosity while two other cases had a distinct DRB1 heterozygous combination of 0401 and 1301, with the latter DRB1 genotype not being present in any other nuclear family member). Based on our results from this somewhat unique family, we conclude that the development of T1D resulted from key HLA alleles being concentrated into the latter generation via incomers to this family. We therefore hypothesize that LADA may have, at least in part, arisen from genetic influences beyond the HLA locus in this family and may reflect a more general mechanism underpinning the genetic etiology of this disease.

1105S

Like Mother, Like Daughter: Analysis of Parent-Child Phenotypic Correlations for Hundreds of Medical and Behavioral Traits. *E. Pierson[1], D. Hinds[1], A. Kleinman[1], N. Eriksson[2].* 1) 23andMe, 1390 Shorebird Way, Mountain View, CA; 2) Coursera, 381 E Evelyn Ave, Mountain View, CA.

We examined correlations between the phenotypes of parents and children in 15,683 trios for 2,235 medical and behavioral phenotypes. While parent and child phenotypes were almost always positively associated, and no phenotypes showed significant negative associations, for most traits that showed significant parent-child associations (58%) the mother's phenotype was more strongly associated with the daughter's than was the father's. There was no disparity for sons. We observed this effect regardless of whether we examined the correlation between parent and child phenotype or used a regression model that controlled for the age, sex, and genetic principal components of the child and included the mother's and father's phenotypes as covariates. When we filtered for phenotypes in which the disparities in parental effect sizes were especially significant ($p < .05$), the mother-daughter effect became more pronounced: the mother showed a stronger association with the daughter than did the father for 35/40 traits. We observed the opposite effect, to a weaker degree, in sons: the father showed a stronger association with the son than did the mother for 24/35 traits. We found specific phenotypes for which children showed significantly more association with one parent after multiple hypothesis correction. Standardized BMI in daughters showed a stronger association with mother's standardized BMI, consistent with previous results; there was no disparity for sons. Smoking behavior in sons (as measured by years smoked) was more strongly associated with father's smoking behavior. We estimated the proportion of these disparities that were attributable to genetics as opposed to environment using SNPs previously associated with the phenotypes and compared parent-sibling correlations with heritability estimates from distant relatives.

1106M

ROBUST MicroRNA EXPRESSION UPREGULATION EXISTS IN INFLAMMATORY BOWEL DISEASE. *S. Ben-Shachar[1], H. yanai[2], L. Baram[2], H. Elad[2], H. Sherman Horev[2], E. Brazowski[3,6], H. Tulchinsky[4,6], M. Pasmanik-Chor[5], N. Shomron[6], I. Dotan[2,6].* 1) Genetic Institute, Tel Aviv Medical Center, Tel Aviv, Israel; 2) IBD Center, Department of Gastroenterology and Liver Diseases, Tel Aviv Medical Center, Tel Aviv, Israel; 3) Department of Pathology, Tel Aviv Medical Center, Tel Aviv, Israel; 4) Proctology Unit, Department of Surgery, Tel Aviv Medical Center, Tel Aviv, Israel; 5) Bioinformatics unit, G.S.W. Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel; 6) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

Introduction: The etiology of inflammatory bowel diseases (IBD), Crohn's disease (CD) and ulcerative colitis (UC) is yet unknown. Similarly, it is not clear why some patients with UC undergoing large bowel resection and creation of ileal pouch (pouch surgery), develop small intestinal inflammation (pouchitis) reminiscent of CD. We have previously shown that gene expression profiles are associated with ileal inflammation and disease behavior in patients with CD and pouchitis. As microRNAs (miRNAs) regulate gene and protein expression, we hypothesized that miRNAs may have a major role in IBD. **Methods:** Patients with CD ileitis, UC (unoperated and after pouch surgery) and normal controls (NC) were recruited. Pouch patients were stratified according to disease behavior into three groups: normal pouch (NP), chronic pouchitis (CP), and Crohn's-like disease of the pouch (CLDP). miRNAs expression in Ileal and pouch biopsies was analyzed using parallel massive sequencing (next generation sequencing-NGS). Effect of inflammation on miRNAs expression was evaluated in Caco-2 and HCT-116 epithelial cell lines after stimulation with IL1β, INFγ, and TNFα. miRNAs expression as well as validation in biological replicates was performed using real time PCR. **Results:** Fifty-Six subjects: 10 CD, 12 NP, 16 pouchitis (12 CP and 4 CLDP) and 11 unoperated UC as well as 7 NC were recruited. The ileum of unoperated UC patients was comparable to NC. Eight, 25 and 124 significant miRNAs alterations (fold change ≥ 2, corrected p-value ≤ 0.05) were noticed in NP, CP, and CLDP patients compared to NC, respectively, with high overlap between pouch subgroups. The magnitude of change correlated with pouch disease behavior. Only two alterations were noted in CD. The validation study confirmed the results. Interestingly, the expression of more than 95% of the altered miRNAs was increased. Most miRNAs with increased expression target mRNA transcripts that were significantly decreased in pouchitis, suggesting an effect of miRNAs on mRNA expression. Inflammatory stimulation of Caco-2 and HCT-116 cells resulted in an expression upregulation of 8/13 (62%) of tested miRNAs. **Conclusions:** The robust increase in miRNAs expression may result from the inflammatory process in IBD. The consequence may be down regulation of mRNA transcripts and proteins, further affecting disease course.

1107T

KinGen: a partnership of high kinship population resources. *J.F. Wilson[1,2], C. Hayward[2], I. Rudan[1], A.F. Wright[2], H. Campbell[1], A. Johansson[3], U. Gyllensten[3], A.A. Hicks[4], P.P. Pramstaller[4], C. Pattaro[4], P. Gasparini[5], D. Toniolo[6], M. Ciullo[7], A.V. Smith[8,9], E. Zeggini[10], C. Ober[11], J.R. O'Connell[12], M. Perola[13,14], J. Marchini[15,16], M. Abney[11], Y.S. Aulchenko[17,18], G. Dedoussis[19], N.D. Hastie[2], C.S. Haley[2].* 1) Centre for Population Health Sciences, Univ Edinburgh, Edinburgh, United Kingdom; 2) MRC Human Genetics Unit, Institute for Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, Scotland; 3) Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 4) Centre for Biomedicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy, Affiliated Institute of the University of Lübeck, Lübeck, Germany; 5) Medical Genetics, Department of Reproductive Sciences and Development, IRCCS-Burlo Garofolo, University of Trieste, Trieste, Italy; 6) Division of Genetics and Cell Biology, San Raffaele Research Institute, Milano, Italy; 7) Institute of Genetics and Biophysics "A. Buzzati-Traverso" - CNR, Naples, Italy; 8) Icelandic Heart Association Research Institute, Kopavogur, Iceland; 9) University of Iceland, Reykjavik, Iceland; 10) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 11) Department of Human Genetics, University of Chicago, Chicago, IL, USA; 12) Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA; 13) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 14) The Estonian Genome Center, University of Tartu, Tartu, Estonia; 15) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 16) Department of Statistics, University of Oxford, Oxford, UK; 17) Institute of Cytology & Genetics SB RAS, Novosibirsk, Russia; 18) "Yurii Aulchenko" consulting, Groningen, The Netherlands; 19) Harokopio University, Athens, Greece.

The application of genome-wide association analysis to diverse sample cohorts has revolutionised our understanding of the common genetic variants influencing complex diseases and their risk factors. However, as the development of low cost sequencing facilitates greater focus on rarer genetic variants, the particular characteristics of population resources become more important. Specifically populations with increased kinship - across a spectrum spanning nuclear families in a cosmopolitan setting to isolated populations - stand apart from collections of unrelated individuals in different ways: 1. Decreased genetic heterogeneity, 2. Alleles can sometimes rise via drift to tractable frequencies, 3. Increased genomic sharing, 4. Increased linkage disequilibrium, 5. Increased homozygosity, 6. Possibility of parent-of-origin and segregation-based analyses. 7. Availability of deep pedigrees. KinGen is a new partnership of 24 population-based cohorts with high kinship including over 100,000 research subjects, which aims to leverage the special population structure to deliver insights into architecture of complex traits that are difficult to achieve in other settings. Subsets of the KinGen sample have variously whole genome sequence, whole exome sequence, exome and genome-wide chip genotyping, together with very rich phenotyping, some of which is harmonised across cohorts. Here we present a number of analyses of identity-by-descent (IBD) sharing. The predominantly isolated populations in KinGen vary dramatically in the amount and type of kinship observed between pairs of individuals. The number and sum length of runs of homozygosity and of segments shared IBD are greatly enriched over and above typical outbred populations. Analyses of surrogate parents with long range phasing methods reveal striking variation from cohort to cohort in the degree of surrogacy. Drift in allele frequencies has been compared to related cosmopolitan populations, summarizing how often otherwise rare alleles have increased significantly in frequency versus how many alleles have been lost. The population genetic characteristics translate into more accurate imputation of rare and low frequency variants and whole genome sequences and also result in concomitant gains in association power for a subset of variants.

1108S

Meta-Analysis of Glaucoma Genome-Wide Imputed Datasets. *J.N. Cooke Bailey[1], S.J. Loomis[2, 3], M.A. Hauser[4,5], L.R. Pasquale[2,3], J.H. Kang[3], R.R. Allingham[4], R.N. Weinreb[6], J.L. Wiggs[2], J.L. Haines[1], NEIGHBORHOOD Consortium.* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, MA; 3) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 4) Department of Ophthalmology Duke University Medical Center, Durham, NC; 5) Department of Medicine, Duke University Medical Center, Durham, NC; 6) Department of Ophthalmology, Hamilton Eye Center, University of California, San Diego, SD, CA.

Glaucoma is a phenotypically and genetically complex neurodegenerative disease that is the second leading cause of blindness worldwide. Although genetic factors are known to contribute to glaucoma and to endophenotypes associated with it (e.g., intraocular pressure, cup-to-disc ratio, optic nerve parameters, and central corneal thickness), identified risk loci fail to fully account for the genetic component of glaucoma. To extend the proportion of the genome testable for association with glaucoma as well as the power to detect associations at numerous loci, we evaluated imputed genome-wide data in the NEIGHBORHOOD Consortium dataset, which consists primarily of individuals of European descent. Eight datasets, typed on different genome-wide SNP arrays, were imputed to the March 2012 version of the 1000 Genomes data using IMPUTE2 and/or MaCH/miniMaC. Each dataset was individually evaluated for genomic inflation and then filtered for minor allele frequency (MAF≥0.05) and imputation quality (r2 or info metric ≥ 0.7). We then performed study-specific dosage analysis (using ProbABEL) evaluating the estimated genotypic probabilities from the imputation step. The study-specific logistic regression models included as covariates, age, gender, and significant Eigenvectors, along with, in some cases, study-specific covariates. These study-specific results were meta-analyzed applying the inverse variance weighted method in METAL and applying genomic control correction. The meta-analyzed dataset includes 3,873 POAG cases and 33,642 controls, and a total of 7,301,525 variants that passed quality control parameters in at least one dataset. Significant associations were found for genomic regions previously associated with POAG (TMCO1, CDKN2BAS, SIX6) as well as novel regions on 6p and 14q. Imputation of genome-wide array data extends the genomic coverage beyond what can be interrogated by a single or multiple arrays. Meta-analyzing multiple imputed datasets can successfully be implemented to study common diseases with complex inheritance, such as glaucoma, and will help fully define the underlying genetic architecture.

1109M

Influence of BMI- and lipid-associated variants on longitudinal phenotypes. *R.M. Salem[1,2,3], J.N Hirschhorn[1,2,3].* 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Department of Endocrinology, Boston Children's Hospital, Boston, MA.

Genome-wide association studies (GWAS) have allowed geneticists to gain significant insights into the genetic architecture of many polygenic traits and diseases of public health importance. However, these studies have tended to focus on a single quantitative or dichotomous phenotype at a single time point. While informative, these simple characterizations disregard important complexities of polygenic traits. Most polygenic traits, such as BMI or lipid levels, change over time, and genetic influences in these traits may be exerted at different stages in the life cycle. By considering more informative phenotypes, study designs and analytical approaches that take time into account, it may be possible to shed light on new or different aspects of biology. Methods: We assessed the longitudinal influence of 187 genetic variants known to be associated with BMI (GIANT 2010) or one or more lipid traits (HDL, LDL, total cholesterol (TC) and triglycerides (TG), Global Lipids Consortium 2013). We analyzed data from 10,950 European Ancestry subjects from the ARIC and MESA cohorts downloaded from dbGaP. The longitudinal analysis was performed using a mixed model analysis implemented in SAS, optimized for faster run time. Lipid phenotypic measures with prevalent or incident diabetes or concurrent statin medication use were excluded from analysis. SNPs were considered significant if SNPxAGE p-value was <0.05 in both cohorts. Results: None of the 31 BMI-associated SNPs influenced longitudinal change in BMI (SNP x Age p<0.05). For the lipids traits, two loci were identified that influenced longitudinal change in HDL and TG. Decreasing longitudinal HDL level was significantly influenced by the HLD decreasing G allele of rs4148008 in locus *ABCA8* (SNPxAge, ARIC p=0.016 and MESA p=0.013). The TC decreasing C allele of rs7515577 in locus *EVI5* was associated with decreasing TG levels across time (SNPx AGE, ARIC p=0.028 and MESA p=0.041). Additional large longitudinal samples or methods to utilize cross-sectional data may be required to validate results. Conclusions and future directions: Using a mixed-model approach and longitudinal data, we showed that two lipids associated variants influence longitudinal change in HDL and TG. Genome-wide application of these and other approaches may reveal SNPs whose longitudinal effect is more prominent than or independent of main effects.

1110T

ABCG2 dysfunction causes renal underexcretion hyperuricemia as well as renal overload hyperuricemia. *H. Matsuo[1], A. Nakayama[1], M. Sakiyama[1], T. Chiba[1], S. Shimizu[1], H. Nakashima[2], H. Nakaoka[3], K. Wakai[4], T. Ito[5], K. Yamamoto[6], Y. Sakurai[2], K. Ichida[7], T. Shimizu[8], N. Shinomiya[1].* 1) Department of Integrative Physiology and Bio-Nano Medicine National Defense Medical College, 3-2 Namiki, Tokorozawa, Saitama 359-8513, Japan; 2) Department of Preventive Medicine and Public Health, National Defense Medical College, 3-2 Namiki, Tokorozawa, Saitama 359-8513, Japan; 3) Department of Integrated Genetics, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan; 4) Department of Preventive Medicine, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan; 5) Department of Internal Medicine, Self-Defense Forces Central Hospital, 1-2-24 Ikejiri, Setagaya-ku, Tokyo 154-8532, Japan; 6) Division of Genome Analysis, Research Center for Genetic Information, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan; 7) Department of Pathophysiology, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachiouji, Tokyo 192-0392, Japan; 8) Midorigaoka Hospital, 3-13-1 Makami-cho, Takatsuki, Osaka 569-1121, Japan.

Background and objectives Gout is a common disease which results from hyperuricemia. We have reported that the decrease in extra-renal (intestinal) urate excretion by ABCG2 dysfunction induced renal urate overload (ROL) hyperuricemia. This mechanism, however, does not give a sufficient explanation for all ABCG2 dysfunction cases as a major cause of hyperuricemia and gout, because renal underexcretion (RUE) hyperuricemia is the most prevalent subtype of hyperuricemia. In this study, we focus on the involvement of ABCG2 dysfunction in RUE hyperuricemia. **Design, setting, participants & measurements** We performed genotyping for 2,267 Japanese male participants, who consisted of 644 hyperuricemia cases (SUA>7.0 mg/dl) and 1,623 controls. Their functional ABCG2 activities were estimated from their genotype combinations of its two dysfunctional missense variants, Q126X (rs72552713) and Q141K (rs2231142). All participants were divided into four groups by ABCG2 function, and all cases are classified into RUE hyperuricemia or ROL hyperuricemia by the patients' fractional excretion of urate (FE_{UA} < 5.5%) and urinary urate excretion (UUE > 25.0 mg/hr/1.73m^2), respectively. **Results** The association analysis revealed that ABCG2 dysfunction significantly raised the risk of both RUE and ROL hyperuricemia. The risk of hyperuricemia was increased as the ABCG2 function decreased. When hyperuricemia was divided into three distinct types, all but RUE type with the severe ABCG2 dysfunction showed the significant risk for hyperuricemia. Nevertheless, moderate and mild dysfunction still contributed to increase the risk of RUE type hyperuricemia. These data imply that ABCG2 dysfunction under certain conditions causes renal urate underexcretion and leads to hyperuricemia even without renal urate overload. **Discussion** Our study show that ABCG2 plays physiologically important roles in renal urate excretion as well as extra-renal urate excretion, and its dysfunctional mutations are involved in all types of hyperuricemia as their major genetic causes. *ABCG2* genotyping in combination with FE_{UA} and UUE tests is sufficient for screening high-risk individuals with hyperuricemia and gout. Our findings will therefore serve to build up the health of people predisposed to hyperuricemia and gout.

1111S

Coherent Somatic Mutation in Autoimmune Disease. *K.A. Ross.* Computer Science, Columbia University, New York, NY.

Many aspects of autoimmune disease are not well understood, including the specificities of autoimmune targets and patterns of co-morbidity and cross-heritability across diseases. We hypothesize that many autoimmune diseases are triggered by immune responses to proteins whose DNA sequence mutates somatically in a coherent, consistent fashion. Simple tandem repeat (STR) sequence is highly mutable, both somatically and in the germ-line, and somatic STR mutations are observed under inflammation. Protein-coding genes spanning STRs having markers of mutability, including high total length, high repeat count, high repeat similarity, and germ-line variability, are evaluated in the context of autoimmunity. For the initiation of autoimmune disease, antigens whose autoantibodies are the first observed in a disease, termed primary autoantigens, are informative. Known primary autoantigens include thyroid peroxidase (TPO) in Hashimoto's thyroiditis, phogrin (PTPRN2) in type-1 diabetes, and filaggrin (FLG) in rheumatoid arthritis. The genes encoding these three autoantigens include STRs that are among the eleven longest STRs spanned by protein-coding genes. This association of primary autoantigens with long STR sequence is highly significant ($p<3.0\times10^{-7}$). Long STRs within twenty genes are associated with sixteen common autoimmune diseases and atherosclerosis. The TTC34 gene is an outlier in terms of repeat length and count, and is proposed as a candidate gene for autoimmune induction in systemic lupus erythematosus. The results support the hypothesis that many autoimmune diseases are triggered by immune responses to somatically mutated proteins. Other autoimmune diseases may be caused by coherent somatic mutations in immune cells. The coherent somatic mutation hypothesis, in which similar somatic mutations happen both within and across individuals, has the potential to be a comprehensive explanation for the initiation of many autoimmune diseases.

1112M

Secular change in 13 metabolic phenotypes: A Chinese longitudinal twin study. *S. Li[1], Z. Pang[2], D. Zhang[3], H. Duan[2], Q. Tan[4], J. Hjelmborg[4], T. Kruse[1], K. Kyvik[5].* 1) Human Genetics, Institute of Clinical Research, University of Southern Denmark, Odense, Odense, Denmark; 2) Qingdao Center for Disease Control and Prevention, Qingdao, China; 3) Dept. of Public Health, Qingdao University Medical College, Qingdao, China; 4) Epidemiology, Biostatistics and Biodemography, Institute of Public Health, University of Southern Denmark; 5) Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark.

Aims: The genetic and environmental influences on metabolic phenotypes have been intensively studied by twin modeling in different populations. However, twin studies on secular change in metabolic phenotypes have been rare due to high expenses, losses of follow up, and long waiting time in prospective investigations. Based on Chinese twin data collected from Danish-Chinese collaboration research, we perform twin modeling on 13 metabolic phenotypes (total cholesterol; triglyceride; high density lipoprotein (HDL); low density lipoprotein (LDL); urine acid (UA); glucose; weight; body mass index (BMI); waist and hip circumstance and ratio (WHR); systolic blood pressure (SBP); diastolic blood pressure (DBP)) measured longitudinally in a period of 7 years in 254 middle aged Chinese twins (126 monozygotic; 128 dizygotic twins). **Methods:** Univariate ACE, ADE models and their nested models were fitted to the secular changes in each of the 13 phenotypes with best fitting model selected based on model performance. Age and sex were included as covariates in the models to adjust for their effects on secular trend. **Results:** Variations in secular change in 3 lipids (total cholesterol; triglyceride; LDL), weight and blood pressure (SBP, DBP) have AE model as the best fitting model with low to moderate (a^2: 0.25 - 0.46) genetic control. Variations in secular change in the rest of the phenotypes have CE model as the best fitting model with low to moderate (c^2: 0.21 - 0.55) control by shared environmental factors. Secular changes in all phenotypes are under moderate to high (e^2: 0.45 - 0.79) control by unique environmental factors. **Conclusions:** Variations in secular change in the 13 metabolic phenotypes show limited genetic control. Our results emphasize the special importance of unique environment in determining individual change in metabolic phenotypes in adult Chinese.

1113T

Latent variable adjustment of NIH Epigenomics Roadmap ChIP-seq data for utilization in tissue-specific polygenic analysis of type II diabetes in the DIAGRAMv3 GWAS meta-analysis. *A.L. Dobbyn[1], B.F. Voight[2], P. Roussos[1,6], A. Morris[3], S. Raychaudhuri[4,5], M. McCarthy[3], E.A. Stahl[1,5,6], the DIAGRAM consortium.* 1) Icahn School of Medicine at Mount Sinai, New York City, NY; 2) Departments of Pharmacology and Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 4) Partners HealthCare Center for Personalized Genetic Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA; 5) Broad Institute, Cambridge, MA; 6) Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY.

INTRODUCTION: Previous analyses have shown that tissue specific chromatin marks can be used to inform association studies in order to dissect the genetic architecture of complex disease. Enrichment for the H3K4me3 chromatin mark in GWAS identified variants has implicated causal tissues in the etiology of complex disease, including both Liver and Pancreatic Islets for type II diabetes (Trynka 2013). However, we have detected similar levels of enrichment in non-causal tissues in separate analyses, and observe a lack of correlation between H3K4me3 peaks from samples of the same tissues, consistent with the presence of latent confounding variables. We outline a method for the normalization of ChIP-seq data that adjusts for confounders, allowing for improved detection of enrichment signal in downstream analyses, and then utilize the normalized data in polygenic risk score analysis of T2D. METHODS: We constructed a matrix of the union of all peaks in all samples from the NIH Epigenomics Roadmap Project (82 samples, 44 tissues/cell types), merging peaks with any overlap, resulting in a total of 214,704 peaks. We then conducted surrogate variable analysis (Leek & Storey, 2007) using this matrix in order to identify and adjust peak fold enrichments for latent variables. We will make publicly available our normalized adjusted Epigenomics Roadmap data. We used H3K4me3 ChIP-seq peaks to define tissue-specific H3K4me3 enriched sets of genes, and selected SNPs within 20kb of each gene's transcription start site for use in polygenic analysis. We then conducted polygenic risk score analysis of T2D using discovery SNPs from the DIAGRAMv3 GWAS meta-analysis (Morris 2012; 12,171/56,862 cases/controls), and tested polygenic risk scores in the WTCCC (2007; 1,924/2,938 cases/controls). RESULTS: After normalization, maximum enrichment (observed/expected variance explained) for Liver and Pancreas SNP sets, based on a small number of tissue-specific peaks, increased from 4.37 and 7.08 to 38.6 and 17.5, respectively. Liver and Pancreas within-tissue correlations increased from R_{Liver}=0.785 and R_{Panc}= 0.570 to R_{Liver}=0.858 and R_{Panc}=0.859. CONCLUSIONS: We have implemented surrogate variable analysis for the normalization and adjustment of ChIP-seq data, reducing the effects of hidden covariates and improving detection of tissue-specific compartments of polygenic signal. The adjustment used is applicable to all downstream analyses that utilize tissue specific ChIP-seq data.

1114S

Excess of Runs of Homozygosity is associated with severe cognitive impairment in Intellectual Disability. *I. Gandin[1], F. Faletra[2], F. Faletra[2], M. Carella[3], V. Pecile[2], G.B. Ferrero[4], E. Biamino[4], P. Palumbo[3], O. Palumbo[3], P. Bosco[5], C. Romano[6], C. Belcaro[1], D. Vozzi[2], A.P. d'Adamo[1].* 1) Medical Sciences, University of Trieste, Trieste, Trieste, Italy; 2) Institute for Maternal and Child Health, IRCCS "Burlo Garofolo", Trieste, Italy; 3) Medical Genetics Unit, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG), Italy; 4) Department of Pediatrics, University of Torino, Torino, Italy; 5) Oasi Institute for Research on Mental Retardation and Brain Aging (IRCCS), Troina, Italy; 6) Unit of Pediatrics and Medical Genetics, IRCCS Associazione Oasi Maria Santissima, 94018 Unit of Pediatrics and Medical Genetics, IRCCS Associazione Oasi Maria Santissima, 94018 Troina (EN), Italy.

The harmful effect of inbreeding is well known by geneticists and several studies have already reported cases of Intellectual Disability (ID) in consanguineous families caused by recessive variants. Nevertheless, despite the recent findings for other neurological disorders like Alzheimer disease, Parkinson disease and schizophrenia, the effect of inbreeding on ID is still poorly investigated in outbred populations. Here we present a Runs of Homozygosity (ROHs) study performed in a 612 individual cohort of ID patients. We first investigated the ROHs distribution considering the complexity of the phenotype (presence of malformations, seizures, micro/macrocephaly, etc.), and we didn't find a different amount of homozygosity comparing syndromic and non-syndromic ID patients. In a second stage, we focused on the effect of ROHs on the ID degree, starting from the IQ. Our data revealed the presence of significantly larger ROH stretches in severe ID cases compared to non-severe ID ones (p=7.1×10-3), together with an increase of the percentage of genome covered by ROHs (p=3.0×10-2). According to the recent findings on ID in autism, this study reveals that autosomal recessive variants have an important role in ID and that they mainly and specifically affect the cognitive impairment modulation, in spite of the ID cause. This is not surprising, since ID degree is a highly complex trait whose overall variability is likely not determined by a single pathogenic mutation. Other environmental and genetic factors modulate the phenotype and one of the most important seems to be accounted for by ROHs.

1115M

Impact of genetic burden on the age at onset in bout-onset and progressive multiple sclerosis. *M. Sorosina[1], F. Esposito[1,2], C. Guaschino[1,2], F. Clarelli[1], N. Barizzone[3,4], S. Lupoli[5], V. Martinelli[2], M. Leone[3], G. Comi[2], S. D'Alfonso[3,4], F. Martinelli Boneschi[1,2], PROGEMUS group, PROGRESSO group.* 1) Laboratory of genetics of Neurological complex disorders, Institute of Experimental Neurology (INSPE), San Raffaele Scientific Institute, Milan, Italy; 2) Department of Neurology, San Raffaele Scientific Institute, Milan, Italy; 3) Interdisciplinary Research Center of Autoimmune Disease IRCAD, University of Eastern Piedmont, Novara, Italy; 4) Department of Health Sciences, University of Eastern Piedmont, Novara, Italy; 5) Department of Health Sciences, University of Milan and Genomics & Bioinformatics Unit, c/o Fondazione Filarete, Milan, Italy.

Multiple sclerosis (MS [MIM 126200]) is an inflammatory disease of the central nervous system (CNS) characterized by myelin damage and neurodegeneration. It is a highly heterogeneous disorder and two main clinical courses exist: the bout onset MS (BOMS) and the progressive onset MS (PrMS). While the susceptibility to the disease has been widely explored from the genetic point of view, conflicting evidence exist on the role played by genetic risk factors in influencing the clinical features. We calculated a weighted genetic risk score (wGRS) in 1450 Italian multiple sclerosis patients using 107 MS-associated genetic risk variants recently reported by the IMSGC consortium to explore the influence of genetic burden on the age at onset (AAO) in the two MS disease courses. By analysing the entire MS cohort, no effect of wGRS was detected on AAO (rho=-0.04, p=0.15; comparison of patients stratified according to the median age at progression: p=0.13), although the trend was toward a negative correlation. After stratification based on the clinical course we observed an opposite relationship between the wGRS and AAO in the two groups: negative in BOMS (n= 796; correlation: rho=-0.1; $p=5\times10^{-3}$; comparison of patients stratified according to the median AAO: $p=9.5\times10^{-3}$), while positive in PrMS (n=474) and in Secondary Progressive MS (SPMS, n=94) referring to the onset of the progressive phase (correlation: rho=0.08; p=0.07; comparison of patients stratified according to the median age at progression: p=0.01). In SPMS females this association seems to be due to a longer inflammatory phase preceding the progressive phase (correlation: rho=0.26, p=0.047; survival analyses comparing patients stratified according to the wGRS: p=0.043), instead to the age of progression itself. While the relationship between wGRS and AAO seems to be mediated also by the non-HLA loci, the effect on age at progression seems to be mediated mainly by HLA. Our results suggest that the MS-associated risk variants differentially influence the onset of the inflammatory and progressive phases in multiple sclerosis.

1116T

Association of Mitochondrial DNA levels with Frailty and All-Cause Mortality. *F.N. Ashar[1], A. Moes[1], A.Z. Moore[2], M.L. Grove[3], P.H.M. Chaves[4], J. Coresh[2], A.B. Newman[5], A.M. Matteini[6], K. Bandeen-Roche[7], E. Boerwinkle[3], J.D. Walston[6], D.E. Arking[1].* 1) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD., USA; 2) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; 3) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA; 4) Benjamin Leon Center for Geriatric Research and Education and Department of Medicine, Florida International University, Miami, USA; 5) Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, USA; 6) Division of Geriatric Medicine and Gerontology, Johns Hopkins University School of Medicine, Baltimore, USA; 7) Department of Biostatistics, Johns Hopkins University, Baltimore, USA.

Age-related declines in mitochondrial function have long been hypothesized to underlie multiple biological changes that increase vulnerability to multiple disease states, functional and cognitive decline, and ultimately, mortality. Evidence to support this comes from the fact that specific variants in mitochondrial DNA (mtDNA) have been shown to modulate risk for several age-related disease states. While the association of mitochondrial variants with age-related disorders is well-established, the effect of mtDNA copy number, which reflects energy reserves and oxidative stress, on aging and mortality in the general population has not been addressed. To address this gap, we examined mtDNA copy number in two large multi-center prospective studies—the Cardiovascular Health Study (CHS) and the Atherosclerosis Risk in Communities (ARIC) study—with a total of 16,401 samples of European and African descent focusing on two primary phenotypes—prevalent frailty in CHS, and all-cause mortality in ARIC and CHS. In race-stratified meta-analyses, we demonstrate a significant inverse association of mtDNA copy number with age, with a reduction of 0.12 standard deviation units with a 10 year differences in age ($P=2.78 \times 10^{-25}$), and higher mtDNA copy number in women relative to men (meta-analysis OR=1.33, 95% CI 1.14-1.51, $P=6.91 \times 10^{-45}$). Furthermore, we show that lower mtDNA copy number is significantly associated with prevalent frailty in 4,109 self-identified white participants from CHS (OR=0.91, 95% CI, 0.85-0.97, P=0.005). Finally, mtDNA copy number is a strong predictor of all-cause mortality in an age and sex-adjusted, race-stratified analysis of 16,401 participants from both cohorts with a pooled hazard ratio of 1.47 (95% CI 1.33-1.62, $P=4.24 \times 10^{-14}$) for the lowest quintile of mtDNA copy number relative to the highest quintile. Based on the association with mortality and observation of a higher mtDNA copy number in women relative to men, we tested whether a mito-protective effect may be related to the differences in life expectancy between men and women. We estimate that mtDNA copy number is associated with 3-10% of the effect of sex on mortality, suggesting that mtDNA copy number may be a contributing factor to the disparity of the phenotype. In summary, we report that a single, easily implemented measure of mtDNA copy number, isolated from whole blood decades before the event of interest (death), is predictive of all-cause mortality.

1117S

Significant role of height-associated variants in the variation of intracranial volume. *R. Shafee[1,2], G. Genovese[2], A.J. Holmes[3], P.H. Lee[4], L. Germine[4], J.L. Roffman[5], J.W. Smoller[4], R.L. Buckner[6,7], S.A. McCarroll[1,2].* 1) Genetics, Harvard Medical School, Boston, MA; 2) Stanley Ctr. for Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA; 3) Dept. of Psychology, Yale University, New Haven, CT; 4) Psychiatric and Neurodevelopmental Genet. Unit, Massachusetts Gen. Hosp., Boston, MA; 5) Psychiatry, Massachusetts Gen. Hosp., Boston, MA; 6) Athinoula A. Martinos Ctr. for Biomed. Imaging, Dept. of Radiology, Massachusetts Gen. Hosp., Charlestown, MA; 7) Psychology and Ctr. for Brain Sci., Harvard Univ., Cambridge, MA.

Height and intracranial volume (ICV) are both highly heritable traits as shown by twin studies. Additionally, it has been observed that the two quantities are positively correlated. We investigated the contribution of height-associated variants to the heritability of ICV and tested whether height-increasing alleles also increase ICV using 963 18-35 year old healthy Caucasian subjects. ICV was measured from the magnetic resonance imaging (MRI) scans of the subjects. The genotype data of these subjects comprised of ~531,000 SNPs after standard quality control. The height-association p-values of these SNPs were acquired from the publicly available results of the GIANT consortium. First, using GCTA we estimated the SNP-heritability of ICV due to all the genotyped SNPs to be $h_g^2 \sim 0.29 - 1$ (95% CI, LRT p-value = 0.0002). Next, we categorized the SNPs as height-associated (HA) or not-height-associated (NHA) based on varying GIANT association p-value thresholds and estimated heritability contributions from each component for each case using a joint model in GCTA. For a threshold of $p = 0.5$, 49% of the genotyped SNPs were HA ($p < 0.5$) and 51% of the SNPs were NHA ($p \geq 0.5$) with the HA SNPs showing a trend of enrichment (1.9X enrichment, $p = 0.18$, 95% CI of $h_g^2 = 0.3 - 1$) over the NHA SNPs. Lowering the GIANT p-value threshold for defining HA/NHA SNPs strengthened the trend. For a threshold of $p = 10^{-4}$ HA SNPs (0.6% of all SNPs) showed statistically significant enrichment (20X, $p = 0.0042$, $h_g^2 = 0.1 - 0.8$) over the NHA SNPs in explaining the SNP-heritability of ICV. Next, we performed association tests of ICV using all genotyped SNPs and investigated whether the height-increasing alleles of the SNPs also increased ICV in our sample. We tested whether the average ICV association test regression coefficient (β) of the height-increasing alleles of the genotyped SNPs was positive. Once again, we divided the SNPs as HA (GIANT p-value <0.5) or NHA (GIANT p-value ≥ 0.5) and found the mean β of the height-increasing alleles to be 96.4 ± 14.5 mm^3/allele (mean ± SEM; permutation p-value = 0.04 for mean ≥ 96.4) for the HA SNPs and 15.2 ± 15.2 mm^3/allele (with permutation p-value = 0.4 for mean ≥ 15.2) for the NHA SNPs indicating high probability of the mean positive β arising by chance in the latter case. Together these results indicate that height-associated SNPs possibly play an important role in determining ICV.

1118M

MultiBLUP: Improved Prediction for Complex Traits. *D. Speed, D.J. Balding.* UCL Genetics Institute, University College London, London, United Kingdom.

Prediction of phenotype from genome-wide SNPs is challenging. It is now well established that for many complex traits, in addition to a handful of individually-significant SNPs, substantial heritability is carried in a polygenic component that is tagged by a much larger set of SNPs. However, there are currently no statistical tools for prediction that can exploit this genetic architecture and remain computationally tractable. BLUP (Best Linear Unbiased Prediction) is a classic prediction tool widely used in plant and animal breeding. It is computationally fast, but is suited only to simple polygenic traits, and not to complex genetic architectures. We therefore introduce MultiBLUP, which generalises BLUP in a way that retains its good computational properties while increasing its flexibility so that it can perform well for a wide range of genomic architectures. In tests on many disease traits, it consistently achieves better risk prediction than rival methods, while being many times faster than the next best method and able to handle much larger datasets. For example, applied to imputed data for inflammatory bowel disease (13,000 individuals, 1.5M SNPs), MultiBLUP runs in a few hours, and achieves better prediction performance than BLUP or genetic risk scores (AUC 0.68 compared to 0.58 and 0.61, respectively). Similar improvement is achieved for Celiac Disease (AUC 0.86 compared to 0.79 for both BLUP and genetic risk scores). MultiBLUP can also be used for predicting continuous traits (we demonstrate this for cholesterol phenotypes), and for non-human data (we consider 139 mice phenotypes). The low prevalence of most diseases means that prediction in a general population remains difficult, but there are many settings in which subgroups of patients can be identified using classical risk factors. For example, the prevalence of Type 2 diabetes among obese individuals is almost 50%, while the chance of developing epilepsy following a single seizure is about 45%, which are both much higher than the population prevalences. SNP-based prediction using MultiBLUP can be very powerful for such subgroups. Moreover, with meta-analysis consortia established for many major diseases, MultiBLUP can take advantage of the very large sample sizes now available. The MultiBLUP software will be freely available at www.ldak.org after publication.

1119T

Hair e-QTLs - delineating the genetic basis of gene-expression in human hair follicle and its implication for the interpretation of hair loss disorders. *S. Heilmann[1,2], H. Schulz[3], N. Karbalai[4], A. Hofmann[1,2], L.M. Hochfeld[1,2], F.A. Degenhardt[1,2], S. Herms[1,2,5], J. Becker[1,2], T. Kätzel[1,2], F.F. Brockschmidt[1,2], A.M. Hillmer[6], P. Hoffmann[1,2,5,7], B. Müller-Myhsok[4], M.M. Nöthen[1,2].* 1) Institute of Human Genetics, University Bonn, Bonn, Germany; 2) Department of Genomics, Life & Brain Center, University of Bonn, Germany; 3) Cologne Center for Genomics (CCG), University Cologne, Cologne, Germany; 4) Max-Planck-Institute of Psychiatry, Munich, Germany; 5) Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel, Switzerland; 6) Cancer Therapeutics and Stratified Oncology, Genome Institute of Singapore, Singapore, Singapore; 7) Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany.

The human scalp hair plays an important role in our social and cultural life and its undesirable loss is often perceived as psychologically stressful in affected individuals resulting in a reduced quality of life. This has provoked many studies on the genetic and biological causes of hair loss. Candidate gene and genome-wide association studies have been successful in the identification of genetic risk factors that predispose to hair loss disorders such as alopecia areata and male-pattern baldness. However, the biological context in which these genetic risk factors exert their biological effect is often unknown. Although the functional annotation of the human genome is becoming increasingly detailed, further experiments are still warranted to elucidate the often tissue- and cell-specific biological contexts. Here, the analysis of genetic variants that influence gene expression (eQTLs) has gained major importance. The aim of our study therefore was to systematically map eQTLs in human hair follicle. For this purpose we performed genome-wide genotyping of blood-DNA samples and transcriptome profiling of hair follicle RNA samples from 100 healthy male donors using the Illumina OmniExpress v1.1 and HT12v4 array. After quality control, a total of 562,176 autosomal SNPs (MAF≥5%, call rate≥95%) and 21,696 expression probes (quantile normalized, detection P-value<0.05 in at least one sample, no SNP with MAF>1% within probe) for 97 individuals remained for eQTL analysis. A total of 5,438 SNPs were found to regulate expression of 856 transcripts in cis or trans, corresponding to almost 1% of all SNPs analyzed. Among the top cis-regulated genes are *CHURC1*, *KCTD10*, *WDR41*. Top trans-regulated genes include *RPS26L*, *PEX16* and the long-non-coding RNA *MGC57346*. Elaborate data analysis is currently underway and the results will be presented at the meeting. The present hair-eQTL data set will help to further elucidate the underlying biological mechanisms of hair loss disorders such as male-pattern baldness and alopecia areata and hair morphological traits such as hair thickness or curliness. Promising strategies include the interpretation of GWAS findings using the hair-eQTL data, tests for association of the limited set of functionally relevant hair-eQTLs in case-control samples for hair related traits and disorders as well as pathway- and score-based approaches.

1120S

Next generation sequencing approaches for the identification of novel genes in spinocerebellar degeneration. *M. Coutelier[1,2,3,4,5,6], L. Raymond[1,2,3,4,5], C. Tesson[1,2,3,4,5], M. Mairey[1,2,3,4,5], J. Konop[1,2,3,4,5], M. Jacoupy[1,2,3,4], T. Esteves[1,2,3,4,5], C. Goizet[7], M. Gaussen[1,2,3,4], F. Darios[1,2,3,4], A. Durr[1,2,3,4,8], JF. Deleuze[9], G. Rouleau[10], S. Zuchner[11], A. Brice[1,2,3,4,8], G. Stevanin[1,2,3,4,5].* 1) Sorbonne Universités, UPMC Univ Paris 06, UMR S 1127, ICM, F-75013, Paris, France; 2) Inserm, U 1127, F-75013, Paris, France; 3) CNRS, UMR 7225, F-75013, Paris, France; 4) Institut du Cerveau et de la Moelle Epinière, Paris, France; 5) Laboratoire de Neurogénétique, Ecole Pratique des Hautes Etudes, ICM, CHU Pitié-Salpêtrière, Paris, France; 6) Laboratoire de Génétique Humaine, Institut de Duve, UCL, Bruxelles, Belgique; 7) Univ. Bordeaux, Maladies Rares: Génétique et Métabolisme (MRGM), EA 4576, Talence, France; CHU Bordeaux, Hôpital Pellegrin, Service de Génétique Médicale, Bordeaux, France; 8) AP-HP, Hôpital de la Pitié Salpêtrière, Fédération des maladies du système nerveux, F-75013, Paris, France; 9) Commissariat à l'Energie Atomique/Direction des Sciences du Vivant/Institut de Génomique, Centre National de Génotypage, 91057 Evry, France; 10) Montreal Neurological Institute and Hospital, Department of Neurology and Neurosurgery, McGill University, Montréal, QC H3A 2B4, Canada; 11) Department of Human Genetics and Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, FL.

Cerebellar Ataxias (CA) and Hereditary Spastic Paraplegias (HSP) are neurodegenerative diseases, which belong to the family of Spinocerebellar Degenerations (SCD). They are characterized by major clinical and genetic heterogeneity, which renders both the molecular diagnosis and the phenotype-genotype correlations really complex. As of today, only 60% of SCD cases are genetically elucidated even though more than 150 causal genes have been described. Aiming at the identification of new causative genes, we gathered the strengths of Next Generation Sequencing technologies (NGS), and the SPATAX network with its cohort of more that 10000 SCD patients. Whole Exome Sequencing (WES) was performed in 54 families (1-3 patients/family). Along with classical criteria based on database frequencies and effect on the encoded protein, variants were sorted with linkage analysis data in the most informative families, or with gene network analysis in the smaller ones. In 9 families, variants in genes previously involved in SCD were described (FTL, SCA14, C10orf2, SETX, ATM, COL6A3, PIK3R5). Another noteworthy family presented with both heterozygous and homozygous mutations of the same gene, leading to interesting intra-familial genotype-phenotype comparison. In yet another family, a heterozygous mutation of a gene already described in an autosomal recessive heavier neurocutaneous phenotype shed the light on the putative broadening of phenotypes that will be brought along by NGS. Two new genes in autosomal recessive CA are published (Synofzik et al., Brain 2014) or submitted. In the remaining families, we isolated 41 candidate genes in autosomal dominant CA, one of which is involved in 2 families; 7 in HSP, again with mutations of one gene in 2 separated families; and 4 in autosomal recessive CA with associated hypogonadism. We designed targeted gene panels aimed at screening our extended cohort for these genes. Functional analyses are already ongoing for 6 of these genes. In conclusion, we report results from WES in a cohort of 54 families with SCD, with clues on frequency of already known genes, novel elements on the complexity of phenotype-genotype correlations, as well as several promising candidate genes. We furthermore demonstrate how high-throughput and medium-throughput sequencing may be used efficiently in novel genes identification and validation. Further results on extended cohorts will finally be available shortly.

1121M

Homozygosity mapping and candidate gene screening in Attention Deficit/Hyperactivity Disorder (ADHD) in Highly Inbred Saudi Arabian Families. *F. Alnaemi[1], D. Abebe[2], A. Adi[1], A. Almagrashi[1], A. Tahir[1], S. Qasem[1], M. Ghaziuddin[2], N. AlTassan[1].* 1) King Faisal Specialist Hospital & research Center, Riyadh, Saudi Arabia; 2) Psychiatry Department, King Faisal Specialist Hospital& research Center.

Attention Deficit/Hyperactivity Disorder (ADHD) is the most common neurodevelopmental disorder affecting at least 7% of all children in Saudi Arabia. ADHD, is not considered to have a single genetic cause but rather a complex interaction of genetic and environmental factors. Association studies have identified a number of susceptibility loci on 4q13.2, 5q33.3, 7p, 11q22,14q12 and 17p11. Moreover, meta-analysis of candidate genes has implicated genes for dopamine, norepinephrine, and serotonin neurotransmitter systems in the etiology of ADHD. The objective of this study is to identify ADHD susceptibility in highly consanguineous and inbred families of Saudi Arabia. To achieve this goal genotyping using high-resolution chip array and candidate gene sequencing has been performed. Regions of homozygosity on chromosomes 3,4,5,6, 7,11,10,13,17 and 18 were identified in different families . Subsequent bidirectional sequencing of candidate genes in these regions identified a number of novel and previously reported single nucleotide variants (SNV's) in *DDR1*, *GAB2*, *INTS4*, *ALG8*, *GDPD4*, *NRM* and *BTNL2* genes. None of these variants segregated with the autosomal recessive inheritance in these families. The large number of LOH regions identified and the lack of segregation of variants identified in candidate genes reinforces the complex genetics of ADHD, and points to other possible genetic mechanisms as an explanation to the phenotypes observed.

1122T

Linkage analysis, homozygosity mapping and whole exome sequencing to identify new genes in consanguineous families with juvenile myoclonic epilepsy. *B. ouled amar bencheikh[1,2,3], F. Lahjouji[2], K. Khaldi[2], H. Lamghari[2], D. Spiegelman[3], A. Dionne Laporte[3], H. Belaidi[2], P. Dion[3], R. Ouazzani[2], G. Rouleau[3], P. Cossette[1].* 1) Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Notre Dame Hospital, University of Montreal, Montreal, Quebec, Canada; 2) Clinical Neurophysiology Department, Speciality Hospital, CHU Ibn Sina, Rabat, Morocco; 3) Montreal Neurological Institute and Hospital, Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada.

Background. The genetic (idiopathic) generalized (GGEs) epilepsies account for 40 to 60% of the etiologies of epilepsy. Most of these syndromes have a complex transmission, with phenotypic and genetic heterogeneity as well as incomplete penetrance of the causative genes. However, despite this complexity, many genes and loci responsible for epilepsies have been identified. The identification of these loci and genes has been possible by genetic linkage studies and recently by whole exome sequencing. Methods. To identify new genes of GGEs, we characterized clinically three consanguineous families from Morocco with juvenile myoclonic epilepsy (JME) and autosomal recessive transmission. Genotyping of each family subjects was done by Illumina Omni-express SNP chips for the linkage mapping. The Multipoint linkage analysis was carried out with Merlin by using an autosomal-recessive model with complete penetrance. The homozygosity mapper was used to identify intervals of homozygosity. Then, the exome sequencing was performed for 8 patients from the three families. The exome capture was done by the Agilent SureSelect V4 and the sequencing on Illumina HiSeq2000. The annotation and the calling of the variants was done by using an in-house data pipeline using the GATK software. Resutls. We identify one locus in each family confirmed by homozygosity mapper. Then, we prioritize the variants identified by exome sequencing within these intervals of homozygosity. We filtered the variants considering the frequency of the disease at 1% and excluded all variants with frequency greater than 1% in public and in our In-house database (2000 exomes). We have identified a total of 328869, 281434 and 325231 shared variants in family 1, 2 and 3 respectively. Some of these variants segregate well with the disease. All the variants filtered were validated by Sanger sequencing. Conclusion. The use of Linkage analysis and homozyosity mapping, combined to whole exome sequencing in consanguineous families with juvenile myoclonic epilepsy have great potential to identify the causative genes of this disease and provide us novel therapeutic issue for epilepsy.

† Trainee Award Finalist

1123S

Genetic characterization of a homozygous 9p deletion in a patient with hyper IgE syndrome and progressive multifocal leukencephalopathy supports deficiency of DOCK8 as a causal factor for both diseases. *C. Sun[1], A. Day-Williams[1], H. McLaughlin[1], I. Jelcic[2], T. Harris[1], R. Martin[2], J. Carulli[1].* 1) Translational medicine, Biogen Idec, Cambridge, MA; 2) Department of Neurology, University Hospital Zurich, Frauenklinikstrasse 26, Zurich, Switzerland.

A 30 year old male Caucasian patient with eczematoid dermatitis since early childhood was found having massively elevated serum immunoglobulin E (IgE) level (26'800 kU/L), and slightly increased peripheral blood eosinophiles (1.6 x 109/L). Therefore, he was diagnosed with hyper IgE syndrome (HIES). In fall 2008, the patient developed progressive neurological deficits caused by large intracerebral lesions as detected by MRI, suggestive of progressive multifocal leukencephalopathy (PML), an infection of the brain caused by JC virus (JCV). CSF analysis showed presence of JCV DNA (500 copies/mL), proving PML. HIES-associated CD4+ and CD8+ T cell lymphopenia was detected repeatedly. Family history revealed that the parents of the patient are first cousins.

Whole genome sequencing and microarray analysis revealed large deletions on both copies of chromosome 9, near the p telomere. The sequence at 9ptel as determined by NGS technology (Complete Genomics Incorporated) is somewhat ambiguous due to the existence of multiple copies of chromosome 9p paralogous sequences in the human genome, and the repetitive sequence nature near telomere. Therefore, we defined the breakpoint using focused molecular genetic assays. Using WGS and whole genome SNP array data as guidance, we PCR-tested multiple chromosome 9ptel-specific SNPs near the potential telomeric end of the break points, on this PML patient and controls. Analysis showed that he has a homozygous deletion and that the breakpoint is very close to telomere repetitive sequence. Cloning and dideoxy sequencing of a long range PCR product from primers flanking the hypothetical centromeric and telomeric breakpoints defined them unambiguously at the nucleotide level (10,047-586751). This interval includes 7 known genes: WASH1, FAM138C, FOXD4, CWBD1, C9orf66, KANK1, and DOCK8, with DOCK8 being the primary candidate for causing immune dysfunction in this patient. DOCK8 deficiency causes a pleiotropic combined T and B cell immunodeficiency characterized by recurrent cutaneous viral infections, susceptibility to cancer and elevated serum levels of IgE. Additionally, cases of PML in DOCK8 deficiency have been described elsewhere. Our data provide further evidence that DOCK8 deletion is causal to HIES and predisposes to PML development.

1124M

Next-generation sequence analysis of neurodegeneration on Guam. *I. Guella[1], J. Steele[2], C. Szu-Tu[1], M.K. Lin[1], C. Thompson[1], D.M. Evans[1], H.E. Sherman[1], C. Vilariño-Güell[1], K. Gwinn[3], H. Morris[4], D.W. Dickson[5], D.S. Fenton[6], M.J. Farrer[1].* 1) Djavad Mowafhagian Centre for Brain Heath, University of British Columbia, Vancouver, BC, Canada; 2) Guam Memorial Hospital, Tamuning, Guam; 3) National Institute of Neurologic Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA; 4) Institute of Neurology, Queen Square, University College London, London, England; 5) Department of Neuroscience, Mayo Clinic Jacksonville, Florida, USA; 6) Anthropology Program, University of Guam, Mangilao, Guam.

The amyotrophic lateral sclerosis (ALS)/Parkinsonism-dementia complex (PDC) is a progressive, age-associated neurodegenerative disorder that has been described in Guam, Western Papua and the Kii peninsula of Japan. The etiology and pathogenesis of this complex syndrome remains enigmatic despite decades of research. In this study we have used next-generation targeted sequencing to evaluate the contribution of genetic variability in the pathogenesis of Guamanian ALS/PDC. Thirty-two genes previously linked and/or associated with parkinsonism, dementia, ALS and related neurodegenerative syndromes were sequenced in 101 indigenous Chamorro Islanders including patients with ALS, parkinsonism and/or dementia, and neurologically normal subjects. Pathogenic mutations identified include HTT (42 CAG repeats), homozygous PINK1 (p.L347P), heterozygous DCTN1 (p.T54I) and FUS (p.P431L). The findings help explain the clinical, pathologic and genetic heterogeneity observed in some multi-incident families and contribute to the excess incidence of neurodegeneration reported.

1125T

Sex differences in neuropsychiatric expression of rare deletions overlapping schizophrenia susceptibility gene ZNF804A. *C. Lowther[1,2], A.C. Lionel[3,4], C.R. Marshall[3,4], S.W. Scherer[3,4], A.S. Bassett[1,2,5].* 1) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 2) Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 3) The Centre for Applied Genomics and Program in Genomics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 5) Department of Psychiatry, University of Toronto and University Health Network, Toronto, Ontario, Canada.

Intellectual disability (ID), autism spectrum disorder (ASD), and schizophrenia each have a strong genetic component and significant sex differences. There is a 4:1 male sex bias in ASD, an excess of males institutionalized for ID, and men with schizophrenia tend to have a more severe course of illness compared to females. X-linked disorders have long been posited to contribute in part to the observed sex bias in ASD and ID and new studies suggest males manifesting these disorders may be more susceptible to the damaging effects of copy number variations (CNVs). Few studies as yet have investigated sex bias in the expression of a rare autosomal CNV. We recently identified two distantly related probands ascertained for schizophrenia with a very rare 790 kb deletion overlapping ZNF804A, a top candidate gene for schizophrenia. Using family history and locally available census data we were able to trace the origin of this inherited deletion. We performed in-depth phenotyping of several relatives of these probands including comprehensive medical, psychiatric, and physical assessment to identify major lifetime features and illnesses. DNA samples for ZNF804A deletion testing were collected. We report on results from a large extended family within which two nuclear families could be distinguished. Of the nine adults (6 females, 3 males) tested, six (5 females, 1 male) were found to have the ZNF804A deletion: two females were unaffected with respect to neuropsychiatric conditions while the remaining three had schizophrenia and intellectual disability, personality disorder, and major depressive disorder, respectively. The one male proband with the ZNF804A deletion was diagnosed with schizophrenia and learning difficulties. A second male without the deletion was also diagnosed with schizophrenia. To our knowledge this is the first study to report on the broader neuropsychiatric phenotypes associated with rare deletions overlapping ZNF804A. It appears that the deletion increases the risk for a range of neuropsychiatric conditions in females, though not all females manifest disease. The recurrence of schizophrenia in two males (one without the deletion) in this family suggests that additional factors are likely contributing to the variable expression and incomplete penetrance. Further study, including next generation sequencing, will be critical for understanding the sex related differences in neuropsychiatric expression associated with the ZNF804A deletion.

1126S

Clinically Relevant Candidate and Known Genes for Alcoholism with Representation on High Resolution Chromosome Ideograms. *A.M. Manzardo, A. McGuire, M.G. Butler.* Psychiatry & Behavioral Sciences and Pediatrics, University of Kansas Medical Center, Kansas City, KS., USA.

Alcoholism arises from the combined effects of multiple biological factors including a wide range of possible genetic variations and/or abnormalities, as well as non-genetic causes including interpersonal and psychosocial relationships, and gene/environmental interaction (epigenetics) with the risk of development strongly related to family history, childhood behavioral problems and the presence or absence of life stressors particularly in childhood. The interface of these relationships are complex and involve overlapping and competing effects of possibly hundreds of genes impacting brain development, structure, and function, as well as, alcohol processing and sensitivity. Years of research on alcoholism has generated a wide range of data linking various genes and pathways to the causation and pathophysiology of alcoholism. The list of alcoholism-related genes has significantly increased recently due to better awareness and advances in genetic technology along with expanding searchable genomic databases. Candidate genes for alcoholism related to neurotransmitter pathways are associated with brain reward processes including the dopaminergic (e.g., *DRD2, MAOA*, and *COMT*), serotoninergic (e.g., *HTR3A, HTR1B, HTR3B*, and *SLC6A4*), GABAergic (e.g., *GABRA1, GABRA2*, and *GABRG1*), glutaminergic (e.g., *GAD1, GRIK3*, and *GRIN2C*) and opioid (e.g., *OPRM1, OPRD1*, and *OPRK1*) pathways. These genes presumably influence the reinforcing properties of alcohol and drive the motivation to seek and use alcohol to excess. We have compiled the latest list of highly validated and clinically relevant alcoholism genes from the medical literature and related websites dedicated to gene discovery and characterization. Our list of over 160 genes consisted of the most clinically relevant with symbols placed on high resolution human chromosome ideograms enabling researchers, physicians, and counselors to have a convenient visual image of the location of a particular alcoholism gene and the flanking alcoholism genes therein. Meaningful correlations of the observed phenotype with the suspected/detected alcoholism gene(s) at the chromosome region/breakpoint defect can be used to inform diagnosis, treatment and improve outcomes with gene-based personalized genetic counseling risk assessments and information for families.

1127M

Association analysis of *HLA-DQB1* gene with narcolepsy without cataplexy and idiopathic hypersomnia. *T. Miyagawa[1], H. Toyoda[1], H. Kojima[2], T. Futagami[2], SS. Khor[1], A. Hirataka[1], M. Yamasaki[1], H. Saji[2], Y. Honda[3], M. Honda[3, 4], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) HLA Foundation Laboratory, Kyoto, Japan; 3) Japan Somnology Center, Neuropsychiatric Research Institute, Tokyo, Japan; 4) Sleep Disorders Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan.

Narcolepsy without cataplexy (NA w/o CA) is a lifelong disorder characterized by excessive daytime sleepiness and rapid eye movement (REM) sleep abnormalities, but does not exhibit cataplexy. Several studies have reported that NA w/o CA is associated with human leucocyte antigen *HLA-DQB1*06:02* similar to narcolepsy with cataplexy (NA-CA). The sample sizes of the studies were relatively small because NA w/o CA is an infrequent condition, which makes it difficult to find a large number of samples. We have therefore formed a collaborative research group to promote the study in Japan. In the present study, we examined *HLA-DQB1* in 146 Japanese patients with NA w/o CA and 1,418 control subjects. The frequency of *DQB1*06:02* in the patient group was significantly higher than that in the control group (allele frequency: 16.8% vs. 7.6%, $P = 1.6\times10^{-7}$, OR = 2.40; carrier frequency: 31.5% vs. 14.7%, $P = 1.5\times10^{-7}$, OR = 2.68). After controlling for the effect of *DQB1*06:02*, distributions of *HLA-DQB1* alleles were compared between NA w/o CA and NA-CA to assess whether the genetic backgrounds of the two diseases have similarities. The distribution of *HLA-DQB1* alleles in *DQB1*06:02*-negative NA w/o CA was significantly different from that in NA-CA ($P = 8.4\times10^{-7}$). On the other hand, the patterns of the *HLA-DQB1* alleles were similar between *DQB1*06:02*-positive NA w/o CA and NA-CA. *HLA-DQB1* analysis was also performed in 171 Japanese patients with idiopathic hypersomnia (IHS). No significant associations were observed between IHS and *HLA-DQB1* alleles. We found a significant difference in *HLA-DQB1* allele distribution between IHS and NA-CA ($P = 2.0\times10^{-8}$). The findings suggest that *DQB1*06:02*-positive NA w/o CA has an autoimmune pathogenesis in common with NA-CA, but *DQB1*06:02*-negative NA w/o CA and IHS might be unique from NA-CA.

1128T

Targeted Sequencing of Candidate Genes Identified with Exome sequencing in Multiplex Autism Families. *A. Patowary, I. Stanaway, R. Nesbitt, W. Raskind, D. Nickerson, Z. Brkanac.* University of Washington, Seattle, WA.

Autistic disorder is a severe neuropsychiatric disorder with strong genetic component. Recent large scale exome sequencing studies have demonstrated the importance of de-novo mutations in Autism. For familial autism, the contribution of rare highly penetrate genetic variants is not well understood. Identification of genes responsible for familial autism is important as genetic architecture of familial autism might differ from de-novo autism. To identify the contribution of rare variants to familial autism we performed whole exome sequencing in 25 NIMH and University of Washington autism families that include affected cousins and thereby minimizing variant sharing between affected family members. We define candidate variants as variants that are private (not present in dbSNP132, 1000 genomes or ESP6500) and functional (coding, missense, stop-gain, stop-loss, splice). For each family between 2 and 25 variants met our selection criteria. Based on overlap with de-novo variants and literature searches we have selected 25 candidate genes for further evaluation. To establish the association of the identified candidate genes with autism we have performed a large gene-based case control study. For all candidate genes whose protein coding sequence encompasses more than 50 Kbp, we have performed Molecular Inversion Probe (MIP) capture and targeted sequencing in 960 NIMH and 288 UW familial autism cases and 960 NIMH unscreened controls. The variant calling and annotation were performed in the same manner as for exome sequencing with modifications due to MIP capture and multiplexing to ensure comparative coverage representations of the genes in both case and control samples. Enrichment of rare variants in the affected subjects was evaluated using plink/seq tools. Our preliminary analysis has identified one novel gene for which rare functional variants are more common in cases as compared to controls ($P<0.01$). Detail analysis of the of the case-control study and analysis will presented.

1129S

Further evidence for DLGAP2 as an ASD/ ID candidate gene. *H. Poquet[1, 3], L. Faivre[1, 5], S. El Chehadeh[1, 5], J. Morton[10], H. Goel[11], B. Isidor[9], C. Lecaignec[8], J. Andrieux[7], B. Delobel[6], M. Lefebvre[1, 5], C. Juif[2], A. Collinet de la Salle[2], N. Lagarde[2], C. Henry[2, 3], N. Marle[4, 5], P. Callier[4, 5], A. Mosca-Boidron[4, 5].* 1) Centre de Génétique et Centre de Référence Maladies Rares 'Anomalies du Développement et Syndromes Malformatifs de l'Interrégion Est', Hôpital d'Enfants, CHU Dijon, France; 2) Centre de Ressources Autisme de Bourgogne, Hôpital d'Enfants, CHU Dijon, France; 3) Service de Pédopsychiatrie, Hôpital d'Enfants, CHU, Dijon, France; 4) Laboratoire de Cytogénétique, Plateau Technique de Biologie, CHU Dijon, France; 5) EA 4271 GAD « Génétique des Anomalies du Développement », IFR 100 - Santé STIC, Université de Bourgogne, Dijon, France; 6) Laboratoire de Cytogénétique, Centre de Génétique Chromosomique, Hôpital Saint-Vincent de Paul, Groupe Hospitalier de l'Institut Catholique de Lille; 7) Institut de Génétique Médicale, Pôle de Biologie Pathologie Génétique, Centre Hospitalier Régional Universitaire de Lille; 8) Unité de Génétique Chromosomique ou Cytogénétique, Service de génétique médicale, Institut de Biologie, CHU Nantes; 9) Unité de Génétique Clinique, Service de Génétique Médicale, Institut de Biologie, CHU Nantes; 10) Clinical Genetics Unit, Birmingham Women's Hospital, Edgbaston, Birmingham, UK; 11) Hunter Genetics, Newcastle, Australia.

The Autism Spectrum Disorders (ASDs) comprise a range of early-onset persistent neuro developmental conditions, characterized by significant impairments in reciprocal social interaction and communication, accompanied by repetitive restricted stereotyped behaviors. The core symptom of ASD typically coexists with other medical conditions such as Intellectual Disability (ID) and seizures. The involvement of rare (<1% frequency) copy number variations (CNV) of varying expressivity and penetrance as risk factors in ASD/ID phenotypes has been highlighted previously in large series of patients. The DLGAP2 (discs, large (drosophila) homolog-associated protein 2) gene, whose glutamatergic postsynaptic density product may play a role in synaptogenesis and plasticity, has been identified as a novel candidate gene on the basis of a de novo 8p23.3 duplication intersecting DLGAP2 observed in a sporadic nonsyndromic ASD male. Another patient with sporadic de novo duplication involving DLGAP2 and ASD/ID had been previously reported in the literature. It has also been suggested that increased DLGAP2 gene expression may contribute to the pathogenesis of schizophrenia. Based on these results and after fine phenotyping of another patient with a de novo 8p24.3p23.2 1.8 Mb duplication involving DLGAP2 and presenting with ASD intersecting schizophrenia, we gathered an international series of six additional cases via the Decipher database and diagnosed with array-CGH. This series comprised a family with 3 females in which the duplication segregated with mild ID but without ASD features, and 3 sporadic males with ID and ASD with a duplication inherited from an asymptomatic parent. This study supports the hypothesis that rare CNV encompassing DLGAP2 gene could predispose to ASD/ID, with the principle of incomplete penetrance and variable expressivity. It also further supports the existence of common predisposing factors to ASD, ID and schizophrenia.

1130M

Structural equation models of communication endophenotypes suggest that human vocalized speech has a polygenic basis. *C.M. Stein[1,2], A. Avrich Ciesla[1], Y. Song[1], L. Freebairn[3], B. Truitt[1], H.G. Taylor[5], B.A. Lewis[3], N.J. Morris[1], S.K. Iyengar[1,4].* 1) Epidemiology & Biostatistics, Case Western Reserve Univ, Cleveland, OH; 2) Center for Proteomics & Bioinformatics, Case Western Reserve Univ, Cleveland, OH; 3) Psychological Sciences, Case Western Reserve University, Cleveland, OH; 4) Genetics, Case Western Reserve University, Cleveland, OH; 5) Pediatrics, Case Western Reserve University, Cleveland, OH.

The genetic basis of vocalized speech, hypothesized to be a human-limited trait, is unknown. We used extensive phenotype and genotype data from children and families with disordered speech to demonstrate that vocalized speech is set in a complex cognitive network connected through many endophenotypes. Speech sound disorder (SSD) is one of the most common types of communication disorders, with prevalence rates of 16% at 3 years of age, and an estimated 3.8% of children continuing to present speech delay at 6 years of age. SSD is manifested through several correlated endophenotypes, including phonological memory, phonological awareness, vocabulary, speed of processing, oral motor skill, reading decoding, spelling, and spoken language; many of these same domains are affected in language impairment and dyslexia. Several studies have identified promising genetic associations between communication disorders and specific genes, but have not accounted for the correlation among endophenotypes. Previously, we developed a structural equation model (SEM) that depicted how endophenotypes assessed at preschool age influenced endophenotypes at school-age (Lewis et al. 2011). However, that SEM did not explicitly model familial structure and also did not incorporate genetic effects. We have recently released 'strum' software, which implements the SEM framework for family data (Morris et al. 2010). Here, we have refined the speech sound SEM with strum, and identified significant polygenic influences on phonological awareness (p=0.02), phonological memory (p=0.04), and the correlation between phonological memory and phonological awareness (p=0.02) and between phonological awareness and speeded naming (p=0.03). Next, we incorporated candidate gene association data into the model, based on single gene associations we previously observed with DRD2, AVPR1A, and ASPM (Stein et al. 2014), and found that DRD2 has an independent influence on phonological memory (p=0.004) after accounting for polygenic and residual correlations among these phenotypes. By synthesizing comprehensive endophenotype and genetic data with SEM, these findings reveal the genetic complexity of SSD and other communication and learning disorders.

1131T

Second thought about 16p12.1 microdeletion syndrome: is two-hit a valid model? *F. Boyar[1], L. Ross[1], J. Kelly[2], J.C. Wang[1].* 1) Dept Cytogenetics, Quest Diagnostics, Nichols Inst, San Juan Capistrano, CA; 2) Dept Cytogenetics, Quest Diagnostics, Nichols Inst, Chantilly, VA.

The chromosome 16p12.1 deletion syndrome (520kb; OMIM #136570) was suggested to be associated with high incidence of carrying additional large copy-number variants (10 of 42, 24%) as per Girirajan, S et al, 2010. The presence of 16p12.1 microdeletion was considered to predispose patients to develop neuropsychiatric phenotypes, which were exacerbated by epigenetic, environmental or genetic abnormalities such as other large deletions or duplications. A cohort of 37 cases with 16p12.1 microdeletion was studied by oligo-SNP array (CaytoScan HD, Affymetrix) at Quest Diagnostics Nichols Institute. The patients' age ranges from newborn to 23 years old. A wide spectrum of phenotypic presentation was found, including intellectual disability, developmental delay, learning disability, cranio-skeletal anomalies, microcephaly, speech delay, seizures, endocrine abnormalities, and autism. All the deletions spanned the 520 kb commonly deleted region, varying from 478 kb to 909 kb. Four cases had small copy number variants of unclear clinical significance, and only two cases (2/37, 5%) had a large or clinical significant copy number variant. One case had a 1.7 Mb gain of 22q11.2 at the DiGeorge syndrome critical region (OMIM #608363), and another had a 580 kb loss involving multiple exons of the NRXN1 gene (OMIM #600565). Our findings are unable to confirm the two-hit model as previously proposed. The absence of other large or clinically significant microdeletions or microduplications in 95% (35/37) of cases from this cohort demonstrates that the loss at 16p12.1 microdeletion alone may give rise to neuropsychiatric phenotypes as a single event.

1132S

DUF1220 domain copy number is linearly associated with increased speech delay in individuals with autism. *J.M. Davis[1], T.E. Fingerlin[2], V.B. Searles[1], L. Dumas[1], M. O'Bleness[1], J.M. Sikela[1].* 1) Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO; 2) Department of Epidemiology, Colorado School of Public Health, Aurora, CO.

Purpose: Autism is a neurodevelopmental condition characterized behaviorally through social and communicative impairments and repetitive behaviors. DUF1220 protein domain copy number is dramatically increased specifically in the human lineage and has been implicated in brain growth and brain evolution. We previously demonstrated that, in individuals with autism, DUF1220 (subtype CON1) copy number is linearly associated, in a dose response manner, with increasing severity of each of the three core symptoms of autism. Given these findings we investigated associations of CON1 copy number and speech delay, a common marker of severity, in an additional independently assayed group of individuals with autism. Methods: DNA samples from an additional 50 non-Hispanic white unrelated individuals with autism were obtained from the Autism Genetic Resource Exchange (AGRE) and assayed by custom high density arrayCGH. Age of first meaningful word was obtained by the Autism Diagnostic Interview-Revised, a caregiver interview used by clinicians to assist with autism diagnostics. Initially age of first meaningful word was dichotomized where greater than 24 months was classified as delayed first word. Unequal variance t-tests were performed to assess mean differences in DUF1220 copy number between delayed first word and not delayed groups. This was followed with a linear model examining the relationship of CON1 with month of speech acquisition adjusted for sex, age and simplex vs multiplex status. Results: Age of first word in this group ranged from 10 to 66 months with a mean of 28.7 months. On average individuals with delayed first word had a 0.065 (p=0.030) higher CON1 copy ratio (copy number) than the group who did not have a delayed first word. Similarly a linear association was identified where increased copies of CON1 were progressively associated with greater speech delay (p=0.019). Conclusions: Using ddPCR we previously implicated the dosage of DUF1220 CON1 in the severity of the core symptoms of autism. Here we have extended these findings to include speech delay. This partially replicates our previous work in a group of individuals independently drawn from AGRE and assayed with different methods. Taken together, these findings lend further support to the view that DUF1220 CON1 dosage is an important contributor to autism severity, and that the same gene family implicated in human brain expansion is also involved in the symptoms of autism.

1133M

First report on a multiplex, consanguineous, family with autism and chromosome 15 duplication. *H. Mansour[1], D. Eltabei[2], O. Shahin[2], F. El-Chennawi[3], U. Urvashi[4], L. Hoffner[4], M. Sathanoori[4], V. Nimgaonkar[1,5].* 1) Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; 2) Department of Psychiatry, Cairo University, Cairo, Egypt; 3) Department of Psychiatry, Mansoura University, Mansoura, Egypt; 4) Pittsburgh Cytogenetics Laboratory, Department of Pathology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; 5) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.

BACKGROUND: Several genetic risk factors have been detected for Autism spectrum disorders (ASD), including copy number variations (CNVs) such as duplications of the chromosomal 15q region. Maternal deletions or inactivation of genes in this region leads to Angelman syndrome, but if paternal, will cause Prader-Willi syndrome. Consanguinity is reported to be associated with the risk as well. METHOD: We report on a consanguineous, first-cousin marriage, Egyptian family having 4 siblings with ASD. We also recruited parents, an unaffected sib and maternal grandfather. Diagnoses were made following structured interviews, including the ADI-R. Array Comparative Genomic Hybridization (aCGH), physical examination by clinical geneticist, IQ testing and EEG were conducted. Co-morbid seizure, intellectual disability or ADHD have been reported in affected siblings. RESULTS: aCGH indicates that 3 of the 4 affected children have maternally-inherited interstitial duplication 15q11.2-q13. The mother and 3 affected siblings have intellectual disability. Several facial dysmorphic features are shared between mother and affected children, including the autistic child who does not have the duplication. CONCLUSION: To our knowledge, this is the largest multiply affected, inbred family reported to date with 15q interstitial duplication. However, the duplication does not segregate completely with ASD in this family, suggesting the presence of other risk factors, possibly related to deleterious mutations unmasked by regions of homozygosity noted extensively. Further analyses, including phenotypic characterization and whole exome sequencing are warranted.

1134T

Screen for somatic mosaic mutations in unexplained Dravet syndrome patients. *CT. Myers[1], JM. McMahon[2], IE. Scheffer[2], HC. Mefford[1].* 1) University of Washington, Department of Pediatrics, Division of Genetic Medicine University of Washington, Department of Pediatrics, Division of Genetic Medicine, Seattle, WA, USA; 2) University of Melbourne, Department of Medicine, Florey Institute of Neurosciences and Mental Health, Austin Health, Melbourne, Australia.

Dravet syndrome is an infantile-onset epileptic encephalopathy characterized by refractory seizures, cognitive arrest or regression associated with ongoing epileptic activity, and a poor prognosis. 80% of Dravet cases are caused by heterozygous mutations in SCN1A, a gene that encodes the voltage-gated sodium channel alpha 1 subunit. Our lab has recently identified GABRA1 and STXBP1 as novel genetic causes of Dravet syndrome (Carvill et al, 2014), explaining a few additional cases. Despite these advances, in almost 20% of patients with Dravet syndrome, the molecular diagnosis is still unknown. Given the strong correlation between genotype and phenotype in solved Dravet cases, we hypothesize that somatic mutations in these known genes (SCN1A, GABRA1, and STXBP1) are causative in a subset of currently unsolved cases. Somatic mutations, well recognized as disease causing in cancer, are implicated in more than 30 monogenetic disorders. We propose to screen for somatic mutations in a cohort of 51 patients with unexplained Dravet syndrome that have already undergone candidate gene screening or exome sequencing. Targeted sequencing using single-molecule Molecular Inversion Probes (smMIPS) will be used to screen for low-level somatic mosaicism in DNA from blood and buccal swabs. The smMIPs technology has been used to detect allele frequencies as low as 0.2% (Hiatt et al., 2013) and searching for mutations in another tissue type such as buccal cells may yield additional mutations not detected in blood. We have collected 51 blood samples, 10 buccal swabs, and are actively collecting more samples. We are also in the process of testing the smMIPS capture protocol and analysis pipeline for variant calling. It is likely that these studies with identify previously undetected mutations that will explain a subset of Dravet diagnoses.

1135S

Genome-wide linkage analyses of non-Hispanic White families identifies several novel loci for familial late-onset Alzheimer's disease. *J. Jaworski[1], B.W. Kunkle[1], C. Reitz[2], A.C. Naj[3], L.S. Cantwell[3], A. Partch[3], G.W. Beecham[1], E.R. Martin[1], W. Raskind[4], R.P. Mayeux[2], J.L. Haines[5], L.A. Farrer[6], G.D. Schellenberg[3], M.A. Pericak-Vance[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; 2) Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY, USA; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 4) School of Medicine, University of Washington, Seattle, WA, USA; 5) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH USA; 6) School of Medicine, Boston University, Boston University, MA, USA.

While more than two dozen loci that contribute to late-onset Alzheimer disease (LOAD) have been identified, few high penetrance variants (e.g., PSEN2 mutations) that explain risk in families heavily burdened with LOAD have been found. Linkage/identity-by-descent (IBD) sharing analyses in large multiplex pedigrees represent a well-powered set of designs to isolate regions and loci of interest for sequencing studies. In order to identify highly-shared regions among multiply-affected pedigrees for prioritized follow-up in the Alzheimer's Disease Sequencing Project (ADSP), we performed extensive parametric and non-parametric multipoint linkage analysis on 386 individuals in 42 non-Hispanic white (NHW) families (5-15 cases per pedigree).The 42 NHW families selected met three criteria: 1) exhibiting dominant inheritance of LOAD; 2) no mutations at known familial AD loci; and 3) low occurrence of the APOEε4 allele. All were among families selected for whole genome sequencing (WGS) through the ADSP and had available genome-wide genotyping of common variants on either the Illumina HumanHap 550 or 1M arrays. Linkage analyses were performed using MERLIN, and follow-up analyses of multipoint loci with NPLall/NPLpairs>2 included IBD sharing with Olorin in order to narrow regions of interest.Two-point parametric linkage analysis identified 16 regions with suggestive linkage (HLOD>3.5), including four regions (Chr3q25, 3q28, 4q34.3, and 11q13.4) with HLOD>4. Non-parametric multipoint analyses found two additional loci at 14q32.2 (LOD= 3.73) and 1q24.1 (LOD=2.31), and supported the 4q34.3 two-point LOD score (LOD=2.51). Haplotype analysis supported segregation in the top families for these regions. 11q13.4 is a newly reported linkage region and includes the gene SHANK2 which has been linked to regulation of synaptogenesis. IBD analyses for the strongest multipoint signal (14q32.2) implicated the gene BCL11B, a transcription factor and regulator of BDNF signaling, a pathway associated with pathogenesis of several neurodegenerative diseases. While these analyses identified several regions of linkage previously observed in AD, including 1q24.1, 3q25, 3q28, 4q34.3, and 14q32.2, we produced new evidence for linkage at 11q13.4 and narrowed a previously reported signal at 14q32 to the candidate gene BCL11B. Prioritizing these regions for follow-up in the ADSP sequence data may provide opportunities to identify highly penetrant variants for familial LOAD.

1136M

Microsatellites in the 5' flanking region of AVPR1A were associated with social behavior scales of autism spectrum disorders in the Korean population. *J. PARK[1], S. Kim[3], H. Ghim[4], Y. Jung[1], H. Yoo[1,2].* 1) Seoul National University Bundang Hospital, Biomedical Research Institute, Gyeonggi-do, South Korea; 2) Seoul National University College of Medicine, Seoul, South Korea; 3) Department of Pharmacology, Eulji University College of Medicine, Daejon, South Korea; 4) Department of Psychology, Chungbuk University, Chungju, Korea.

Background: Impairment in social interaction and communication is core features in autism spectrum disorder (ASD). The arginine vasopressin receptor 1A gene (AVPR1A) is widely expressed in the brain and is considered to be a key mediator for regulation of social behavior. Evidence from numerous organisms implicates relationship between microsatellites of the 5'-flanking region in the arginine vasopressin receptor 1A gene (AVPR1A) and social behavior and genetic variation at AVPR1A has been reported to be associated with autism. The objective of this study is to evaluate the relationship between microsatellites in the 5' flanking region AVPR1A and specific social phenotypes of ASD. Methods: Two microsatellites (RS3 and RS1) in the 5' flanking region of AVPR1A were examined in 218 Korean family trios comprising children with ASD and their biological parents. Behavioral phenotypes are derived from comprehensive measures of the behaviors, using Social Communication Questionnaire (SCQ), Asperger Syndrome Diagnostic Scale (ASDS), Social Responsiveness Scale (SRS), Child Behavior Checklist (CBCL), Autism Diagnostic Observation Schedule (ADOS), Autism Spectrum Quotient (AQ), Empathy Quotient (EQ) and Systemizing Quotient (SQ). In the family-based association test and haplotype analysis using FBAT, we tested association between microsatellites and 28 quantitative traits related with social behaviors. Potential confounding effects, including age, sex and IQ, were controlled as covariates. Results: We found a statistically significant association ($P < 0.05$) between microsatellites and multiple scales related with social behaviors. RS1 is significantly associated with 18 phenotypes (p's = 0.049 - 0.004) and RS3 markers with 5phenotypes (p's = 0.043 - 0.012). Both markers are significantly associated with Social domain score in ASDS (RS1: $P = 0.004$; RS3: $P = 0.012$) and AQ (RS1: $P = 0.034$; RS3 = 0.022). Total seventeen phenotypes were significantly associated in haplotype association analysis (p's= 0.049 - 0.009). Conclusion: We observed significant relationship between microsatellites and specific social behaviors in Korean ASDs. Our results support that RS1 and RS2 microsatellite markers AVPR1A gene can be possible candidates for diagnosis for social behavioral dysfunction and autistic trait of ASD.

1137T

X-Chromosomal genetic and epigenetic factors in the etiopathogenesis of ADHD. *J. Kapalanga[1,3,4], D. Wong[2,3], A. Gandy[2,3], S. Cato[4], N. Nwebube[4].* 1) Pediatrics/Genetics, Western University/GBHS, London, ON, Canada; 2) Pediatrics, Dalhousie University, Halifax, NS, Canada; 3) Summerside Medical Centre, Summerside, PEI, Canada; 4) Grey Bruce Health Services, Owen Sound, ON, Canada.

ADHD is a common neurobehavioral disorder and genetic factors are known to play a critical role in its etiopathogenesis. Molecular, twin, family and adoption studies have led to the conclusion that genetic factors contribute 70 to 80 percent susceptibility for ADHD. While certain genes have been linked to ADHD, specific mutations in specific genes on specific chromosomal regions have not been conclusively implicated. In clinic settings it is not uncommon to see families with several affected members. Assessment of ADHD patients often involves genetic testing including fragile X testing, array based comparative genomic hybridization and karyotype analysis. We report on 5 patients who presented with severe clinical symptoms of ADHD and were subsequently found to have Klinefelter syndrome. Patients were identified from a large clinic population who underwent laboratory investigations to determine a genetic cause to explain their neurobehavioral symptoms. Selection clinical criteria for genetic testing included at least four of the following: onset before age 6 years, inattention, hyperactivity, learning disability, aggressive and violent behavior, impulsivity, social difficulties, significant comorbidity, poor response to treatment. A total of 557 patients fulfilled the clinical criteria and underwent genetic testing including molecular testing for fragile X syndrome, array based comparative genomic hybridization and karyotype analysis. Five patients were found to have an XXY karyotype, two were found to have tri-allelic mutation in FMR1. All five patients presented with inattention and learning disability. Marked hyperactivity and aggressive behavior were not observed in all five patients. All five patients required at least one stimulant medication change. Despite heightened intensity of inquiry, genome-wide association studies, meta-analyses and individual candidate gene studies have to date not identified strong and consistent association of a specific gene with ADHD. This suggests genetic heterogeneity and/or polygenic influences on symptomatology. Further, epigenetic mechanisms could also be implicated in ADHD. In all cases ecogenetic factors most likely play a critical role. The findings in this study suggest that genetic and/or epigenetic factors located on the X-chromosome are involved in the etiopathogenesis of ADHD in at least a subset of patients, and that a gene dosage effect is the underlying molecular mechanism in these cases.

1138S

Personality Genetics and Health in Super Seniors. *J.M.T. Nelson[1, 2], A.R. Brooks-Wilson[1, 2].* 1) Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada; 2) Genome Science, Canada's Michael Smith Genome Sciences Centre, Vancouver, British Columbia, Canada.

Background: Few seniors are free of chronic disease and the genetic factors that contribute to long term good health are not well understood. We have established a study of super seniors, individuals who are over the age of 85 who have never reported being diagnosed with Alzheimer disease, cancer, diabetes, cardiovascular disease or pulmonary disease. It is important to determine why some individuals age with many comorbidities and others with none at all. Longevity and health are associated with specific personality traits. Personality consists of a person's emotional, attitudinal and behavioural responses, which are related to neurotransmission in the brain. Genes that encode neurotransmitters, their receptors or affect processes in their metabolism are candidates for study on the effects of personality. Genetic variation in personality-related genes may affect health by influencing behaviour and lifestyle. Hypothesis: Genetic variants in genes involved in neurotransmission or that underlie personality disorders influence healthy aging. Objective: To determine if genetic variation in personality-related genes is associated with the super senior phenotype. Methods: 1) A literature search has been conducted for genes involved in the physiology of personality and personality disorders. Five neurotransmitter-related candidate genes have been chosen. In addition, a literature search was conducted for genome wide association studies (GWAS) of personality traits and personality related disorders. Single nucleotide polymorphisms (SNP) have been selected and include tag SNPs, SNPs that meet genome-wide significance in GWAS of personality traits, and SNPs that were associated with personality traits in more than one study. 2) Genotyping of the SNPs will be performed using the Sequenom MassArray method. Variable number random repeats will also be genotyped using PCR and fragment analysis. 3) We will test for association with the super senior phenotype using a set of 493 European-ancestry super seniors and 431 European-ancestry controls. Identifying personality-related genetic variants that influence healthy aging may lead to pharmacological opportunities to improve long term health in our aging population.

1139M

Gene-based pleiotropy across five major psychiatric disorders. *D.R. Nyholt, H. Zhao.* QIMR Berghofer, Brisbane, QLD, Australia.

Background Studies have indicated genetic overlap between the five disorders in the Psychiatric Genomics Consortium (PGC): autism spectrum disorder (ASD), attention deficit-hyperactivity disorder (ADHD), bipolar disorder (BPD), major depressive disorder (MDD), and schizophrenia (SCZ). In this study, we aimed to identify specific genes overlapping the five psychiatric disorders utilizing a novel gene-based approach. **Methods** Optimized gene-based tests were performed utilizing genome-wide association (GWA) results from the PGC analysis of single-nucleotide polymorphism (SNP) data for the five disorders in 33332 cases and 27888 controls of European ancestry. After accounting for correlation (i.e., non-independence) of gene-based test results due to linkage disequilibrium we examined the significance of the proportion of genes nominally associated across the five disorders. Pathway and network based analyses were performed on the sets of genes significantly overlapping the disorders. **Results** We found highly significant overlapping genes between SCZ and BPD, moderate overlap between SCZ and MDD, SCZ and ASD, MDD and ASD, and ADHD and BPD. After combining disorders as discovery sets, we found significant overlap across SCZ, BPD and MDD, across SCZ, BPD, MDD and ASD, and across BPD, MDD and ASD/ADHD. No significant overlap was observed between the individual adult-onset disorders and ADHD. Pathways implicated by the genes overlapping the adult-onset disorders include MAPK signalling, calcium signalling, dorso ventral axis formation, chemokine signaling, and melanogenesis. Pathways for BPD and ASD include glycosphingolipd biosynthesis globo series, and pathogenic Escherichia coli infection. Pathways implicated by genes overlapping BPD and ADHD include glutathione metabolism, and arachidonic acid metabolism. Additionally, combining gene-based association results across disorders identified numerous genes surpassing our cutoff for genome-wide significance. **Discussion** Utilizing a novel approach, we identified numerous genes associated across multiple psychiatric disorders. Our results extend previous findings from single SNP-based genetic overlap analyses by providing important insight into the likely genes and biological mechanisms underlying the observed genetic correlation and co-morbidity between these major psychiatric disorders.

1140T

Fine mapping of schizophrenia risk locus CSMD1 (rs10503253) in Indonesian samples revealed association with indels. *D.B. Wildenauer[1], A.A.A. Kusumawardhani[2], D.M.B. Wildenauer[1], B. Benyamin[3], S.G. Schwab[4], Indonesian Schizophrenia Genetics Consortium.* 1) University of Western Australia, Perth, Australia; 2) Department of Psychiatry, University of Indonesia, Jakarta, Indonesia; 3) Queensland Brain Institute, Brisbane, Australia; 4) Department of Psychiatry, University of Erlangen-Nuremberg, Germany.

Schizophrenia is a devastating complex genetic disorder with a highly polygenic background. Recent GWAS findings in large combined Caucasian samples suggested rs10503253, located in the intronic region of the CUB Sushi and multiple domains 1 gene (CSMD1) on 8p23.2 as a risk locus for schizophrenia. We hypothesized that risk loci detected in Caucasian samples are also evident in samples from a different ethnic background. Our sample comprised 1067 individuals with schizophrenia and 1111 non-affected controls consecutively ascertained in five Mental State Hospitals in the area of greater Jakarta, Indonesia. Structured interviews (DIP in Bahasa translation) were performed in all cases and consensus diagnosis made according to the DSMIV criteria by psychiatrists. Using a test panel with 374 evenly spaced SNPs (Illumina) all samples were tested for genotyping quality and sampling errors. Principal component analysis revealed ethnical homogeneity of cases and controls with close relationship to East Asian populations (genomic inflation factor 1.02). Association of rs10503253 at 8p23.2 was confirmed (P =0.0086, same marker, same allele). For fine mapping, we selected 22 SNPs in the region (4.17-4.225Mb) and genotyped these using Taqman technology. Lowest P-value was obtained for rs10102283 (nominal p=0.00083, p=0.02 adjusted for multiple testing), which is 2.582kb apart from rs10503253. Next we analysed the region further by imputing with variants identified in this region in Asian individuals by the 1000 genome project. We selected 134 variants with info≥0.9. Four indels (MAF around 0.45) in the association peak (1.4kb) revealed nominal p-values from 10^{-4}-10^{-5} (P ~ 0.005 after adjusting for multiple testing). Odds ratios were in the range of 1.4. CSMD1 is expressed predominately in the brain. The protein plays a role in complement activation and has been found to be a binding site for micro RNAs. Even though the indels are intronic, they may affect structure of the gene with consequences for transcription and thus quantitatively influence the function of the protein in complement activation.

1141S

Transethnic HLA comparison in narcolepsy. *H.M. Ollila[1], J. Faraco[1], J. Ravel[1], F. Han[2], L. Lin[1], J. Hallmayer[1], E. Mignot[1].* 1) Stanford University, Palo Alto, CA., Select a Country; 2) Department of Pulmonary, Critical Care Medicine, Peking University People's Hospital, Beijing, China.

Narcolepsy is an autoimmune sleep disorder, which is characterized by severe daytime sleepiness. Recently, the onset of narcolepsy has been associated with the 2009 H1N1 influenza. In addition, it is strongly associated with Human leukocyte antigen (HLA) DQB1*06:02, and over 98 percent of narcoleptics carry this haplotype. However, the role of other HLA loci has remained unresolved. The aims of this study were to study the contribution of various HLA haplotypes to narcolepsy and clinical phenotypes related to narcolepsy. We used direct HLA genotyping and HLA imputation. The study was performed in Caucasian and Asian samples comprising altogether over 3000 cases and 10000 controls. Clinical phenotypes included age of onset, hallucinations, onset after vs. before 2009, absence of cataplexy, sleepiness and sleep paralysis. We found that DQB1*06:02 effect was strongly modulated by DQA1 genotype so that those carrying DQA1*01:02 had highest risk and those with DQA1*01 non 01:02 were protected from narcolepsy, suggesting allelic competition in narcolepsy. Analysis of clinical phenotypes revealed that cases carrying DQB1*03:01 had nearly two years earlier age of onset. DQB1*06:02 was rarer in individuals that got narcolepsy after 2009. The findings highlight the importance of HLA in the etiology of narcolepsy. Analysis of HLA-A, B, C and HLA-DP loci are ongoing.

1142M

Comparative sequencing of the *PARK2/PARCRG/QKI* locus in Early Onset Parkinson's disease. *W.C. Macedo[1,2], M.E.D. Sauer[1], H.A.G. Teive[3], R.P. Munhoz[4], C.M. Probst[5], M.T. Mira[1].* 1) Graduate program in health sciences, Pontifícia Universidade Católica do Paraná - PUCPR, Curitiba, Paraná, Brazil; 2) Department of Genetics, Universidade Federal do Paraná - UFPR, Curitiba, Paraná, Brazil; 3) Department of Neurology, Hospital de Clinicas - UFPR, Curitiba, Paraná, Brazil; 4) Associação Paranaense dos Portadores de Parkinsonismo - APPP, Curitiba, Paraná, Brazil; 5) Department of Genomic and Bioinformatics, Instituto Carlos Chagas - FIOCRUZ-PR, Curitiba, Paraná, Brazil.

Early Onset Parkinson's disease (EOPD [MIM 600116]) is a neurodegenerative disorder characterized by death of midbrain dopaminergic cells and clinical outcome before 45 years of age. EOPD is under strong genetic control; mutations in the *PARK2* gene have been consistently associated with familial and non-familial autosomal recessive EOPD cases. Interestingly, an attempt to reproduce PD in *Park2* knockout mouse strain has failed. On the other hand, the Qk^{v} mice, which develops vigorous limb shaking a few days after birth, are carriers of a spontaneous, recessive deletion of Quaking (*Qk*), *Pacrg* and *Park2* neighboring genes located at *chr17A*. In humans, these three genes are also adjacent to each other and located at *chr6q25-q27*, homologous to the murine *chr17A*. We hypothesized that mutations in human *QKI* and/or *PACRG*, together or not with *PARK2* mutations, are necessary or sufficient for the development of EOPD. To investigate this hypothesis, we performed comparative sequence analysis of the coding sequences of *QKI*, *PACRG* and *PARK2*, seeking for potentially pathogenic variants. Study sample was composed by two DNA pools: one including 36 EOPD cases and the other, 17 individuals with age higher than 65 years, free of any neurodegenerative disease. Sequencing was conducted as implemented at the Ion PGM platform. Primary data was subjected to a filtering protocol designed to eliminate sequencing artifacts; variants approved by the quality filters were subjected to *in silico* prediction analysis, searching for changes in regulatory sites of mRNA, splicing sites and polypeptide sequence. A total of 82 variants were identified and 38 were validated by the quality criteria (seven new and 31 known). From this set, functional high impact variants were only found in *PARK2*; two of them were selected for further investigation: a nonsense c.1198C>T and a frameshift c.1277delA. These novel mutations, located on exon 11, were found only in the patient pool. In both cases, the functional predicted impact was of complete abolishment of enzyme activity. To validate these mutations, Sanger sequencing was conducted on an expanded sample of 55 controls and 41 patients. As a result, the c.1198C>T mutation was validated in two heterozigous patients. Our findings reveal a novel mutation that may predispose humans to non-familial EOPD, corroborating the evidence that suggests that *PARK2* variants alone are responsible for the role of the *6q25-q27* locus in the pathogenesis of EOPD.

1143T

Strategies for identifying new genes in autosomal recessive Parkinson's disease. *S. Lesage[1,6], A. Nicolas[1], M. Sabahtou[2], V. Drouet[1], C. Pujol[1], F. Cormier[1,3], E. Lohmann[4], M. Tazir[5], A. Dürr[1,6], O. Corti[1], E. Kabashi[1], A.L. Leutenegger[2], A. Brice[1,6], French PDG group.* 1) U1127, INSERM, Paris, Paris, France; 2) Inserm UMR_S946, Fondation Jean Dausset, Paris, France; 3) Centre d'Investigation Clinique Pitié Neurosciences, CIC-1422, Paris, France; 4) Behavioural Neurology and Movement Disorders Unit, Department of Neurology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey; 5) Faculté de Médecine, Université d'Alger, Algier, Algeria; 6) AP-HP, Pitié-Salpêtrière Hospital, Department of Genetics and Cytogenetics, Paris, France.

BACKGROUND: In recent years, understanding of the genetics and the physiopathological mechanisms underlying molecular defects in Parkinson's disease (PD) has greatly advanced, since 10 genes are now confirmed to cause monogenic forms of PD, including 4 (Parkin, PINK1, DJ-1 and ATP13A2) associated with early-onset (EO) PD cases with autosomal recessive (AR) inheritance that account for only ~40% of AR EO PD cases. Thus, a significant proportion of cases with inherited PD remain unexplained genetically. AIM: To identify novel genes implicated in AR forms of PD. PATIENTS AND METHODS: 1) Subjects: We selected 140 EO (<55 years) PD families or isolated cases with known or suspected consanguinity, excluded for Parkin, PINK1, DJ-1 and the LRRK2 G2019S mutation and fulfilled quality and quantity controls; 2) Genome-wide SNP genotyping and homozygosity mapping: SNP genotyping was performed using Illumina HumanCytoSNP-12 BeadChip. Inbreeding coefficients F estimated by FEstim and multipoint genetic linkage analyses were computed with FLOD; 3) Search for rare rearrangements using the same DNA microarrays and the Illumina cnvPartition module; 4) Whole exome sequencing was performed in a subset of 60 PD families with confirmed consanguinity, using Agilent technologies and Illumina Hiseq2000; 5) Validation of 10 best candidate genes by functional testing candidate genes in zebrafish. Morphans are phenotyped for lethality, abnormalities in the body axis, touched-induced and spontaneous motor behaviour and loss of TH+ cells; 6) Replication in an independent set of 200 EO AR PD families and 400 patients with apparently sporadic EO PD using medium-throughput sequencing and establishment of the mutational spectrum and frequency of each new gene and phenotype/genotype correlations. RESULTS: We looked for rare homozygous mutations that are assumed to affect the protein sequence and fell within identified regions of homozygosity. Preliminary analyses of exome data from consanguineous families have led to the identification of deleterious mutations in ATP13A2, FBXO7 in 2 affected sibpairs and a list of 40 best candidate genes. Zebrafish testing for one of them showed a phenotype. CONCLUSION: As a proof of concept, we identified a list of known and novel candidate genes for PD which validation is in progress. The identification of new genes involved in AR PD will represent a major advancement that will permit the development of new molecular diagnostic strategies.

1144S

Could somatic copy number alterations contribute to sporadic Parkinson's disease? *C. Proukakis[1], K. Gancheva[2], C. Grace[2], D. Pease[1], D. Brazma[2], E. Kara[1], J.W. Taanman[1], H. Houlden[1], A.H. Schapira[1], E. Nacheva[2].* 1) Institute of Neurology, University College London, London, United Kingdom; 2) 2Academic Department of Haematology, University College London, United Kingdom.

Objective: to determine feasibility and sensitivity of detecting somatic copy number alterations (CNAs) in Parkinson's disease brain. Background: Parkinson's disease (PD) is usually sporadic, and despite heritability estimates of around 30%, the majority of the cases are unexplained. Alpha-synuclein (SNCA) protein is central to pathogenesis, but *SNCA* missense mutations and copy number variants (CNV) are rare. We have recently hypothesised that mosaicism due to post-zygotic somatic mutations in early embryogenesis in the neuronal cell lineage could contribute to pathology. We have already screened over 500 brain samples for somatic coding *SNCA* mutations, without detecting any. There is, however, recent clear evidence of somatic CNAs in brain, which are methods would not have detected, but their role in diseases remains unknown. Methods: We have designed a custom Agilent CGH array with dense spacing of probes for PD-related genes, including *SNCA* and *PARK2*. Dilution of DNA from a *SNCA* duplication carrier with wild-type to give various levels of increased *SNCA* DNA (thus creating artificial mosaics for CNAs) allows determination of the sensitivity, and comparison of different analytical methods. To determine if interphase cell FISH (fluorescent in situ hybridization) can allow reliable detection of lower level mosaicism, probes for *SNCA* and *PARK2* will first be tested on fibroblasts with known CNVs. Results: Our array can detect 7.5-15% increase of *SNCA* genomic DNA, depending on the analytical settings used. FISH probes for *SNCA* and *PARK2* have been verified in both normal and positive control fibroblasts. Reproducible data can be obtained by FISH analysis of substantia nigra. Conclusions: We have developed methods and tools to allow us to further investigate our hypothesis of somatic CNAs in sporadic PD cases.

1145M

Whole-genome sequencing to identify genes implicated in Familial Parkinsonian Tauopathy. *M. Sanchez-Contreras[1], S. Fujioka[1], C. Pottier[1], A.J. Strongosky[2], B.F. Boeve[3], J.E. Parisi[3], P.M. Tacik[1], N. Aoki[1], M. Baker[1], V. Sossi[4], D.W. Dickson[1], A.J. Stoessl[5], O.A. Ross[1], Z.K. Wszolek[2], R. Rademakers[1].* 1) Department of Neuroscience, Mayo Clinic Jacksonville, FL; 2) Department of Neurology, Mayo Clinic Jacksonville, FL; 3) Department of Neurology, Mayo Clinic, Rochester, MN; 4) Department of Physics and Astronomy, University of British Columbia, BC, Canada; 5) Pacific Parkinson's Research Centre, Division of Neurology, University of British Columbia & Vancouver Coastal Health, BC, Canada.

Tauopathies are a group of neurodegenerative disorders characterized by the pathological inclusions of phosphylated tau protein. In some entities, as Alzheimer's disease, genetic variants have been proposed to contribute to tau pathology. However, genetic causes of other less common tauopathies, as Progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD), are poorly understood. To identify novel causal genes implicated in tauopathies, we provide a detailed study of a large family of 64 individuals with 8 affected patients with an inherited tauopathy that is likely autosomal dominant with reduced penetrance. The index case developed progressive speech and language difficulties at age 64 years. Examination 4 years later disclosed non-fluent aphasia, word-finding difficulties, circumlocution, frontal release signs, and right-sided bradykinesia, rigidity, and pyramidal signs. She died 5 years after the symptomatic onset. Neuropathologic features included numerous ballooned neurofilament-positive neurons, tau-positive astrocytic plaques, and oligodendroglial coiled bodies, all typical of CBD. Two other family members were diagnosed clinically with Parkinsonism and behavioral problems, 2 with Parkinson's disease, 1 with amyotrophic lateral sclerosis, 1 with dementia, and 1 with progressive gait and speech problem. DNA was available from the proband and one first-degree cousin, clinically diagnosed with dementia and Parkinsonism at the age of 54. After exclusion of mutations in MAPT, PGRN and LRRK2 in these affecteds, we performed whole-genome sequencing. Analysis of the genomes of these patients resulted in a list of novel and rare variants that were shared among the affecteds and follow an autosomal dominant inheritance pattern. This list of potential tauopathy variants was confirmed by Sanger sequencing and subsequently screened in a series of >750 population controls which resulted in a total of 6 confirmed-variants which were absent or very rare in controls: SCN10A, OPRK1, CAPRIN2, UBN1, NEURL4 and CCDC9. The presence of additional variants in these candidate tauopathy genes currently is being studied in a series of pathologically-confirmed CBD and PSP cases as well as in 3 additional tauopathy families. This approach provides a select list of potential tauopathy genes by whole-genome sequencing in a family with pathologically confirmed CBD that may help identify novel pathways involved in pathological tau aggregation.

1146T

Distinct genetic variants in Alzheimer's disease, Parkinson's disease and type 2 diabetes. *S.J. Chung[1], S.Y. Kim[2], J. Kim[1], M.J. Kim[3], Y. Kim[1], J.H. Lee[1].* 1) Neurology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; 2) Psychiatry, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; 3) Neurology, Bobath Memorial Hospital, Seongnam, South Korea.

Insulin and insulin-like growth factor contribute to normal brain function. Recent experimental and clinical studies demonstrated the possible link between T2DM and Alzheimer's disease (AD). Experimental evidence also indicates that T2DM and Parkinson's disease (PD), both age-related chronic diseases, share similar dysregulated pathways. Epidemiological researches also demonstrate that T2DM is associated with an increased risk of PD. We aimed to investigate whether genome-wide significant loci of type 2 diabetes mellitus (T2DM) are associated with AD and PD. Study subjects were 400 AD cases, 500 PD cases, and 500 unrelated controls without the evidence of any cognitive impairment or neurological diseases. All subjects were ethnic Koreans. We selected 32 genetic variants from 10 genes (*CDKAL1, CDKN2B, FTO, GLIS3, HHEX, IGF2BP2, KCNJ11, KCNQ1, SLC30A8,* and *TCF7L2*) and intergenic regions based on results of the recent genome-wide association studies (GWAS) in T2DM. These variants were reported to be T2DM-susceptibility loci and were replicated by other independent studies. Two *APOE* SNPs (rs429358 and rs7412) were also genotyped to adjust the *APOE* ε4 status in association analyses using AD cases. All association analyses were performed using logistic regression models, adjusting for age, sex. There was no significant association of each genetic variant of GWAS-linked loci in T2DM with AD or PD. The *KCNQ1* SNP rs163182 showed nominal significance (P = 0.0084) in unadjusted analysis using AD cases and controls, but it did not remain significant after the adjustment for age, sex, and *APOE* ε4 status. Two *APOE* SNPs showed significant associations with AD. Our results suggest that genome-wide significant loci of T2DM have no major role in AD or PD. Further studies are needed to confirm these findings.

1147S

Defects of *ARHGAP36* in patients with developmental delay and autism. *S. Fan[1], L. Dukes-Rimsky[1], R.C. Rogers[1], A. Chaubey[1], C. Skinner[1], L. Wang[2], A.K. Srivastava[1,2].* 1) J.C. Self Research Institute of Human Genetics, Greenwood Genetic Center, Greenwood, SC; 2) Department of Genetics and Biochemistry, Clemson University, Clemson, SC.

Rho family GTPase activating proteins (GAPs) are known to regulate physiological processes including embryogenesis, neurodevelopment, cytokinesis and differentiation and act as negative regulators of Rho GTPases. Several genes implicated in intellectual disability (ID) and autism spectrum disorders code for proteins that function as regulators or effectors of Rho GTPases. ARHGAP36 is one of approximately 70 Rho family GAPs in human with uncharacterized biological functions. Using the Affymetrix Genome-wide SNP 6.0 Microarray system, we identified a copy number loss of approximately 770 kb on chromosome Xq25q26.1 in a 5-year-old male patient with mild fine motor and speech delay, macrocephaly, hypothyroidism, autism, and significant ADHD. The deletion included the *ENOX2*, *ARHGAP36*, and *IGSF1* genes. The proband's normal mother carried the deletion and showed a random X inactivation pattern. We found the identical deletion in the proband's 28-month-old brother with speech delay and macrocephaly. Furthermore, the NCBI ClinVar database lists a copy number loss located in a similar genome region encomapssing *ARHGAP36*, *IGSF1*, and *OR13H1* genes in a patient with hearing impairment and language delay. We screened the *ARHGAP36* gene for mutations in approximately 249 male patients with ID and 245 male patients with developmental delay/autism and identified a novel missense alteration, c.1316G>A (p.R439H), in a 16-year-old male patient with absent speech, hyperactivity and autism. The alteration was absent in 6500 individuals in the NHLBI dataset. Defects of *IGSF1* have been reported in male patients with congenital central hypothyroidism. Thus, the deletion of the *IGSF1* gene is likely responsible for hypothyroidism in the proband and we speculate that his younger brother is also likely to develop hypothyroidism. To understand the physiological function of *ARHGAP36*, we further determined *ARHGAP36* expression in several tissues including human fetal brain. We identified and analyzed genes that are co-expressed with *ARHGAP36*. On the basis of probe-to-probe correlation coefficient calculated from an integrated set of 2,968 microarray expression profiles of healthy human tissue samples, we found that *ARHGAP36* was co-expressed with genes significantly enriched with Gene Ontology terms including synapse, postsynaptic membrane, synaptic transmission and neurogenesis. Altogether, our findings suggest a potential role for the *ARHGAP36* gene in developmental delay and autism.

1148M

HIV-related cognitive impairment shows association with polymorphisms within the dopaminergic system in substance dependent and independent populations. *M.M. Jacobs[1], M. Pizzirusso[1], J. Murray[1], S. Morgello[1,2,3].* 1) Neurology, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Pathology, Icahn School of Medicine at Mount Sinai, New York, NY.

It has been postulated that drugs of abuse act synergistically with HIV, leading to increased neurotoxicity and neurocognitive impairment. The CNS impacts of HIV and drug use converge on the mesocorticolimbic dopamine (DA) system. Using an advanced-stage HIV+ population, previous studies in our laboratory have implicated a role of polymorphisms of two receptors within this system (DRD1 and DRD2) and neuropsychological performance as well as opiate and cocaine dependence. To replicate and expand these studies, we have increased our sample size and added additional polymorphisms within several genes of the dopaminergic system (DRD1-5, COMT, DAT). We observed that polymorphisms of DRD2 are associated with opiate and cocaine dependence in our population. In Caucasian subjects, we observe significant associations for opiate and cocaine dependence with polymorphisms within DRD2 and COMT genes, while a significant association within the DRD1 gene was observed in African-American individuals. Using linear regression analysis, we next examined these polymorphisms for associations with neuropsychological performance in global and cognitive domain T-scores (Motor, Processing Speed, Verbal Fluency, Learning, Memory, Executive Functioning, Working Memory) while controlling for opiate and cocaine dependency. While significant associations were observed in nearly every domain across both populations for multiple polymorphisms, the most significant effects in Caucasian subjects were observed in the motor domain with several DRD2 polymorphisms while African-American subjects had its most significant associations in working memory (COMT and DRD3) and memory (DRD1) domains. For all of these associations, the effects differed for substance dependence groups as the direction of the correlations were opposite to what was seen in subjects without dependency. Future studies will focus on increasing the sample size of the population as well as increasing the number of genes within the dopaminergic circuitry. We conclude that studies to examine genetic risk for HAND must carefully account for substance dependence patterns when assaying dopaminergic systems, as the neurobiological substrates of cognition in HIV populations may vary with tonic alterations secondary to chronic substance exposures.

1149T

HLA-DRA is strongly associated with Parkinson's disease in Iranian population. *J. Jamshidi[1], H. Darvish[2], B. Emamalizadeh[2], A. Movafagh[2].* 1) Department of Biochemistry, Fasa University of Medical Sciences, Fasa, Iran; 2) Department of Medical Genetics, Faculty of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Background and Objective: Parkinson's disease (PD) is a progressive neurodegenerative disorder which impairs the patient's motor skills. The rs3129882, a noncoding variant in HLA-DR, was found to be associated with PD using several genome-wide association studies (GWAS). The aim of this replication study was to explore the relationship between this variant and PD in Iranian population Materials and methods: The study was performed on 520 unrelated patients and 520 healthy controls. The mean age of patients was 59.5±14.5 years and were composed of 278 men and 242 women. The diagnostic criteria for PD were based upon the UK Parkinson's disease Society Brain Bank Clinical Diagnostic Criteria. Genomic DNA was extracted from peripheral blood using a standard salting out method. The rs3129882 variant was genotyped through polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). We examined the association of rs3129882 (A/G) with the risk of PD, under four models (codominant (G/G vs. G/A vs. A/A), dominant (A/A or G/A vs. G/G), recessive (G/G or G/A vs. A/A) and overdominant (G/G or A/A vs. G/A) using SNPassoc package of R version 3.0.1. Odds ratio (OR) together with 95% confidence interval (CI) was Estimated; and a $p < 0.05$ was considered as statistically significant for the tests. Results: Distributions of the rs3129882 (A/G) polymorphism in both PD and control group were in Hardy-Weinberg equilibrium. The alleles frequencies of rs3129882 (A/G) were distributed differently in PD group and control group ($\chi 2 = 4.641$, OR=1.209, 95% CI: 1.017-1.436, $P = 0.035$), and under codominant, dominant and overdominant models the association of the SNP with PD risk was significant, where the A allele was observed to be protective. Discussion: HLA is associated with numerous neurologic disorders and association of HLA with Parkinson's disease was recently revealed. Our result suggests that the rs3129882 (A/G) polymorphism may be a risk factor for PD in Iranian population. It is assumed that the intronic variant plays a role as a cis-acting regulatory element which correlates significantly with overexpression of DR antigens in substantia nigra. The finding is not so far beyond expectation because of the previous evidences for involvement of neuroinflammation and adaptive immunity in PD pathogenesis. This association emphasizes an important biological pathway in the etiology of the disease and points to a potential target for new therapies.

1150S

Ancestral haplotypes of BHLHE40 in non-24-hour rhythms and bipolar disorder. *D.F. Kripke[1], S.A. Ament[2], C.M. Nievergelt[1], J.R. Kelsoe[1].* 1) Psychiatry, UCSD, La Jolla, CA; 2) Institute for Systems Biology, Seattle, WA.

A previous study (Kripke et al., Psychiatry Investigation, in press, 2014) demonstrated a haplotype group in the circadian gene BHLHE40 that was associated with patients displaying non-24-hour free-running circadian rhythms. This haplotype group and several of its SNVs were also nominally associated with bipolar disorder. The only exonic variant of the 16 in the haplotype group, rs908078, is a synonymous variant which may alter splicing. Also, several of the variants may alter transcription binding sites. With further study of the haplotypes, we now found that 7 of the 16 key minor variants were identical with the chimpanzee reference, and 6 matched the reference allele of the majority of 9 non-human primates. Each of these 7 variant alleles was also reported at least once in available early sequencing data from 3 Neanderthals and the Denisovan, so they likely have an ancestral primate origin. According to 1000 Genome data (Phase 1), 5 of the 7 minor alleles are distinctly more common among Africans than Europeans or Asians. In sum, current data suggest that these 7 SNVs are of ancestral origin but have been negatively selected in modern humans, particularly outside the tropics. Perhaps altered circadian dynamics were favored after the discovery of fire and especially after migration to more northern latitudes where more time must be spent inside shelters. This would not be surprising, since Drosophila and several small mammals have recognized genetic variability among populations related to adaptations to latitude and photoperiod. Whole-genome sequencing of BHLHE40 was now examined 325 members of multiplex bipolar disorder pedigrees, among whom 74 individuals displayed these SNVs, but the sequenced pedigrees are not yet sufficient to examine association powerfully. Recent PGC GWAS analyses have given no support to any BHLHE40 variant association with bipolar disorder, but Seifuddin et al. (2013) found a nominally-significant association with down-regulated brain BHLHE40 by summarizing studies of bipolar disorder. Since the ancestral haplotype group has not been well represented in GWAS studies, further replication of expression studies and expanded studies with whole-gene-region sequencing will be needed to clarify the import of BHLHE40 for circadian and affective disorders.

1151M

***MAPT* non-coding variation in neurodegenerative disorders.** *C. Labbé[1], M. Heckman[2], K. Ogaki[1], O. Lorenzo-Betancor[1], A. Ortolaza[1], R. Walton[1], D. Serie[2], R. Uitti[3], Z. Wszolek[3], O. Ross[1].* 1) Neuroscience, Mayo Clinic Florida, Jacksonville, FL; 2) Section of Biostatistics, Mayo Clinic, Jacksonville, FL; 3) Neurology, Mayo Clinic, Jacksonville, FL.

Genome-wide association studies have identified over 20 loci associated with increased risk to sporadic Parkinson's disease (PD). The *MAPT* gene is in one of the most systematically replicated locus in PD. The gene encodes protein tau which aggregate in brain inclusions and define the neurodegenerative diseases called tauopathies. A common non-recombining *MAPT* haplotype (*MAPT* H1) has been associated to several tauopathies including progressive supranuclear palsy (PSP); yet *MAPT* H1 has also been implicated in the risk to Parkinson's disease (which is not a traditional tauopathy). Preliminary sub-haplotype analyses suggest that different genetic variants on the *MAPT* H1 haplotype associate with each of these disorders. To date it remains unclear which variant(s) at the *MAPT* locus is(are) responsible for the risk and what is the underlying pathomechanism of disease. We set out to identify causal variants for PD and PSP within the *MAPT* region using next generation sequencing technologies. We captured the entire *MAPT* gene and 10kb on each side, a 154kb genomic region that was sequenced in 300 patients with PD, 300 patients with PSP, and 300 controls using a pooling strategy (10 DNA samples/pool). We designed 4248 amplicons using the Haloplex platform followed by Illumina sequencing for a total coverage of 96.3%. We identified over 4000 variants in the coding, intronic and surrounding region of the *MAPT* gene. Among these variants 39% are singletons, 19% have a minor allele frequency between 1% and 5%, and 42%, over 5%. We identified 43 variants located within 25 base pairs of an intron/exon boundary, some of which map to boundaries of the well-known excised exons 2, 3, 4a, 6, 8 and 10. We selected SNPs for stage 2 based on association tests and are using the OpenArray platform to genotype the variants in our replication series consisting of a total of 4500 independent samples. We will present results using the latest sequencing and genotyping technologies to comprehensively define the *MAPT* locus associated with PD and PSP and thus identify novel targets for both neuroprotective and symptomatic therapies.

1152T

9.6% of mouse gene knockouts show abnormal neuroanatomy: a resource to identify genes related to intellectual disability in human. *A. Mikhaleva[1], A. Baud[2], V. Vancollie[3], A. Edwards[4], M. Kannan[5], H. Whitley[5], C. Wagner[5], A. Duret[5], I. Herr[1], K. Navarro[1], A. Delay[1], S. Jayaram[1], J. Estabel[3], D. Wright[6], D. Fairley[6], C. Lelliott[3], J. White[3], D. Adams[3], D. Keays[7], J. Flint[4], Y. Herault[5], A. Reymond[1], B. Yalcin[1,5].* 1) Center for Integrative Genomics, University of Lausanne, Switzerland; 2) EMBL-European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, UK; 3) Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1HH, UK; 4) Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK; 5) Institute of Genetics and Molecular and Cellular Biology, Illkirch, 67404, France; 6) HistologiX Ltd, Biocity, Nottingham, NG1 1GF, UK; 7) Research Institute of Molecular Pathology, 1030 Vienna, Austria.

Although intellectual disability affects 1-3% of the population, it is one of the least understood health problems. It is estimated that genetic lesions account for half of the currently undiagnosed cases. Despite recent successes in identifying some of the mutations responsible, it has been suggested that up to 1,000 more genes remain to be uncovered. The large number of intellectual disability syndromes is due to many causal pathophysiological mechanisms. The diversity of mechanisms results in an array of quantifiable neuroanatomical abnormalities. To identify genes related to intellectual disability, we are collaborating with the Sanger Mouse Genetics Project (MGP), allied to the International Mouse Phenotyping Consortium (IMPC), to systematically study the neuroanatomy of the MGP/IMPC knockout mouse strains using a standardized set of 78 brain parameters. So far, we have assessed brain defects in 425 knockout mouse mutants (this number will double by the end of 2014). These preliminary data yielded success with the identification of 20 known intellectual disability genes including Ap4e1, Cenpj, Chd7, Mcph1, Sc4mol and Ube3b demonstrating the pertinence of our approach. We also discovered 21 other genes including Mta1, Ccdc104, Caprin2 and Dusp3, which when disrupted caused modification of brain structures. Our study is the largest screen of brain morphology from the MGP/IMPC. It shows that we can detect abnormalities in about 10% of knockout mouse mutants, and that these translate into human pathology. This offers a complementary resource to human genetic studies.

1153S

Homozygous deletions of non-coding transcriptional control sites in autism spectrum disorder. *K. Schmitz-Abe[1,2,3,4,7], M. Chahrour[1,2,3,4,5], S. Hill[1,2,3], G. Sanchez-Schmitz[1,3], J. Partlow[1,2,3], B. Mehta[1,2,3], S. Servattalab[1,2,3], A. Lam[1,2,3], E. Morrow[8], T. Yu[1,2,3,4,5,6], C. Walsh[1,2,3,4,5], K. Markianos[1,2,3].* 1) Division of Genetics and Genomics, Department of Medicine, Boston Children's Hospital, Boston, Massachusetts, USA, 02115; 2) Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, Massachusetts, USA, 02115; 3) Harvard Medical School, Boston, Massachusetts, USA, 02115; 4) The Autism Consortium, Boston, Massachusetts, USA, 02115; 5) Howard Hughes Medical Institute, Boston Children's Hospital, Boston, Massachusetts, USA, 02115; 6) Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA, 02114; 7) Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA;; 8) Department of Genetics and Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.

Autism spectrum disorder (ASD) is characterized by extensive genetic heterogeneity. A subset of ASD has been linked to disruptions of gene coding regions by de novo copy number variants (CNVs), de novo single nucleotide variants (SNVs), or inherited biallelic SNVs but most cases remain unexplained. We analyzed CNVs in 183 consanguineous families with ASD and found that individuals affected with ASD have significantly more homozygous deletions compared to unaffected siblings [22% vs. 15%]. This excess suggests that homozygous deletions might account for up to 7% of ASDs in our cohort. In contrast to most de novo CNVs and SNVs, we found that the homozygous deletions often did not include exons, but were enriched for DNA regulatory regions, with 12 of 16 disrupting ENCODE regulatory elements, a rate much higher than expected by chance (bootstrap $p<0.05$). Specifically, we cross-referenced the 16 deletions with histone modification marks from the ENCODE project. We selected H3K4Me1, H3K4Me3, and H3K27Ac as three of the best characterized histone modifications corresponding to states of gene transcription and enhancer activity, and then examined ChIP-Seq data gathered from 7 cell lines (GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK, NHLF). Our data suggest an important new mechanism of ASD, and highlight the importance of patterned gene activation in cognitive and social function.

1154M

VGF as a potential target for Night Eating Syndrome. *G.J. Wyckoff, S.B. Mullegama.* MBB, SBS, University of Missouri- Kansas City, MO.

In recent years, clinicians and researchers have realized that a significant number of individuals with eating disorders did not fit into DSM categories of anorexia nervosa and bulimia nervosa. By default, many received a diagnosis of “eating disorder not otherwise specified.” Nonetheless Eating Disorders(ED) have biochemical indicators such serotonin, norepinephrine, dopamine. Deeper analyses of EDs in recent years have shown that there are genes that may play a role in ED. Night Eating Syndrome is specifically characterized as a “delayed circadian pattern of food intake”, and it may be part of a broader spectrum of sleep-related eating disorders. Identification of genes that may be related to Night Eating syndrome should be undertaken to help determine treatment and diagnosis options. To gather preliminary data for this question, we collected data from the Gene Expression Omnibus (GEO). The GEO website has a feature named GEO2R that does some analysis of datasets. The GEO2R does a normalization of data with a log2 transformation along with a Benjamini & Hochberg (False discovery rate) plus adding annotation. GEO2R provides the R script of the online process that GEO2R performed for the particular dataset. In order for the GEO2R to perform analysis, experimental groups must be assigned. The datasets that were selected for the most part had the design of control versus experimental group. Thirteen relevant datasets from GEO were passed through the GEO2R online application. Following analysis, genes of interests were selected by using the criterion of $|z| \geq 5.5$, which is well above $|z| \approx 5$ that corresponds to a Bonferroni-corrected P value of 0.001. The analysis of the datasets resulted in 44 genes that are possible candidates for further studies. An extensive literature search and further review suggests that the most interesting of the 44 genes collected is the gene VGF, which is associated with biological processes such as glucose homeostasis, insulin secretion, response to dietary excess, and has been shown to be under the control of the circadian clock in mammals.

1155T

Association of HTR2C gene variants with suicidal behavior: A cases-control study and meta-analysis. *C.A. TOVILLA-ZÁRATE[1], M.I. López-Narvaez[5], I.E. Juárez-Rojop[2], J. Bles-Castillo[2], S. Pool-García[3], M. Velazquez-Sanchez[3], M. Villar-Soto[4], M.A. Jimenes-Santos[1], D. Bermudez-Ocaña[1], T.B. González-Castro[2].* 1) Universidad Juárez Autónoma de Tabasco, División Académica Multidisciplinaria de Comalcalco, Comalcalco, Tabasco, Mexico; 2) Universidad Juárez Autónoma de Tabasco, División Académica de Ciencias de la salud. Villahermosa, Tabasco, México; 3) Hospital General de Comalcalco, Secretaría de Salud, Comalcalco, Tabasco, México; 4) Hospital de Alta Especialidad "Rovirosa", Villahermosa, Tabasco, México; 5) Hospital General de Yajalón, Secretaría de Salud, Yajalon, Chiapas, México.

Background: Our aim was explore the role of HTR2C gene in pathogenesis of suicidal behavior through a study association in Mexican population and to get a more comprehensive data we conducted a meta-analysis and a systematic-review with up-do-date outcomes. Methods: A number of 192 suicide attempters and 207 healthy volunteers were included. Regarding to meta-analysis and systematic review was collected 1,604 cases and 2,143 controls for them study. Results:In Mexican population we found a significant association in the allelic distribution of the polymorphism ($p=0.03$ and $X2= 4.34$) and also in the common haplotype AACGC ($p = 0.01$ and $X2= 5.51$). In the meta-analysis not associatión was observed between HTR2C variant gene and allelic, additive, dominant and recessive models. Conclusions: Even we did not find an association in the meta-analysis, in Mexican population it could be suggest a possible role of the HTR2C gene that collaborate in the pathology of suicidal behavior. The lack association of the meta-analysis it could be because may be the expression of this gene could depend of the ethnic.

1156S

Replicative association analysis of schizophrenia in Russian population of Siberian region. *A. Bocharova, A. Marusin, V. Stepanov.* Research Institute of Medical Genetics, Tomsk, Russian Federation.

Introduction. Schizophrenia is a common neuropsychiatric disorder. The etiology of schizophrenia as presently defined is not well understood. It is a complex disease with genetic, environmental and lifestyle factors. Common psychiatric and neurological disorders including schizophrenia are the subject of intensive genetic research based on genome-wide association studies (GWAS) and targeted resequencing of genomic regions of interest. Genetic variants associated with cognitive impairments, which are an important endophenotypes for schizophrenia, also have been revealed by GWAS. The aim of this study was to analyze associations of 15 SNPs reported in GWAS with schizophrenia in Russian population of Siberian region. **Subjects and Methods.** 15 SNPs strongly associated with schizophrenia and cognitive performance according to recent GWA studies were genotyped by real-time PCR in Russian patients with schizophrenia (N = 114) and in healthy control group (N = 285). DNA samples were genotyped using TaqMan® SNP Genotyping Assays (Applied Biosystems) under condition recommended by the manufacturer. Allele-specific ORs and associated p values were calculated. **Results.** The association of rs2616984 located in the region of CUB and Sushi multiple domains 1 (*CSMD1*) gene, previously reported in GWAS by Cirulli et al. (2010), was confirmed in Russians (OR = 1.4, p = 0.04). This gene is of unknown function but has been implicated in multiple neurodevelopmental disorders. Genetic markers in loci for *VRK2, SLCO6A1, NOTCH4, TCF4, ZNF804A, AGBL1, TLR4, RELN, ZFP64P1, KCNB2, CPVL, NRGN* and *NRIP1* previously reported in GWAS were not associated with the disease in Russian population of Siberian region. **Conclusions.** Our findings replicate association of schizophrenia with rs2616984 at *CSMD1* gene and demonstrate that genetic variability in cognitive performance and schizophrenia have overlapping genetic background. This work is partially supported by RFBR grant #12-04-00595.

1157M

GABAergic interneuron origin of Schizophrenia - a genetic association analysis in South Indian population. *KR. Saradalekshmi[1], Balan. Shabeesh[1], Sathyan. Sanish[1], NV. Neetha[1], Nair. Indu[2], Nair. Chandrasekharan[3], Moinak. Banerjee[1]*. 1) Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, India; 2) Mental Heath Centre, Thiruvananthapuram, India; 3) Nairs Hospital, Cochin, India.

Schizophrenia is one of the most debilitating disorders affecting about 1% of the world population. The etiology is complex involving a major genetic contribution as well as environmental factors interacting with the genetic susceptibility. Several lines of evidence from post-mortem and brain imaging in schizophrenia patients suggest that Gamma-amino butyric acid (GABA) deficits may contribute to the pathophysiology of schizophrenia. Changes in GABA receptors in schizophrenic brain may represent a primary pathogenesis, such that modifications to GABA receptor subunit assembly and/or structure contribute to the symptoms of schizophrenia. The GABA receptors are a class of receptors that respond to the neurotransmitter gamma-aminobutyric acid (GABA), the chief inhibitory neurotransmitter in the vertebrate central nervous system. Brain-derived neurotrophic factor (BDNF) is one of the most important modulators of glutamatergic and GABAergic synapses. BDNF can play a permissive role in shaping synaptic networks, making them more susceptible for the occurrence of plastic changes. Reelin(RELN), a neuromodulator secreted by GABAergic interneurons regulates the synaptic plasticity. Together, these genes play an important role in regulation of neurotransmission and synaptic plasticity .The objective of the study was to investigate the association of polymorphisms in GABA (A) receptors, BDNF and RELN genes with schizophrenia in a south Indian population. DNA was isolated from 300 patients and 300 healthy controls after obtaining informed consent. Only patients suffering from schizophrenia diagnosed by DSM-IV/ICD 10 have been enrolled for this study. Symptom severity was rated using BPRS-E and PANNS. Age, sex and ethnicity matched controls were recruited for the study. We investigated the association of 10 single-nucleotide polymorphisms: rs2279020 (GABRA1), rs3219151 (GABRA6), rs2229944 (GABRB2), and rs211037 (GABRG2), BDNF -270C/T, rs6265(BDNF), rs727531(RELN), rs7341475(RELN), rs6951875(RELN), rs362719 (RELN)with predisposition to schizophrenia. Samples were genotyped using allele specific amplification followed by fluorescence detection (KASPar), PCR-RFLP and PCR sequencing. Allele and genotypic frequencies were calculated for patients and controls and association with predisposition to schizophrenia was tested.

1158T

Parsing genetic associations in the MHC in schizophrenia. *S. Mukherjee[1,2], S. Ripke[3,11], O. Andreassen[8,15], A. Corvin[6], P. deBakker[16], J. Knight[12], S. McCarroll[3, 13], B. Neale[3,4,5,11], V. Nimgaonkar[17], R. Ophoff[14], J. Pouget[12], P. Sullivan[7, 9,10], Y. Wang[8], T. Lencz[1, 2], MHC- Schizophrenia Working Group of the Psychiatric Genomics Consortium.* 1) Psychiatry Research, The Zucker Hillside Hospital, Glen Oaks, New York, NY; 2) The Feinstein Institute for Medical Research, Manhasset, NY, USA; 3) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 6) Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity College Dublin, Ireland; 7) Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; 8) NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 9) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 10) Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA; 11) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 12) Department of Psychiatry, University of Toronto; 13) Department of Genetics, Harvard Medical School, Boston, MA, USA; 14) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA; 15) Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; 16) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America; 17) UPMC/University of Pittsburgh Schools of the Health Sciences,Pittsburgh, PA.

The major histocompatibility complex (MHC) has emerged as a region of major interest in schizophrenia genetics. Large-scale genome-wide association studies (GWAS) in schizophrenia have converged to demonstrate that the MHC contains the strongest association signal for illness susceptibility. However, prior GWAS have been unable to precisely localize the source of this signal, due to the extensive long-range linkage disequilibrium throughout the MHC; different studies have identified top SNPs ranging across a nearly 10Mb extent (coordinates ranging from 25-35Mb on Chromosome 6). A subcommittee within the PGC SCZ working group has been formed to parse the signal within the MHC using HLA imputation and conditional analysis. Methods: Of the 52 PGC-SCZ cohorts, raw genotype data were available for 38 cohorts of European ancestry, with a total n = 64,631 (29,148 cases and 35,483 controls). Imputation of classical HLA alleles was performed using SNP2HLA (Jia et al. 2013) applied to raw genotype data from each cohort. A total of 267 HLA alleles were successfully imputed. Regression and conditional regression analyses, covarying for top 10 PCAs and study site, were performed in PLINK. Results: Six HLA alleles that form the so-called 8.1 ancestral haplotype (AH8.1) attained genomewide significance (10^{-15} < all p-values < 10^{-9}) and were protective (frequencies in cases ~10.5-11% vs ~12-12.5% for controls; OR~0.87). Notably, these associations were much less strong than those observed for individual SNPs across the extended MHC, and conditional analyses covarying for AH8.1 components revealed genomewide significant SNPs remaining throughout the region, with top signals at rs34661691 (an eQTL for BTN3A2) and rs2523721 (an eQTL for HLA-A and VARS2). Additional conditional analyses, covarying for top individual SNPs rather than HLA alleles, are ongoing and results will be presented at the meeting. Discussion: Initial results demonstrate a pattern that is markedly different from that observed for autoimmune disorders. In most autoimmune disorders, AH8.1 is associated with risk; whereas in schizophrenia, the opposite relationship is observed. . Moreover, in autoimmune disorders, SNP effects in the MHC are entirely accounted for by HLA alleles and corresponding amino acid changes. By contrast, SNP effects in schizophrenia are likely to play a significant independent role, probably regulatory in nature, and potentially implicating non-HLA genes.

1159S

Common and rare genetic risk factors converge in protein interaction networks underlying schizophrenia. *X. Chang[1], L. Lima[1], Z. Wei[2], J. Li[1], P. Sleiman[1], H. Hakonarson[1].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, USA.

Schizophrenia is an idiopathic brain disorder with profound genetic heterogeneity. The major forms of genetic risk factors in schizophrenia are common single nucleotide polymorphisms (SNPs) and rare copy number variants (CNVs). Recent studies have shed new light on the genetic basis of schizophrenia that de novo mutations play a prominent part in sporadic form of schizophrenia. However, the underlying genetic mechanisms remain unclear. We first performed a network-based pathway analysis on our previous meta-analysis of GWAS in schizophrenia and bipolar disease, which contains 13,394 cases and 34,676 controls from 16 cohorts. The largest connected component (LCC) identified in our PPI network is significantly enriched in many pathways relevant to synaptic plasticity, neural development and signaling transduction such as long-term potentiation, RAP1 signaling pathway, GnRH signaling pathway. We also observed the proteasome pathway and ubiquitin mediated proteolysis pathway to be significantly enriched. Dysfunction of the ubiquitin-proteasome pathway has been implicated in the pathology of various neurodegenerative diseases. However, growing evidence revealed that schizophrenia patients have aberrant gene expression patterns in the UPP suggesting the UPP may also contribute to the deficits in schizophrenia. In order to add more genetic pieces to the schizophrenia puzzle, we collected literature reported genes disrupted with CNVs and de novo mutations in schizophrenia patients for the network analysis. We found the size of LCC were significantly increased, bigger than 10000 simulation based on genes randomly selected. To pinpoint a small group of interacted genes with a significant combined effect to schizophrenia, we developed an edge-based network search algorithm for detecting casual gene modules in PPI networks. The most significant module is comprised by DISC1 connectors and N-methyl-D-aspartate (NMDA) receptor associated genes. We then combined the top three modules and constructed a small sub-network of 47 nodes and 76 edges, which is highly enriched in long-term potentiation and Calcium signaling pathway. In summary, we demonstrated that multiple types of genetic risk variants converge on a PPI network enriched with schizophrenia susceptibility genes involved in synaptic plasticity and neural development. We also provided evidence that de novo mutations and common variants are likely to influence same genes or functionally interacted genes.

1160M

Alzheimer's Disease: Analyzing the Missing Heritability. *K.L. Hoyt, P.G Ridge, K. Boehme, L.A. Staley, H. Smith, J.S.K. Kauwe, Alzheimer's Disease Genetics Consortium.* Biology, Brigham Young University, Provo, UT., USA.

Alzheimer's disease (AD) is a complex disorder influenced by genetic and environmental factors. We previously analyzed the phenotypic and genetic variance explained by SNPs known to be associated with AD, APOE alone, and by ~2 million genotyped or imputed SNPs from the Alzheimer's Disease Genetics Consortium (ADGC), using the Genome-wide Complex Trait Analysis (GCTA) software. We found that 33% of total phenotypic variance is explained by common SNPs (n=~2 million), 6% by APOE alone (including the e2 and e4 alleles), and 2% by other known markers—meaning 25% of phenotypic variance remains unexplained by known AD markers. In this study we follow-up our previous work using an expanded version of the ADGC dataset. In this expanded version, SNPs were imputed using the 1000 Genomes dataset, resulting in more complete imputation of both common and rare variation. In the present analysis we estimate phenotypic and genetic variance explained by the SNPs. Additionally, while our previous study included only APOE and nine other loci, the present study includes all the IGAP and previously known AD SNPs (n=23, includes APOE e4 and e2). In addition we will evaluate the contribution of SNPs in genes known to harbor rare variants for AD (APP, PSEN1, PSEN2, TREM2, PLD3, ADAM10). We discuss the current state of known AD heritability and suggest where resources should be focused to identify the remaining undiscovered AD SNPs.

1161T

Probing the shared polygenic underpinnings of anorexia nervosa and five other major psychiatric disorders. *L.M. Huckins[1], K.S. Mitchell[2], L. Thornton[3], D. Collier[4], P.F. Sullivan[3], C.M. Bulik[3], E. Zeggini[1], WTCCC3 Consortium; GCAN Consortium.* 1) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Boston University, MA, USA; 3) University of North Carolina at Chapel Hill, NC, USA; 4) King's College London, London, UK.

Anorexia nervosa (AN) is marked by extremely low body weight and intense fear of gaining weight. Comorbidity with other psychiatric disorders (PsyD) is common (up to 56.2%) We evaluated shared genetic determinants of AN and comorbid PsyD by testing whether polygenic risk scores derived from genome-wide data of other PsyD can predict AN status. We obtained allele risk scores for major depressive disorder (MDD), bipolar disorder (BPD), autism (AUT), attention deficit hyperactivity disorder (ADHD) and schizophrenia (SCZ), and a set of cross-disorder risk alleles, from the Psychiatric Genomics Consortium (PGC). We divided each of these sets into 10 Pt significance level thresholds. The test set comprised AN cases and controls from a published WTCCC3 AN GWAS study. For every sample in this test set, we produced a polygenic risk score as a weighted sum of risk allele scores. Logistic regression was used to assess whether each of the five polygenic scores predicted AN case-control status. We computed pseudo R^2 values, and compared these to the values obtained when randomly permuting case-control status to measure the proportion of variance in AN explained by each PsyD risk score. Our pseudo R^2 values were ~0.5-1%, comparable to pseudo R^2 values found by the PGC in a recent cross-disorder polygenic score analysis of five PsyD [1]. We used the baseline results to compute an empirical p-value for every significance threshold and found significant pseudo R^2 values at the lowest Pt threshold (Pt< 0.001) for AN vs AUT (p=0.0009), MDD (p=0.009), and SCZ (0.0008), and at the second lowest Pt threshold (Pt <0.01) for BPD (p=0.0004). We demonstrated for the first time a shared genetic etiology between AN and other PsyD using genome-wide data. Cross-disorder analyses may prove to be a powerful approach to identifying overlapping susceptibility loci across PsyD. [1] Cross-Disorder Group of the Psychiatric Genomics Consortium (2013) "Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis" The Lancet 381:1371-1379.

1162S

Genes regulated by epigenetic mechanisms in determining general intelligence (g) are over-represented in disorders that affect cognition. *P. Cha[1], K. Kobayashi[1], Y. Ando[1], K. Takao[2], T. Miyakawa[2,3], T. Toda[1,4].* 1) Division of Molecular Brain Science, Kobe University Graduate School of Medicine, Kobe, Hyogo, Japan; 2) Center for Genetic Analysis of Behavior, Section of Behavior Patterns, National Institute for Physiological Sciences, 38 Aza-Nishigonaka, Myodaiji-cho, Okazaki, Aichi 444-8585, Japan; 3) Division of Systems Medical Science, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan; 4) Division of Neurology, Kobe University Graduate School of Medicine, 7-5-1 Kusunoki-chou, Chuo-ku, Kobe 650-0017, Japan.

General intelligence (g) is a common core shared by cognitive tasks. Despite the high heritability of "g" that increases with age, molecular mechanisms that underlie "g" remains poorly understood. To study epigenetic regulation of "g", we conducted 5 mouse behavioral studies on 41 inbred mice and identified mice with high, medium, and low cognition. Through subsequent genome-wide gene expression profiling by using hippocampal RNAs of these cognitively discordant mice followed by gene-set enrichment analysis (GSEA), we identified 17 genes and several pathways that expressed differentially between mice with high and low cognition. Since inbred mice are genetically homogeneous, we expect most of the differences that we observed are resulting from epigenetic mechanisms. In addition to memory and learning, the role of epigenetic mechanisms in cognition and neurological disorders has recently come to lights. To investigate if genes implicated in our study provide insights into mechanisms underlying "g" and disorders that affect cognition, we evaluated enrichment of genes reported to be associated with each of the cognitive function/intelligent quotient (IQ), intellectual disability (ID), autism, Parkinson's disease (PD), Alzheimer's disease (AD), Rett Syndrome, and Schizophrenia in the top 100 genes, as well as their interacting proteins, identified in our study. Information on genes reported to be associated with IQ, ID and each of the aforementioned disorders was retrieved from published studies or disease-related databases (AutDB, AlzGene etc), whereas interacting proteins were identified from Protein Interaction Network Analysis (PINA) platform. Overlap enrichment analysis indicated that the 100 genes or genes encoding their interacting proteins are significantly enriched with candidate genes reported to be associated with IQ, ID, PD and AD (p<0.05 in 100,000 permutation test). Subsequent investigation of the commonly shared genes may shed light on common mechanisms and pathways that underlie "g" and the aforementioned disorders.

1163M

Novel Locus in 15q23 Implicated in Recovery after Severe Traumatic Brain Injury. *Y.P. Conley[1,2], D. Ren[1,4], S. Beers[5], A. Puccio[3], D. Okonkwo[3], D.E. Weeks[2,4].* 1) School of Nursing, Univ Pittsburgh, Pittsburgh, PA; 2) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh PA; 3) Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA; 4) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA; 5) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA.

Traumatic Brain Injury (TBI) often results in chronic disability, however there is great variability in symptom development and severity of disability with little knowledge to date to explain the biological underpinnings for this variability. Genomics, though rarely used in the context of recovery following injury, provides a toolbox for assessment of the underlying biology that may explain the variability in phenotypes post-TBI. This project utilized a cohort of subjects recruited immediately after a severe TBI and followed for up to 24 months post injury using the neurobehavioral rating scale (NRS) to conduct a genomic evaluation using the Illumina Human Exome Beadchip. After data quality checks and filtration using the GWASTools R package, 386 subjects were analyzed and six SNPs within a 414kb region in 15q23 were implicated in NRS scores at six months post-TBI (p=3.8×10-7 through p=4.6×10-5) after controlling for age, sex, and extent of initial injury. This region has been implicated previously with rate of progression in subjects with mild cognitive impairment and contains several novel, biologically relevant candidate genes. This is the first study to conduct a non-candidate driven association study to identify genes potentially implicated in recovery after TBI.

1164T

Long non-coding RNAs associated with synapse are differentially expressed in autistic bloods. *W. Ju[1], Y. Wang[2], X.L. Zhao[3], M. Flory[1], E.C. Jenkins[1], W.T. Brown[1], N. Zhong[1,2,3].* 1) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY; 2) Shanghai Jiaotong University Children's Hospital, Shanghai, China; 3) Peking University Center of Medical Genetics, Beijing, China.

Autism spectrum disorders (ASD) are characterized by significant impairments in reciprocal social communication and the presence of repetitive and/or restricted behaviors. The most recent surveillance data showed that ASD affects as many as 1 in 68 children within the United States. The increased incidence of ASD may reflect environmental exposures interaction(s) with genetic factor(s). Several studies have demonstrated that noncoding RNAs may involve in the development of ASD and the long noncoding RNA (lncRNA) may be differentially expressed in ASD brains and contributes to ASD risk. We have studied 25 pairs of blood samples collected from ASD vs. controls to investigate the lncRNA in the ASD bloods. We identified both up-regulated and down-regulated lncRNAs from ASD bloods. Among the up-regulated 2,329 lncRNAs, 1,053 were intergenic, 210 exon sense-overlapping, 69 intron sense-overlapping, 592 natural antisense, 392 intronic antisense, and 13 bidirectional. Among the 1,408 down-regulated lncRNAs, 644 were intergenic, 129 exon sense-lapping, 96 intron sense-overlapping, 443 intronic antisense, and 96 bidirectional with no natural antisense. Six pathways, constructed with three up-regulated lncRNAs involved in long-term depression, long-term potentiation, and synaptic vesicle cycling and three in down-regulated lncRNAs involved in Huntington's disease, Alzheimer diseases, and Parkinson's disease, showed neurological lncRNAs may be differentially expressed in peripheral bloods. Among the lncRNAs identified, 27 (11 up-regulated and 16 down-regulated) were found as the synapse associated. Functionally, they are localized at the loci of gene SYT (synaptotagmin), SYN (synapsin), STX (syntaxin), SV2C (synaptic vesicle glycoprotein 2C), SDCBP2 (syntenin), SYNM (synemin), SYNRG (synergin gamma), and SYNPO2L (synaptopodin 2-like protein). Accordingly, a subset of synaptic mRNAs, transcribed from genes SYT, SYN, STX, SYNGR, SDCBP2, SYNPO2L, SNAP25, SYDE1, STXBP6, SYNJ2BP, SYNCRIP, SNCA, SDC2, SYCP3, SYPL2, SYBU, and SYCE1, were identified. Identification of synapse-associated lncRNAs that were differentially expressed in the ASD peripheral bloods may have opened a new avenue to investigate the epigenetic mechanisms underlying ASD and to explore the potential of applying the differentially expressed lncRNA as a biomarker for early detection of ASD clinically.

1165S

Evidence pointing to abnormal energy metabolism in two genetic animal models of epilepsy. *A.H.B. Matos[1], A.S. Vieira[1], V.D.B. Pascoal[1,3], C.S. Rocha[1], M.F.D. Moraes[2], C.V. Maurer-Morelli[1], D.R. Nascimento[2], S. Martins[2], A.S. Martins[2], A.C. Valle[4], A.L.B. Godard[2], I. Lopes-Cendes[1].* 1) UNICAMP, Campinas, Brazil; 2) UFMG, Belo Horizonte, Brazil; 3) UFF, Nova Friburgo, Brazil; 4) USP, Sao Paulo, Brazil.

Background: Wistar audiogenic rat (WAR) is a genetic epilepsy model susceptible to audiogenic seizures, after high-intensity sound stimulation. Another genetic model we have recently identified is the generalized epilepsy with absence seizures (GEAS) rat. The aim of this study was to determine the molecular pathways involved in the susceptibility to seizures of these two strains using gene expression analysis. Methods: We obtained total RNA from five susceptible WAR [hippocampus and corpora quadrigemina (CQ)], five control Wistar and five WAR naive (without stimulation) as well as from hippocampus and somatosensory cortex (SC) of five GEAS rats and five control Wistar. Gene expression analysis was performed using microarray technology, and analyzed in R environment using the Affy and RankProd packages from Bioconductor, as well as the MetaCore® platform to identify molecular networks, gene ontology categories and gene interactions. Genes with differential expression and a possible biological role in epileptogenesis were validated by qRT-PCR. Results: In WAR, expression profileshowed a total of 1624 differentially expressed transcripts in the CQ and 1351 differentially expressed in the hippocampus compared with controls, with 616 upregulated and 1008 downregulated in CQ and 660 upregulated and 691 downregulated in the hippocampus. Enriched gene ontology categories identified in WAR were involved in oxidative phosphorylation, neurophysiological process GABA-A receptor life cycle. Genes validated by qRT-PCR were Grin1, Nedd8, Il18 and Slc1a3. The comparison of these genes between WAR naive and WAR showed that overall these genes are upregulated in WAR. In GEAS rats expression profile showed a total of 2307 differentially expressed transcripts in the hippocampus and 2282 differentially expressed in the SC compared with controls, with 1039 upregulated and 1268 downregulated in hippocampus and 991 upregulated and 1291 downregulated in the SC. The top enriched gene ontology categories included: oxidative phosphorylation, LRRK2 in neurons in Parkinsons disease. Genes validated by qRT-PCR were Grin1, Gabbr1 and Slc6a1. Conclusion: Interestingly, among the gene ontology categories abnormally expressed in both models (WAR and GEAS rats) we found oxidative phosphorylation, indicating that an abnormal energy metabolism in the central system may be an important mechanism leading to epileptogenesis in these two genetic models of epilepsy. Supported by FAPESP, BRAZIL.

1166M

Genetic influences on metabolite levels in Alzheimer's Disease. *P. Proitsi[1], M. Kim[2], L. Whiley[2], P. Sham[3], S. Lovestone[4], J. Powell*[1], C. Legido-Quigley*[2], R. Dobson*[5], AddNeuroMed Consortium.* 1) King's College London, Institute of Psychiatry, De Crespigny Park, London, UK; 2) King's College London, Institute of Pharmaceutical Science, London, UK; 3) Department of Psychiatry, The University of Hong Kong, Hong Kong; 4) Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK; 5) NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at South London and Maudsley NHS Foundation Trust.

A better understanding of the biological mechanisms underlying Alzheimer's Disease (AD) is required. AD is one of the major challenges for healthcare in the 21st century and with estimated longer life expectancy, the worldwide numbers of demented patients are expected to reach 81.1 million in 2040. Studies have now demonstrated the promise of using associations with blood metabolites, the repertoire of small molecules present in cells and tissue, as functional intermediate phenotypes in biomedical and pharmaceutical research. In AD, a number of non-targeted blood metabolomic studies have been performed highlighting the role of lipid compounds in disease initiation and progression. In addition, recent studies investigating the genetics of the human metabolome have identified genetic variants in metabolism-related genes that lead to clearly differentiated metabolic phenotypes, termed 'genetically influenced metabotypes' (GIMs), which provide new insights into the role of inherited blood metabolic variation. The aim of this study was to investigate genetic influences on human plasma metabolites in a sample of >400 AD patients and controls, in a two stage approach (Stage 1 N=102, Stage 2 N=315), in order to survey regions of the genome associated with metabolic traits and identify AD specific associations. We performed a comprehensive untargeted lipidomic analysis, using Ultra-Performance Liquid Chromatography/Mass Spectrometry generating >600 features and a Genome Wide Association (GWA) study followed by imputation (>6 million SNPs). Following genetic and metabolomic data quality control, linear regression analyses were run to identify genetic influences on each metabolic feature. Results were pooled together by inverse-weighted meta-analysis. We report novel and previously reported associations between genetic loci and metabolites. We identified >30 associations at p<1*10-6 and 3 associations at p< 5*10-8 in stage 1, as well as evidence for associations between loci and two metabolites previously shown to be altered in AD (p<5*10-8). Results of stage 2 and meta-analysis will be reported. This is the first study to investigate the genetic influences on metabolite levels in an AD cohort and allow us to gain further insights into the genetic control of metabolites in plasma related to AD pathogenesis, which lead to a greater understanding of disease processes and help identify potential new opportunities for drug development.

1167T

Genome-wide meta-analysis identifies three loci for common forms of epilepsy. *J.P. Bradfield, The International League Against Epilepsy Consortium on Complex Epilepsies.* Center for Applied Genomics, Childrens Hospital of Philadelphia, Philadelphia, PA.

Epilepsy is a common complex disorder characterized by recurrent seizures of different etiology. While genome-wide association studies have been successful in identifying associations in a number of complex diseases, they have had limited success in identifying risk loci in epilepsy due in part to relatively small sample sizes and lack of power. To search for common variants associated with this disorder, we performed a meta-analysis of genome wide association studies of common forms of epilepsy. Here we report three meta-analyses conducted for all epilepsy and the two largest clinical subgroups - genetic generalized epilepsy (GGE) and focal epilepsy. We combined 9 different GWAS's comprising a total of 8,696 (2606 GGE/5310 Focal) cases and 26,157 controls that were imputed to the 1000 Genomes Phase I (interim), using the June 2011 reference. Each genome wide association was carried out using linear mixed models to control for potential population stratification and meta-analyzed with METAL. Meta-analysis of the "all epilepsy" phenotype identified 2 loci at genome wide significance ($P<1.66\times10^{-8}$). The locus at 2q24.3 (rs6732655, $P=8.7\times10^{-10}$) contained numerous genes with the strongest association coming from within *SCN1A*, while the locus at 4p15.1 (rs28498976, $P=5.44\times10^{-9}$) harbored just one gene (*PCDH7*). For the GGE cohort, a single locus passing genome wide significance was observed at 2p16.1 (rs2947349, $P=9.99\times10^{-9}$), which contained the genes *VRK2* and *FANCL*. Even though the focal epilepsy cohort was more powered than the GGE cohort, no genome wide significant SNP's were observed. *SCN1A* is a voltage gated sodium ion channel that has been implicated in numerous monogenic forms of epilepsy such as Dravet syndrome and genetic epilepsy with febrile seizures plus (GEFS+). *PCDH7*, a protocadherin gene, would be a novel candidate locus for epilepsy but another gene in this family (*PCDH19*) is a known cause of epilepsy and mental retardation limited to females (EFMR). The region at 2p16.1 that was associated in the GGE cohort was previously reported by another GWAS that is a constituent of our meta-analysis.

1168S

Investigating polygenic contributions of common hippocampal variants to epilepsy predisposition. *C.D. Whelan[1], D.P. Hibar[2], J. Stein[2], D. Speed[3], S. Sisodiya[4], M. Johnson[5], D. Goldstein[6], N. Delanty[7], S.E. Medland[1], P.M. Thompson[8], G.L. Cavalleri[2], The International League Against Epilepsy Consortium on Genetics of Complex Epilepsies.* 1) Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Leinster, Ireland; 2) Imaging Genetics Center, Keck School of Medicine, University of Southern California; 3) UCL Genetics Institute, University College London; 4) Institute of Neurology, University College London; 5) Division of Brain Sciences, Imperial College London; 6) Duke Centre for Human Genetic Variation, North Carolina; 7) Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland, Australia; 8) Departments of Human Genetics and Psychiatry, Radboud university medical center, Nijmegen, The Netherlands.

Hippocampal sclerosis (HS) is a common feature of localisation-related epilepsies (LREs), present in 50-75% of all surgical resections in the disorder. The underlying cause of HS is debated. Animal models and post-mortem cell counts suggest that HS can result from epileptogenesis, and some MRI investigations have highlighted a familial component to HS and concomitant neuronal loss within hippocampal regions. Recently, the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium identified genome-wide significant signals correlating with hippocampal volume, in a study of 29,037 individuals. We tested the hypothesis that variants predisposing to changes in hippocampal volume may, as a group, contribute to epilepsy predisposition. To investigate this, we summarised variation across nominally-associated ENIGMA SNPs into quantitative 'risk' scores, weighted for local linkage disequilibrium and effect size, and related these scores to disease state in (i) a phenotypically mixed sample of epilepsy patients (n=2,502), (ii) four epilepsy 'subtypes', including LREs (n=1,801), lesional epilepsies (n=280), nonlesional epilepsies (n=614) and idiopathic generalised epilepsies (IGEs; n=194) and (iii) an independent sample of healthy controls (n=5,191). Results did not reveal a significant association between disease state and risk score: observed scores only explained a small fraction (0-0.2%) of total variance in our risk model. Our findings suggest that being genetically predisposed to having smaller hippocampal volume may not be a strong risk factor for epilepsy. However, further analyses in additional epilepsy patients and healthy control cohorts are required to conclusively confirm or reject this position.

1169M

Using polygenic risk scores of psychiatric disorders to predict Neuroticism. *L. Colodro Conde[1,2], K. Verweij[3], E. Byrne[4], R.A. Power[5], N. Martin[1], S. Medland[1].* 1) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 2) University of Murcia, Spain; 3) Vrije Universiteit Amsterdam, The Netherlands; 4) University of Queensland, Australia; 5) SGDP Centre, Institute of Psychiatry, King's College London, United Kingdom.

Neuroticism is conceived of as an endophenotype for major depressive disorder (MDD). Neuroticism and MDD are moderately heritable and 45-55% of the genetic liability of MDD is shared with Neuroticism. To test if Neuroticism provides an index of the risk for MDD, we examined the extent to which Neuroticism can be predicted by a polygenic risk score (PRS) derived from the Psychiatric Genomics Consortium MDD analyses. Additionally, we sought to determine if including the PRSs for ADHD, Autism (AUT), bipolar disorder (BP) and schizophrenia (SCZ) accounted for additional variation in Neuroticism. Phenotypic information was obtained from four studies undertaken at the QIMR Berghofer Medical Research Institute. Phenotypes were available for 20702 individuals, of whom 5295 were genotyped. Personality data were collected using the Eysenck Personality Questionnaire. Scale scores in Neuroticism were transformed by taking the arcsine of the square root, corrected for age and sex effects and standardised. After selecting one individual per family, data were available for 2549 individuals (36.2% males, mean age = 35.83, SD = 12.14). PRSs were calculated using the PLINK profile score method A multiple regression on the profile scores of Neuroticism was performed controlling for ancestry (using three principal components). Preliminary analyses found no effects of ancestry on Neuroticism (p=.55). PRSs were calculated for clumped SNPSs using thresholds of $p \leq 0.001$, 0.01, 0.1, 0.5 or 1. As hypothesised, the MDD PRSs predicted Neuroticism explaining 0.17, 0.31, 0.49, 0.48 and 0.45% of the variance (p= 0.04, 0.005, 4.2×10-4, 5.0×10-4 and 7.2×10-4 respectively). However, adding the corresponding PRSs for ADHD, AUT, SCZ and BP did not result in a significant improvement in fit. To determine if BP PRSs could predict Neuroticism in the absence of the MDD PRS, we repeated this sequence of analyses using the BP risk scores. While nominally significant at a $p \leq 0.001$, 0.01 and 0.1 thresholds, these results did not survive correction for multiple testing. These results confirm the findings of previous studies which have shown a moderate genetic correlation between MDD and Neuroticism. However, adding the PRSs for ADHD, AUT, SCZ and BP did not improve the prediction of Neuroticism. Although limited by the power of both the PRS analysis and the original PGC GWAS, the present analyses support the use of Neuroticism as a specific endophenotype for MDD.

1170T

Machine Learning Derived Disease Risk Prediction for Anorexia Nervosa. *Y. Guo[1], Z. Wei[2], B. Keating[1,3], H. Hakonarson[1,3], GCAN, WTCCC3.* 1) The Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 2) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA; 3) Department of Pediatrics, School of Medicine University of Pennsylvania, Philadelphia, PA 19104, USA.

Anorexia nervosa (AN) is a serious complex psychiatric disease with a strong contribution from the patients' genetic background. In addition to identifying signals through genome wide association (GWA) studies, researchers have been using genetic information and machine learning methods to predict risk of diseases in which genetics play an important role. In this report we collected whole genome genotyping data on 3,940 AN cases and 9,266 controls from the Genetic Consortium for Anorexia Nervosa (GCAN), the Wellcome Trust Case Control Consortium 3 (WTCCC3) and the Children's Hospital of Philadelphia (CHOP), and applied machine learning methods for predicting AN disease risk. The prediction performance is measured by area under the receiver operating characteristic curve (AUC), indicating how well the model distinguishes cases versus controls. Logistic regression model with the lasso penalty technique generated an AUC of 0.693, while Support Vector Machines and Gradient Boosted Trees reached AUC's of 0.691 and 0.623, respectively. Our results of different sample sizes also suggested that larger datasets are required to optimize the machine learning models and achieve higher AUC values. To our knowledge, this is the first attempt to assess AN risk by incorporating genetic data, and it will pave the way for improved AN risk evaluation, eventually benefitting AN patients and families.

1171S

Genome-wide Association Study of Quantitative Autistic Traits in the General Population. *T. Nishiyama[1], S. Hosono[2], M. Watanabe[2], I. Oze[2], H. Ito[2], H. Tanaka[2], H. Kishino[3], T. Kawaguchi[4], F. Matsuda[4], K. Matsuo[5].* 1) Department of Public Health, Aichi Medical University School of Medicine, Nagakute, Japan; 2) Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan; 3) Laboratory of Biometry and Bioinformatics, Graduate School of Agriculture and Life Sciences, University of Tokyo, Tokyo, Japan; 4) Center for Genomic Medicine, Kyoto University, Kyoto, Japan; 5) Department of Preventive Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan.

Copy number variations and single nucleotide variants, many of which are highly penetrant, are increasingly identified to confer susceptibility to autism spectrum disorder (ASD). In contrast, only a modest number of common variants associated with ASD have been reported through genome-wide association studies (GWAS) (Weiss LA 2009, Wang K 2009, Anney R 2010, Xia K 2013). These common variants have weak effects on ASD susceptibility, and thus sample size would be prohibitively large in practice. Recent studies both of the general population and of twins have shown that quantitative autistic traits are continuously distributed with no natural boundary between normality and ASD (Baron-Cohen 2001, Constantino & Todd 2003), leading to the notions that the extreme of quantitative autistic traits represents the clinical syndrome of autism. Based on this dimensional conceptualization, we conducted GWAS of quantitative autistic traits in the general population, instead of relying on case-control sample of ASD. This methodology seems promising because of easy sampling.

In this study, we used non-cancer participants in the Hospital-based Epidemiological Research Program II at Aichi Cancer Center Hospital (HERPACC-II) (Hamajima N 2001). Quantitative autistic traits were measured using the Japanese version of the Subthreshold Autistic Traits Questionnaire (SATQ) (Nishiyama T 2014) in the second wave of surveys. The sample was genotyped on an IlluminaHuman 610 Quad BeadChip with 576,736 SNP markers. The present study was approved by the Ethics Committee of Aichi Cancer Center and informed consent was obtained from all participants.

After removing SNPs that failed the quality control criteria a priori defined, 488,823 markers were used for the analysis. The snpStats package in R was used to test the association between SNP genotypes and the SATQ score, using additive models adjusted for gender. Using only 358 subjects currently available (we have a plan to increase up to around 1000 in this year), we found four SNPs (rs11577575, rs10071163, rs11862060 and rs10446439) significant at p-value $< 1 \times 10^{-5}$. These four SNPs are located in/near ARHGAP29, PDE4D, ABAT and LSM3-SLC6A6, respectively, and were reported to be associated with ASD in some way. This result suggests potential advantages of the methodology using quantitative autistic traits in the general population.

1172M

Diagnostic exome sequencing of patients with Autism Spectrum Disorder overwhelmingly detects mutations in newly characterized genes, which supports a *de novo* paradigm and the convergence of disrupted pathways in neurodevelopmental disease. *Z. Powis[1], C. Mroske[1], K. Radtke[1], D. Shinde[1], K.F. Gonzalez[1], L. Shahmirzadi[1], D. El-Khechen[1], B. Tippin[1], E. Chao[1,2], R. Baxter[1], S. Tang[1].* 1) Ambry Genetics, Aliso Viejo, CA., Select a Country; 2) University of California Irvine, Irvine, CA 92697.

Diagnostic exome sequencing (DES) is an increasingly effective tool for diagnosis in intractable cases where the underlying cause is believed be genetic. Autism Spectrum Disorders (ASDs) have traditionally been thought to be highly heritable but often the exact genetic origin has been difficult to elucidate. It is estimated that the positive rate for ASD patients is around 15% with the three first tier tests (karyotyping, CMA, and fragile X). However, the diagnostic yield for the other identifiable etiologies for ASD remains unknown. Of the first 500 unselected cases referred for DES to Ambry Genetics (Aliso Viejo CA), 50 (10%) were reported to have an ASD. All of these individuals had additional features such as seizures, intellectual disability, dysmorphic features or other organ involvement, and 40 (80%) had previous uninformative results in at least one of the three first-tier tests. For these 50 ASD patients, trio DES analyses were performed for 41 families. Among them, mutations in characterized genes have been detected in 12 (24%) (*ANK2, CDKL5, CHD8, ELP2, FHL1, IQSEC2* (2 cases), *KMT2A, PANK2, SET5D, SHANK3*, and one individual with mutations in both *ANO3* and *NALCN*). 10 of these cases (83%) were *de novo* alterations. In addition, 7 patients (14.0%) had an uncertain result (*SYP, KDM6B, MTRR, PIGA, ZBTB20, ADCY5*, and *SHROOM4*). 6 of these 7 cases were inherited (85.7%). 62 (4%) of these cases had a novel result (*HDAC1* and *MTOR*), and 29 (58%) had a negative result. Interestingly, among the remaining 450 patients without ASD as an indication for testing, de novo loss-of-function mutations in recently discovered high-confidence autism genes were identified in 7 cases: *ADNP* and *ARID1B*, 2 cases each, *CHD2, DYRK1A, GRIN2B*). These observations indicated that 24% of syndromic ASD patients received a definitive molecular diagnosis via DES after negative first-tier genetic testing. The vast majority of the associated genes were newly identified, non-classical ASD genes and thus might evade traditional sequential or panel testing. The rapid progress being made to the understanding of ASD genetics accentuates the great potential and continuing challenges for genotype-phenotype correlation in the clinical setting. Our clinical data also strongly support the *de novo* paradigm as well as the convergence of disrupted pathways in neurodevelopmental disorders.

1173T

Common polygenic variation and risk for childhood-onset schizophrenia. *K. Ahn[1], S. An[2], J. Rapoport[1].* 1) Child Psychiatry Branch, National Inst Mental Health, NIH, Bethesda, MD; 2) Department of Environmental Health Sciences, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD.

Childhood-onset schizophrenia (COS) is a rare and severe form of the disorder, with more striking abnormalities in brain development and more striking uneven prepsychotic development compared to later onset disorder. We previously documented that COS patients, compared with their healthy siblings and with adult-onset patients (AOS), carry substantially more rare (0.1% in control) chromosomal copy number variations (CNVs), spanning large genomic regions (> 100kb). (Ahn et al. in press). Here we interrogated the contribution of common polygenic variation to the genetic susceptibility for schizophrenia. We examined the association between a direct measure of genetic risk of schizophrenia in 126 COS probands and 98 healthy siblings. Using data from the schizophrenia and autism GWAS of the Psychiatric Genomic Consortia, we selected risk related sets of SNPs from which we conducted polygenic risk score comparisons for COS probands and their healthy siblings. COS probands had higher genetic risk scores of schizophrenia and autism than their siblings ($P<0.05$). These initial results show that childhood onset schizophrenia shares common risk alleles with autism. Finally, these findings suggest that COS patients may have more salient genetic risk than do adult onset patients.

1174S

A psychometric GWAS finds specificity of variants associated with level and change in immediate and delayed verbal memory after age 60. *T.E. Arpawong[1], J.J. McArdle[1,2], C.A. Prescott[1,2].* 1) Psychology, University of Southern California, Los Angeles, CA; 2) Gerontology, University of Southern California, Los Angeles, CA.

Substantial individual differences exist in memory ability and the rate of age-related memory decline among older adults. Phenotypically, there is a strong association between immediate recall (IR) and delayed recall (DR), but candidate gene findings suggest that genetic influences on IR and DR may be partially distinct, with stronger genetic influence on initial level of recall than on rate of decline. However, this may be due to measurement error; with limited measurement occasions, there is greater precision for estimating level than change. Our goal was to combine a psychometric approach with GWAS to identify whether unique genetic variants are associated with level and change in IR and DR in older age. The sample comprised 12,326 participants (mean age=66.7, range of 50 to107) in the Health and Retirement Study (HRS) who provided DNA and cognitive data (77,693 data points) and were assessed up to 8 times between 1996 through 2012. IR and DR were assessed with an auditory learning task in which individuals were read a list of 10 nouns, asked to immediately recall the list, then asked to recount them after a several-minute delay. Genetic data comprised 2.4 million SNPs from the Illumina Human Omni-2.5 Quad beadchip. We used the following analytical steps: (1) Applied a two-spline mixed effects models to estimate each individual's: level at age 60 and rate of change after 60 in both IR and DR; (2) Conducted four GWAS for IR and DR level and decline, adjusting for sex and ancestral informative markers; (3) Examined whether the same SNPs associated with level and slope of IR were similarly associated with level and slope of DR. The results await replication; however, several SNPs were strongly associated with each outcome (P<5.0×10e-6), and the implicated regions for IR and DR level and decline were mostly non-overlapping.

1175M

GWAS analysis of Insight into illness in Schizophrenia. *V. De Luca, A. Bani-Fatemi, A. Graff, P. Gerretsen.* Dept Neurogenetics, Univ Toronto, Toronto, ON, Canada.

Impaired insight into illness is a core feature of schizophrenia characterized by a failure to recognize the presence of a mental illness, symptom misattribution, and unawareness of the need for treatment or negative consequences associated with the disorder. Impaired insight is associated with treatment non-adherence and poor clinical outcomes. The degree of insight impairment tends to remain stable over the illness course, improving modestly with age, but can fluctuate with illness exacerbations, i.e. psychotic episodes. Impaired insight is associated with a number of premorbid factors, such as lower functioning, education, IQ, and personality traits. In particular, impaired insight into illness is consistently associated with illness severity, global cognitive impairment, lower premorbid intelligence (i.e. IQ), executive dysfunction, and memory deficits. A number of genes are implicated in schizophrenia, including DISC1 and NRG1. Yet, despite the high heritability of schizophrenia (81%), few genes are reliably replicated and together explain only a small proportion of the disorder's variance. Due to the heterogeneity of schizophrenia and the overlap between susceptibility genes for schizophrenia and related conditions, it is possible that these genes may contribute more substantially to certain clinical subtypes, such as schizophrenia with impaired insight. As such, the aim of this study was to perform an exploratory genome wide association study (GWAS) to identify candidate genes for impaired insight into illness in a sample of 123 schizophrenia patients. No variants passed the genome-wide significance threshold. However, insight can be used as valuable tool to identify subtype of schizophrenia with higher genetic susceptibility.

1176T

Genetic influences of language development in typically developing children and children with autism spectrum disorders. *J.D. Eicher[1], J.R. Gruen[1,2], Pediatric Imaging Neurocognition Genetics study.* 1) Department of Genetics, Yale University, New Haven, CT; 2) Departments of Pediatrics and Investigative Medicine, Yale University School of Medicine, New Haven, CT.

The ability to effectively communicate is an essential milestone of child development, particularly in social interaction and academic achievement. Deficits in communication skills severely hamper development and lead to long-term negative consequences. One of these vital communication skills is receptive vocabulary, the ability to comprehend and respond to written/verbal language. Children with various language disorders, including dyslexia, language impairment, and autism spectrum disorders, have deficits in receptive vocabulary. As each of these disorders is a complex trait with a significant genetic component, we examined the genetics of receptive vocabulary skills in typically developing children and in children with autism spectrum disorders. First, we performed a genome wide association study (GWAS) of receptive vocabulary in 440 typically developing children in the Pediatric Imaging Neurocognition and Genetics (PING) study. There were associations with markers in *MAGI-1* ($p=3.66\times10^{-8}$) and *BMP7* ($p=8.25\times10^{-7}$). *MAGI-1,* a scaffolding protein important in cell-cell junctions, has previously been implicated with bipolar disorder, schizophrenia, associative learning, and modulation of AMPA synaptic behavior. *BMP7* is important in neural development and was recently identified as a prospective candidate gene for language related disorders (Boeckx and Benitez-Burraco 2014). Next, we assessed whether genes previously associated with language—including *MAGI-1*, *BMP7*, *DCDC2*, *KIAA0319*, *MRPL19*/*GCFC2*, *ATP2C2*, *CMIP*, and *KIAA0319L*, among others—were associated with receptive vocabulary in a meta-analysis of two autism spectrum disorders cohorts: the Autism Genome Research Exchange (AGRE) and Simons Simplex Collection (SSC). We found evidence of associations with *ATP2C2* ($p=4.89\times10^{-6}$), *KIAA0319L* ($p=1.00\times10^{-5}$), *MAGI-1* ($p=3.00\times10^{-5}$), and *MRPL19* ($p=3.91\times10^{-5}$), indicating that genes previously associated with other language disorders (such as dyslexia and language impairment) also influence communication skills in children with autism spectrum disorders. These studies identify genes that influence childhood communication development and indicate that disorders with language-associated deficits share genetic etiologies. Future studies should further examine to what extent these disorders are related using human-based and model-based approaches.

1177S

Genome-wide meta-analysis reveals significant association between CHRNA4 variants and nicotine dependence in cohorts of European ancestry. *N.C. Gaddis[1], D.B. Hancock[1], G.W. Reginsson[2], N.L. Saccone[3], S.M. Lutz[4], S. Steinberg[2], P. Kraft[5], J.E. Hokanson[4], L.J. Bierut[3], T.E. Thorgeirsson[2], E.O. Johnson[1], K. Stefansson[2].* 1) RTI International, Research Triangle Park, NC; 2) deCODE Genetics, Reykjavik, Iceland; 3) Washington University, St. Louis, MO; 4) University of Colorado Anschutz Medical Campus, Aurora, CO; 5) Harvard University School of Public Health, Boston, MA.

Cigarette smoking is a major contributor to cancer, vascular disease, and lung disease and is the leading cause of preventable mortality worldwide. Previous genome-wide studies of smoking-related phenotypes have convincingly demonstrated associations with the nicotinic acetylcholine receptor genes on chromosome 15q25 (CHRNA5-CHRNA3-CHRNB4) and chromosome 8p11 (CHRNB3-CHRNA6). To identify novel genetic loci associated with smoking, we conducted a genome-wide association study (GWAS) and meta-analysis of nicotine dependence (ND) in five cohorts of European ancestry totaling 16,244 subjects: the largest GWAS meta-analysis of ND to date. We defined ND using the Fagerstrom Test of Nicotine Dependence (FTND), which is a multi-dimensional measure of physiological dependence to nicotine. After completing standard quality control and 1000 Genomes imputation for each of the five cohorts, we tested approximately 10 million genetic variants for association with a three-level FTND-based phenotype (mild, moderate, and severe ND), separately by cohort, using a linear regression model adjusted for age, sex, eigenvectors to correct for population stratification, and other cohort-specific covariates as needed. We then combined the results from the five cohorts in an inverse variance-weighted meta-analysis with genomic control applied to each cohort. As expected, variants in the CHRNA5-CHRNA3-CHRNB4 region displayed the strongest association with FTND, with p-values as low as 10^{-17}. We also observed genome-wide significant association for a variant in the 3' untranslated region of CHRNA4 ($p = 3.60\times10^{-8}$), with odds ratios for the five cohorts ranging from 1.05-1.20 and 1.11-1.43 for moderate and severe ND compared to mild ND, respectively. This variant is predicted to alter a microRNA binding site that may regulate CHRNA4 expression. In total, we found 22 CHRNA4 variants with p-values less than 10^{-5}, all of which had consistent directions of association across the five cohorts: their minor alleles (frequency range 4-20%) being associated with an increased risk of ND. CHRNA4 is known to have high affinity for nicotine, and our findings build on the evidence from several smaller studies suggesting that CHRNA4 variants contribute to smoking-related phenotypes.

1178M

Identification of novel candidate genes in canine noise phobia -a model for human panic disorder. *O. Hakosalo[1,2], K. Tiira[1,2], R. Sarviaho[1,2], M. Sillanpää[3], J. Kere[4,5], H. Lohi[1,2].* 1) Department of Veterinary Biosciences and Research Programs Unit, Molecular Neurology, University of Helsinki, P.O.Box 63, 00014 University of Helsinki, Helsinki, Finland; 2) The Folkhälsan Research Center, Helsinki, Finland; 3) Department of Biology and Mathematical Sciences, University of Oulu; 4) Department of Biosciences and Nutrition, Karolinska Institutet, and Center for Biotechnology, Sweden; 5) Research Programs Unit, University of Helsinki.

Noise phobia (NP), a fear of loud noises, is a severe anxiety disorder in dogs, which may result in panic attacks. It provides a natural animal model for human panic disorder. NP is common (20-40%) in many breeds with high heritability (h^2 0.56) estimates but its genetic background remains unknown. We aimed to discover the genetic cause by a genome wide association analysis. We focused on German Shepherds because the breed is popular and presents sufficient phenotypic variation in NP. We used our validated owner-completed anxiety questionnaire to create a categorical NP phenotype and genotyped altogether 310 German Shepherds (GS) (86 cases and 224 controls) to map the NP locus. We found a genome widely significant association in a 4 Mb region at CFA20 harboring several interesting candidate genes. Targeted capture and resequencing of the associated region from 39 dogs revealed several disease segregating variants that were enriched in three promising candidate genes. Further validation and functional characterization of variants in a larger GS cohort and other breeds should implicate the causative gene. This study will improve the understanding of the genetic background of noise phobia, and at best reveal a new candidate gene for human anxiety and help in developing a model for potential therapeutic approaches.

1179T

Core-Exome Chip study of low-frequency variants identifies genome-wide significant hits associated with anorexia nervosa. *K. Hatzikotoulas[1], L.M. Huckins[1], L. Thornton[2], L. Southam[1], D. Collier[3], P. Sullivan[2], C.M. Bulik[2], E. Zeggini[1], GCAN Consortium.* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 3) Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, UK.

Anorexia nervosa (AN) is a neuropsychiatric disorder presenting with extremely low body weight, and a marked fear of gaining weight. AN has the highest mortality of any psychiatric disorder, and affects roughly 0.9% of women. Very little is known about the biological mechanisms which underlie AN; two GWAS have been completed and have yet to identify genome-wide significant hits. No effective medications are available, and treatment outcome for AN remains unacceptably poor. Our study comprises 2,376 female AN cases and >22,000 controls, genotyped on the Core-Exome Chip. Samples derive from eight different populations; cases and controls are ancestrally matched. The CoreExome Chip enables us to study both common and low-frequency variants simultaneously; our study is the first to examine the role played by low-frequency and rare variants in AN. Analysis is currently complete across three of the eight contributing populations: Norway, 87 cases, 100 controls; Finland, 163 cases, 5,300 controls; and the UK, 181 cases and 10,034 controls. We have performed a meta-analysis across these three populations and thus far have identified four genome-wide significant signals: exm370124, exm462797, exm464785, exm2116552. These four variants are all low frequency, missense variants. We looked at the frequency of these SNPs in both cases and controls. All SNPs were extremely low frequency in the control populations, with highest MAF between 0.005 and 0.01. SNPs were also low frequency in the cases, with highest MAF between 0.01 and 0.10. Effect sizes for each SNP were high, and the same direction of effect was noted for every SNP in at least 2/3 populations. Maximum effect sizes for each SNP were between 6.6 and 74.5. One of these associated variants (exm464785) lies in RASGRF2, a gene that has previously been associated with eating disorders (Wade et al. 2013), albeit not at a genome-wide significant level. This is the first genome-wide significant variant that has been associated with AN. We hope that this will enable further studies into the functional mechanisms underlying AN, and perhaps be a first step towards establishing effective medications and treatment. Further, all four hits that have been identified are very low frequency and could not possibly have been identified in previous GWAS studies. This may be a good indication that low-frequency, Core-Exome chip type studies have potential to reveal new associated variants across a range of psychiatric disorders.

1180S

Multi-ethnic meta-analysis in a cohort of 110,266 individuals identifies novel shared and sex-specific loci associated with smoking initiation. *E. Jorgenson[1], L. Shen[1], A. Hamidovic[2], H. Choquet[3], T. Hoffmann[3], Y. Banda[3], M. Kvale[3], N. Risch[3], C. Schaefer[1], L. Sakoda[1].* 1) Kaiser Permanente Division of Research, Oakland, CA; 2) Departments of Pharmacy and Psychiatry University of New Mexico, Albuquerque, NM; 3) Institute for Human Genetics University of California, San Francisco San Francisco, CA 94143.

Smoking is a major public health problem in which both genetic and environmental factors play an important role. Genome-wide association studies have identified six loci influencing smoking behavior, and additional investigation is needed to clarify how these and other genetic factors affect the chronology of smoking behavior, specifically smoking initiation, persistence, and cessation. Family and twin studies have suggested that significant sex differences exist in smoking behavior, with the heritability of smoking initiation being higher in women than in men. We conducted a genome-wide association meta-analysis of smoking initiation across race/ethnicity (non-Hispanic white, Latino, Asian, and African American subjects) and sex strata in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. Over 6.6 million SNPs were imputed with imputation r2>0.8 and minor allele frequency > 0.05 using the 1000 Genomes reference panel. We identified four loci that exceeded a genome-wide significance level of 5×10-8. An intergenic locus near TMEM182 on chromosome 2 (p=2.65×10-8) and SNPs in SDK1 (p=2.52×10-8) were associated with smoking initiation in both men and women. Additionally, SNPs in BDNF (p=3.48×10-8) and NCAM1 (p=9.20×10-9) were associated with smoking initiation in women but not in men. While BDNF was previously reported to be associated with both smoking initiation and persistence, our other findings are novel. We examined additional evidence for the association of these loci with smoking initiation in the combined-sex association results among European populations made publicly available by the Psychiatric Genomics Consortium. Those results provided support for both BDNF (p=2.01×10-5) and NCAM1 (p=6.76×10-3), but not for the other two loci (p>0.05), indicating both a need for additional replication in large, multi-ethnic samples and the power of combining evidence across race/ethnicity groups to provide additional insights into the genetics of smoking behavior.

1181M

Genome-Wide Association Study in *APOE* ε4 Negative Subjects Identifies a Novel Locus in 17q21.31 for Alzheimer Disease. *G. Jun[1,2,3], C. Ibrahim-Verbaas[6], C. Bellenguez[7,8,9], M. Vronskaya[11], J. Chung[1], J.C. Bis[12], J. Williams[11], S. Seshadri[5], G.D. Schellenberg[13], K.L. Lunetta[2], P. Amouyel[7,8,9,10], P. Holmans[11], C.M. van Duijn[6], L.A. Farrer[1,2,3,4,5], International Genomics of Alzheimer's Project (IGAP) Consortium.* 1) Medicine, Boston University, Boston, MA, USA; 2) Biostatistics, Boston University, Boston, MA, USA; 3) Ophthalmology, Boston University, Boston, MA, USA; 4) Epidemiology, Boston University, Boston, MA, USA; 5) Neurology, Boston University, Boston, MA, USA; 6) Epidemiology, Clinical Genetics and Neurology, Erasmus MC University Medical Center, Rotterdam, the Netherlands; 7) Inserm, U744, Lille, France; 8) Université Lille 2, Lille, France; 9) Institut Pasteur de Lille, Lille, France; 10) CNR-MAJ, Centre Hospitalier Régional Universitaire de Lille, Lille, France; 11) Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University, UK; 12) Medicine, University of Washington, Seattle, WA, USA; 13) Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA.

APOE ε4 is the most significant genetic risk factor for Alzheimer disease (AD) and may mask effects of other loci. We re-analyzed genome-wide association study (GWAS) data from the International Genomics of Alzheimer's Project (IGAP) Consortium in *APOE* ε4-positive (10,246 cases and 11,924 controls) and *APOE* ε4-negative (7,231 cases and 19,603 controls) subgroups as well as in the total sample allowing for interaction between a SNP and *APOE* ε4 dosage. Suggestive associations ($p<1\times10^{-4}$) in the discovery sample were evaluated in an independent replication sample containing 4,474 subjects (ε4-positive: 1,250 cases and 536 controls, ε4-negative: 718 cases and 1,699 controls). In the discovery set, we observed suggestive associations from 15 loci in either ε4-positive or ε4-negative subjects (5 novel loci in ε4-negative subgroup: *SOX/CLDN18*, *CDC42SE2/ACSL6*, *PFDN1/HBEGF*, *KANSL1/LRRC37A*, and *CDR2L*) and two novel loci in the total sample for interaction (*THSD7A/TMEM106B* and *ABCA13/CDC14BL*). Of these, five distinct loci from 4 known (*CR1*, *BIN1*, *CLU*, and *MS4A* cluster) and one novel locus (*KANSL1/LRRC37A*) were genome-wide significant ($p<5\times10^{-8}$) in meta-analysis of the discovery and replication in either ε4-positive or ε4-negative subjects. Among the loci previously identified by AD GWAS, SNPs in *BIN1* were genome-wide significant in both *APOE* subgroups. SNPs in *CR1* and *CLU* were genome-wide significant in ε4-positive subjects, whereas SNPs in the *MS4A* cluster region (*MS4A6A/MS4A4A/MS4A6E*) were genome-wide significant in ε4-negative subjects. Novel associations among ε4-negative subjects were observed in the *MAPT* region covering approximately 100 kb including *KANSL1* and *LRRC37A* (best SNP, rs2732703, ε4-positive: OR=0.86, CI=0.76-0.98, p=0.02; ε4-negative: OR=0.73, CI=0.65-0.81, $p=5.8\times10^{-9}$). Although 17 SNPs in this region initially showed genome-wide significant p-values, analyses conditioned on the lead SNP (rs2732703) revealed no additional independent associations. However, despite these findings, no SNPs showed statistical evidence of genome-wide significant interaction by ε4 status in the meta-analysis of discovery and replication. Association of variants in the *MAPT* region with AD may represent a novel genetic risk factor among individuals who do not carry the ε4 risk allele. Further examination of this region is warranted to identify functional determinants of AD risk among persons lacking the ε4 allele.

1182T

First GWAS in DBH confirms strong cis-acting variants and lends support for its role as an intermediate phenotype in post-traumatic stress disorder. *A.X. Maihofer[1], M. Mustapic[1,2,4], D.G. Baker[3], D.T. O'Connor[2], C.M. Nievergelt[1,3].* 1) Department of Psychiatry, University of California at San Diego, La Jolla, CA; 2) Department of Medicine, University of California at San Diego, La Jolla, CA; 3) VA Center of Excellence for Stress and Mental Health (CESAMH), VA San Diego Healthcare System, La Jolla, CA; 4) Division of Molecular Medicine, Ruđer Bošković Institute, Zagreb, Croatia.

Dopamine beta-hydroxylase (DBH) catalyzes formation of norepinephrine. DBH is expressed in noradrenergic nerve terminals of the central and peripheral nervous systems, as well as in chromaffin cells of the adrenal medulla. DBH is present in cerebral spinal fluid and plasma as stable heritable trait. Differences in DBH expression or activity might reflect a role in the pathogenesis of cardiovascular and neuropsychiatric disorders. The genetic mechanisms underlying DBH activity and its secretion have been only partially explained. Thus we conducted a genome-wide association search for loci contributing to human plasma DBH (pDBH) activity. In a population sample of European ancestry, we identified 3 common trait-determining variants (top hit rs1611115, $p=7.2\times10^{-51}$) in the proximal DBH promoter. Each of the 3 variants had an effect on transcription and acted additively on gene expression. We replicated the associations in a population sample of Native American descent (top hit rs1611115, $p=4.1\times10^{-15}$). Additionally, we identified yet another genome-wide significant SNP at the LOC338797 locus on chromosome 12 as potential trans-quantitative trait locus (QTL) (rs4255618, $p=4.62\times10^{-8}$). Analysis conditioned on 3 DBH promoter variants identified a third genomic region on chromosome 9q contributing to pDBH variation: a likely cis-QTL adjacent to DBH in SARDH (rs7040170, $p=1.31\times10^{-14}$). The identification of SNPs with such strong effects on pDBH opens the possibility of utilizing Mendelian randomization (MR) approaches to test causal effects of this intermediate trait on disease. Therefore we performed an exploratory MR analysis in a sample of participants recruited from the Marine Resiliency Study (MRS), a large, prospective study of post-traumatic stress disorder (PTSD) involving active-duty United States Marines bound for deployment to Iraq or Afghanistan. The SNP rs1611115, which explained 50% of the variability in pDBH, was employed as a genetic instrument to test for a causal effect of plasma DBH on PTSD symptoms. The MR estimate of the association of pDBH and re-experiencing symptoms was significant (beta = 0.21, p = 0.002), indicating the possibility that pDBH is a causal component in the development of re-experiencing symptoms.

1183S

Genome-wide association study of sensory disturbances in the inferior alveolar nerve after bilateral sagittal split ramus osteotomy. *D. Nishizawa[1], D. Kobayashi[1,2,3], Y. Takasaki[4], S. Kasai[1], Y. Aoki[1,2], J. Hasegawa[1], T. Kakizawa[5], K. Ikeda[1], K. Fukuda[2].* 1) Psychiatry and Behavioral Science (Addictive Substance Project), Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) Department of Dental Anesthesiology, Tokyo Dental College, Tokyo, Japan; 3) Department of Dentistry and Oral surgery, Tokyo Metropolitan Tama Medical Center, Tokyo, Japan; 4) Department of Dentistry and Oral surgery, National Hospital Organization, Takasaki General Medical Center, Gunma, Japan; 5) Department of Oral Health and Clinical Science, Division of Oral and Maxillo-facial Surgery, Tokyo Dental College, Tokyo, Japan.

Background: Bilateral sagittal split ramus osteotomy (BSSRO) is a common orthognatic surgical procedure. Sensory disturbances in the inferior alveolar nerve, including hypoesthesia and dysesthesia, are frequently observed after BSSRO, even without distinct nerve injury. The mechanisms that underlie individual differences in the vulnerability to sensory disturbances have not yet been elucidated. **Methods**: The present study investigated the relationships between genetic polymorphisms and the vulnerability to sensory disturbances after BSSRO in a genome-wide association study (GWAS). A total of 304 and 303 patients who underwent BSSRO were included in the analyses of hypoesthesia and dysesthesia, respectively. Hypoesthesia was evaluated using the tactile test 1 week after surgery. Dysesthesia was evaluated by interview 4 weeks after surgery. Whole-genome genotyping was conducted using Illumina BeadChips including approximately 300,000 polymorphism markers. **Results**: Hypoesthesia and dysesthesia occurred in 51 (16.8%) and 149 (49.2%) subjects, respectively. Significant associations were not observed between the clinical data (i.e., age, sex, body weight, body height, loss of blood volume, migration length of bone fragments, nerve exposure, duration of anesthesia, and duration of surgery) and the frequencies of hypoesthesia and dysesthesia. Significant associations were found between hypoesthesia and the rs502281 polymorphism (recessive model: combined $\chi^2 = 24.72$, nominal $P = 6.633 \times 10^{-7}$), between hypoesthesia and the rs2063640 polymorphism (recessive model: combined $\chi^2 = 23.07$, nominal $P = 1.563 \times 10^{-6}$), and between dysesthesia and the nonsynonymous rs2677879 polymorphism (trend model: combined $\chi^2 = 16.56$, nominal $P = 4.722 \times 10^{-5}$; dominant model: combined $\chi^2 = 16.31$, nominal $P = 5.369 \times 10^{-5}$). The rs502281 and rs2063640 polymorphisms were located in the flanking region of the *ARID1B* and *ZPLD1* genes on chromosomes 6 and 3, whose official names are “AT rich interactive domain 1B (SWI1-like)” and “zona pellucida-like domain containing 1”, respectively. The rs2677879 polymorphism is located in the *METTL4* gene on chromosome 18, whose official name is “methyltransferase like 4”. **Conclusions**: The GWAS of sensory disturbances after BSSRO revealed associations between genetic polymorphisms located in the flanking region of the *ARID1B* and *ZPLD1* genes and hypoesthesia and between a nonsynonymous genetic polymorphism in the *METTL4* gene and dysesthesia.

1184M

Genes Involved in Brain Development Influence Crying Habits -A Genome Wide Association Study. *C. Tian, C.Y. McLean, E.Y. Durand, N. Eriksson, J.Y. Tung, D.A. Hinds.* Research, 23andme, Mountain View, CA.

Crying, for emotional reasons, is considered unique to humans. Crying habits vary greatly--some crying easily and others rarely. Thus far it is not clear why, though emotional stability has been shown to play a role and excessive emotionality appears to have a familial tendency. In this genome-wide association study, we searched for possible genetic associations with crying easily using a large sample of unrelated customers of 23andMe, Inc., with European ancestry. 60848 cases self reported that they ‘cry easily’ and 94871 controls said that they did not ‘cry easily’. Females are three times more likely to self-identify as ‘crying easily’ than males and individuals who cry easily have a significantly higher rate of psychiatric disorders in our data. We identified genome-wide significant associations with 7 loci, including rs62335062 (P=2.6×10-31, OR=1.111) upstream of IRX2, rs7196282 (P=5.0×10-14, OR=0.928) in an intron of ZNF423, rs2178574 (P=2.7×10-11, OR=1.066) in an intron of LRRTM4, rs2744475 (P=2.1×10-10, OR=1.060) upstream of TFAP2B, rs10838125 (P=2.5×10-10, OR=1.058) between TTC17 and HSD17B12, rs876714 (P=5.7×10-10, OR=1.054) in an intron of BIN3 and close to EGR3, and rs16903275 (P=3.9×10-8, OR=0.94) in microRNA9-2 (MIR9-2). The top two associated genes, IRX and ZNF423, have been shown to be involved in cerebellum development and brain regionalization. The cerebellum is an important brain region that has been linked to crying. LRRTM4 gene product is known to trigger the formation of excitatory synapses and shows highly selective expression in the brain. Mutations in LRRTM family genes have been associated with human handedness and schizophrenia. TFAP2B product, an important factor in the development of ectodermal and neural tissues, is involved in monoaminergic regulation and has been associated with neonatal temperament, alcohol addiction, adolescent depressive symptoms, and attention deficit hyperactivity disorder. The expression of EGR3 is rapidly regulated by neural synaptic activity in the brain cortex, which may be important in defining neuroplastic responses following stimulus. MIR9-2 has been associated with neuron-specific expression and neuronal differentiation during brain development. This study highlights some important genes that potentially impact the neural basis of crying in response to emotional stimulus, which may also shed light on the disease etiology of more serious affective disorders.

1185T

Identification of a novel locus for human-directed fear in dogs. *K. Tiira[1, 2], O. Hakosalo[1, 2], R. Sarviaho[1, 2], M. Sillanpää[3], J. Kere[2,4,5], H. Lohi[1, 2].* 1) Department of Veterinary Biosciences and Research Programs Unit, Molecular Neurology, P.O.Box 63, 00014 University of Helsinki, Helsinki, Finland; 2) The Folkhälsan Research Center, Helsinki, Finland; 3) Department of Biology and Mathematical Sciences, University of Oulu; 4) Department of Biosciences and Nutrition, Karolinska Institutet, and Center for Biotechnology, Sweden; 5) Research Programs Unit, University of Helsinki.

Behavioural, pharmacological, clinical and etiological studies suggest that dogs suffer from human anxiety-like disorders with shared neurobiological etiologies. Fearfulness has obtained high heritabilities in dogs (h^2 0.46) and inbred population structures facilitate gene mapping. We have initiated a large program to dissect the genetic and environmental correlates of different canine anxieties. We aimed here to identify the genetic cause of canine fearfulness in relation to various environmental correlates. We utilized our validated anxiety questionnaire to data collection and to derive several continuous and binomial behavioural variables that describe the dog’s fearfulness. We derived a Human fear-variable, which is a continuous variable emphasizing the avoidance behaviour towards strangers and is highly correlated with generalized fearfulness in the data set ($r_s = 0.92$, P<0.0001). We selected 126 Great Danes (52 cases and 74 controls) with extreme ends of the stranger avoidance behaviour for a genome-wide association study to map the fear locus using canine HD SNP chip arrays (173K). We identified a genome wide significant association in a 4 Mb region at CFA11 using a linear mixed model analysis with puppy socialization as a covariate. Our ongoing analyses aim to replicate the association in larger cohorts available across breeds and to identify the causative gene. This is the first report for a fear locus in dogs and provides insights to the genetic and environmental factors in canine anxiety with a translational potential to human anxiety.

1186S

Genetic determinants of survival in patients with Alzheimer's disease. *X. Wang[1, 2], OS. Lopez[3, 4], RA. Sweet[3, 4, 5], JT. Becker[3, 4], MM. Barmada[1], E. Feingold[1, 2], FY. Demirci[1], IM. Kamboh[1, 4, 5].* 1) HUMAN GENETICS, University of Pittsburgh, PITTSBURGH, PA; 2) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA; 3) Department of Neurology, University of Pittsburgh, Pittsburgh, PA; 4) Alzheimer's Disease Research Center, University of Pittsburgh, Pittsburgh, PA; 5) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA.

There is a strong genetic basis for late-onset of Alzheimer's disease (LOAD) and thus far over 20 genes/loci have been identified that affect the risk of LOAD. In addition to disease risk, genetic variation at these loci may also affect components of the natural history of AD, such as survival in AD. However, the role of these known loci in survival of AD patients has not been explored extensively. In the present study, we examined the role of 22 known LOAD genes with the time to death in 983 AD patients recruited from a referral clinic. In a secondary analysis, genome-wide association was examined to identify novel loci that may influence survival in AD. Survival analysis was conducted using Cox proportional hazards regression under an additive genetic model adjusted for the baseline MMSE score, education, gender, psychosis and top four principal components. We found multiple nominally significant associations (P<0.05) either within or adjacent to 10 LOAD genes and 3 of them remained significant in gene-based analyses (*BIN1*, P=4.79E-04; *INPP5D*, P=4.49E-02 and *APOE*, P=3.23E-02). Genome-wide association analysis revealed 8 suggestive novel loci at P<1E-05 (*ALDH4A1*, P=3.24E-06; *IL19*, P=6.62E-07; *NCKAP5*, P=1.37E-06; *CCDC85C*, P= 2.25E-06; *NARS2*, P=3.41E-06; *PKNOX2*, P=8.03E-06; *SDR9C7*, P=2.42E-06; and *SALL4*, P=2.25E-06). These results indicate that in addition to some known LOAD genes, genetic variation in novel loci may also affect survival of AD patients. Additional large follow-up studies in independent samples are required to confirm our potentially novel candidate loci findings.

1187M

Integrative systems approaches to deciphering the genetic landscape of late-onset Alzheimer's disease. *Y. Zhao[1], V. Mäkinen[2], Q. Meng[1], X. Yang[1].* 1) Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, CA 90095; 2) EMBL Australia, South Australian Health and Medical Research Institute, Adelaide Australia 5000.

Late-onset Alzheimer's disease (LOAD) is a complex neurodegenerative disorder, with the genetic components estimated to account for 60-80% of the disease variability. Recent genome-wide association studies (GWAS) implicated a number of susceptibility loci for LOAD. However, the identified susceptibility loci are substantially inconsistent across GWAS and the underlying genetic mechanisms are still largely unknown. In this study, we utilized an integrative systems approach that leveraged a multitude of genetic and genomic datasets, including 1) LOAD GWAS from the Alzheimer Disease Genetics Consortium (ADGC) from nine cohorts), 2) expression quantitative trait loci (eQTLs) from genetics of gene expression studies of human tissues related to LOAD (such as brain, blood, liver, and adipose tissue), 3) knowledge-driven biological pathways, 4) brain transcriptomic profiles of LOAD, and 5) data-driven gene regulatory networks from multiple human and mouse cohorts. The integration of these diverse data sources enabled tissue-specific investigations on whether the genetic variants associated with LOAD concentrated on gene subnetworks (i.e., specific parts of gene regulatory networks) and whether novel key regulators in the subnetworks could be identified based on data-driven network structures. Our results confirmed the involvement of many well-known LOAD-related processes, such as lipid metabolism, immune and inflammatory response, endocytosis/intracellular trafficking, and cell migration to be significantly enriched for LOAD risk variants across the nine LOAD GWAS datasets. More interestingly, we detected several novel subnetworks related to mitochondrial protein import, viral infection, and pyrimidine metabolism, which show significant expression changes between LOAD patients and nondemented subjects. Finally, by utilizing the gene-gene relationships revealed by the network architecture, we detected key regulator genes, both known (e.g. NCKAP1L and FYB) and novel (e.g. VAV1, and PTPN6), in these LOAD subnetworks. Our results shed lights on the complex mechanisms underlying LOAD and highlight potential novel targets for the highly debilitating disease.

1188T

***TMEM106B* is a genetic modifier of frontotemporal lobar degeneration with *C9ORF72* hexanucleotide repeat expansions.** *M.D. Gallagher[1,2], E. Suh[3], M. Grossman[2], L. Elman[2], L. McCluskey[2], J.Q. Trojanowski[3], V.M. Lee[3], V.M. Van Deerlin[3], A.S. Chen-Plotkin[2], International Collaboration for FTLD.* 1) Cell & Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, PA; 2) Department of Neurology, Perelman School of Medicine, University of Pennsylvania, PA; 3) Department of Pathology & Laboratory Medicine, Center for Neurodegenerative Research, Perelman School of Medicine, University of Pennsylvania, PA.

Hexanucleotide repeat expansions in chromosome 9 open reading frame 72 (*C9ORF72*) have recently been linked to frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS), and may be the most common genetic cause of both neurodegenerative diseases. Genetic variants at *TMEM106B* influence risk for the most common neuropathological subtype of FTLD, characterized by inclusions of TAR DNA binding protein of 43kDa (FTLD-TDP). Previous reports have shown that *TMEM106B* is a genetic modifier of FTLD-TDP caused by mutations in the progranulin gene (*GRN*), with the major (risk) allele of rs1990622 associating with earlier age at onset of disease. Here we report that rs1990622 genotype affects age at death in a single-site discovery cohort of FTLD patients with *C9ORF72* expansions (n=14), with the minor allele correlated with earlier age at death (p=0.024). We replicate this modifier effect in a 30-site international neuropathological cohort of FTLD-TDP patients with *C9ORF72* expansions (n=75), again finding that the minor allele associates with earlier age at death (p=0.016), as well as earlier age at onset (p=0.019). In contrast, *TMEM106B* genotype does not affect age at onset or death in 241 FTLD-TDP cases negative for *GRN* mutations or *C9ORF72* expansions. Thus, *TMEM106B* is a genetic modifier of FTLD with *C9ORF72* expansions. Intriguingly, the genotype that confers decreased risk for developing FTLD-TDP and delayed age of disease onset in *GRN*-associated FTLD (minor, or C, allele of rs1990622) is associated with earlier age at onset and death in *C9ORF72* expansion carriers, providing an example of sign epistasis in human neurodegenerative disease.

1189S

Gene subnetworks in cocaine-induced paranoia: Convergence between populations. *C. Phokaew[1], H. Kranzler[2], L. Farrer[3], H. Zhao[4], J. Gelernter[1,5].* 1) Department of Psychiatry, Yale University School of Medicine, New Haven, CT and VA Connecticut Healthcare, West Haven, CT, USA; 2) Center for Studies of Addiction, Department of Psychiatry, Perelman School of Medicine of the University of Pennsylvania and VISN4 MIRECC, Philadelphia VAMC, Philadelphia, PA, USA; 3) Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, USA; 4) Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA; 5) Departments of Genetics and Neurobiology, Yale University School of Medicine, New Haven, CT, USA.

Cocaine induce paranoia (CIP) is a common consequence of cocaine use that can cause serious morbidity. We used pathway analysis to identify target genes to aid in our understanding of CIP pathophysiology, based on the principle of identifying reproducible GWAS-derived gene lists from two independent datasets. GWAS data set was generated by using the Illumina Omni1-Quad microarray, as described in our published cocaine dependence (CD) GWAS. The two populations in this study were African American (AA; 1273 cases and 824 controls, 52% male) and European American (EA; 951 cases and 749 controls, 58.9% male). All sample subjects were interviewed using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA). The CIP phenotype was ascertained via the question : "Have you ever had a paranoid experience when you were using cocaine?" The control group included only subjects who reported lifetime cocaine use. Gene sub-networks were obtained by using the dmGWAS R package on GWAS data and the Reactomes database. P-values from the GWAS (briefly, logistic regression on sex, age, and the first three PCs - Gelernter et al, 2013) was combined with gene-gene interaction data to identify significant gene-gene interaction modules. The AA and EA lists were evaluated for overlapping gene content (which could result from different SNPs In the two populations), and genes the appeared in both group networks was tested by SKAT-O. Gene lists were submitted to the Database for Annotation, Visualization and Integrated Discovery (DAVID) for check for enrichment pathways according to KEGG (Kyoto Encyclopedia of Genes and Genomes) gene interaction database. The pathway analysis considered a total of 22061 genes. From these, four genes out of the 40 identified in AAs and the 88 In EAs were identical between the two populations: GABBR2 (AA p = 0.00056, EA p = 0.0076), ADCY8 (AA p = 0.037, EA p = 0.0099), DLC1 (AA p = 0.0015 , EA p = 0.0065) and ABLIM2 (AA p = 0.0023, EA p = 0.00013). Gene lists from both populations were enriched in several KEGG pathways, including Focal Adhesion (Padj 1.20E-09), ECM-Receptor Interaction (Padj 6.30E-07), etcs Separately, the AA group was most enriched in the Neuroactive Ligand-Receptor Interaction Pathway (Padj 1.40E-03), while the EA group was most enriched in Focal Adhesion Pathway (Padj 2.40E-07). Interaction pathways including GABBR2, ADCY8, DLC1 and ABLIM2 appear the most promising for CIP risk, based on our dataset.

1190M

A genome-wide screen for fear of heights susceptibility loci in a Finnish isolate. *I. Hovatta[1,2], Z. Misiewicz[1], T. Hiekkalinna[3,4], T. Paunio[3,5], T. Varilo[6].* 1) Department of Biosciences, University of Helsinki, Helsinki, Finland; 2) Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland; 3) Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki, Finland; 4) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 5) Department of Psychiatry, University of Helsinki, Helsinki, Finland; 6) Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki, Finland.

Acrophobia (fear of heights) is an anxiety disorder characterized by disproportional or irrational feeling of apprehension cued by the occurrence or anticipation of elevated spaces. It is classified as a specific phobia of naturalistic type under the DSM-V. No specific genetic variants predisposing to acrophobia have been identified. Our goal was to map genes predisposing to acrophobia. We genotyped and analyzed 575 microsatellite markers across the genome in the Finnish schizophrenia cohort, in which acrophobia segregates independently of schizophrenia. The sample constitute mostly of large multigenerational pedigrees from an internal genetic isolate comprising 57 nuclear families including 643 people, 105 of them affected with acrophobia. We first undertook parametric and non-parametric linkage analysis using statistical software packages FASTLINK 4.1P and MERLIN. As a follow-up, we are performing joint linkage and association analyses with PSEUDOMARKER to fine-map the identified loci and to find putative founder alleles predisposing to acrophobia. Finally, we will perform haplotype analysis using MERLIN to identify potential risk haplotype(s) shared among affected family members. Our results from single two-point linkage analysis revealed two suggestive loci with recessive model: 13q34 (Zmax= 2.88) and 5q32-q34 (Zmax=2.15). Variants in D-amino acid oxidase activator (DAOA) gene located at the 13q34 locus were previously associated with schizophrenia. Consistently, we found linkage in individuals affected both with schizophrenia and acrophobia and not those affected only by the latter. However, the 5q32-34 seems specific for acrophobia as omitting individuals with comorbid schizophrenia did not significantly affect the LOD score. The singlepoint non-parametric linkage analysis showed the strongest linkage for region 5q32-q34 [(NPL (all) LOD score 1.51)]. Our study revealed several suggestive loci segregating independently in schizophrenia and acrophobia. The concordance between results from parametric and non-parametric linkage analysis for region 5q32-q34 with individuals only affected with acrophobia indicates this region as potentially harboring acrophobia-associated gene(s). Our results will require follow-up analyses to fine-map the regions of interest. Identification of risk variants would provide novel insight into the genetic basis of the fear of heights phenotype and contribute to the molecular-level understanding of its etiology.

1191T

Ancient human mtDNA variation is associated with Autism spectrum disorder in Europeans. *D. Chalkia[1], O. Derbeneva[1], M. Lvova[1], A. Lakatos[2], J. Leipzig[3], D. Hadley[4], H. Hakonarson[4], D. Wallace[1,5].* 1) Center for Mitochondrial & Epigenomic Medicine, Children's Hospital of Philadelphia Research Institute, Philadelphia, PA; 2) Department of Psychiatry and Human Behavior, University of California at Irvine, Irvine, CA; 3) Center for Biomedical Informatics, Children's Hospital of Philadelphia Research Institute, Philadelphia, PA; 4) Center for Applied Genomics, Division of Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 5) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA.

Autism spectrum disorders (ASD) are thought to result from genetic defects in synaptic biogenesis, structure, and maintenance. GWAS and CNV studies in autism have led to the identification of a number of chromosomal loci involved in the structure of the synapse suggesting a synaptic assembly pathway. However, when all of the synaptic loci are added together they still account for only a small proportion of ASD cases. Therefore, additional pathways must be relevant in the genetics of ASD. To explain the complex genetics of ASD, alternative ASD pathways would need to encompass a large number of genes, affect a quantitative trait, and be relevant to the synapse. Mitochondrial bioenergetics fulfills these criteria. Therefore, we are testing the hypothesis that a significant proportion of the risk for developing ASD is the result of inheriting partial defects in genes for mitochondrial bioenergetics. To test our hypothesis, first we determined if ancient human mtDNA variation is associated with ASD risk. To accomplish this, we used two independently generated and genotyped cohorts, the Autism Genetic Resource Exchange (AGRE) and the CHOP Autism Case-Control (ACC). These pre-existing mtDNA single nucleotide polymorphism (SNP) data were generated by Illumina chip analysis. We deduced the mtDNA haplogroups for all the AGRE and ACC samples by first converting the Yoruban mtDNA SNPs included in these arrays to their corresponding positions on the rCRS and then using Mitomaster to assign an mtDNA haplogroup to each sample. To limit the number of variables and thus increase the statistical power of the study, we only utilized cases and controls that harbored European haplogroups. Generalized linear modeling analysis of the AGRE cohort suggests that J mitochondrial lineage and T2 haplogroup are risk factors for ASD (J lineage: OR = 2.22, CI=1.21-4.05, p-value =0.0094; T2 haplogroup: OR= 1.79, CI= 1.26-2.55, p-value=0.0012). Congruently, in the ACC cohort we found that haplogroup T2 is strongly associated with ASD and represents a risk factor (OR=1.45, CI=1.04-2.03, p-value-=0.0284). These data demonstrate that ASD risk is modified by mtDNA haplogroup thus supporting our mitochondrial bioenergetic hypothesis of ASD.

1192S

Genome-wide association study of dementia with Lewy bodies. *J. Bras[1], C.M. Molony[2], A.B. Singleton[3], D.J. Stone[4], International DLB Consortium.* 1) Dept Molecular Neuroscience, University College of London, London, United Kingdom; 2) Merck Research Laboratories, Department of Genetics & Pharmacogenomics (GpGx), Boston, MA; 3) Laboratory of Neurogenetics, National Institutes on Aging, NIH, Bethesda, MD; 4) Merck Research Laboratories, Department of Genetics & Pharmacogenomics (GpGx), Phila, PA.

Dementia with Lewy Bodies (DLB [MIM 127750]) is a complex brain disorder that has been severely underserved and whose etiology remains unclear. A growing body of evidence supports the notion that Alzheimer's disease (AD), DLB, and Parkinson's disease (PD) are members of the same disease continuum. Pure DLB appears to exist in only 20-30% of clinical DLB cases, with the remainder showing some level of AD pathology. Similarly, as many as 20% of AD patients show comorbid DLB pathology. Genetic studies in AD have focused on the "pure" form of the disease when possible, excluding patients showing DLB/ α-synuclein (SNCA) inclusions. Also, drug-development efforts for AD disease modification have almost exclusively focused on the amyloid and tau pathologies. Taken together, these raise a critical concern that the current compounds in clinical trials for AD disease modification, even if successful, may not have full efficacy in up to 20% of patients. Understanding of the genetics of DLB and mixed AD/DLB is therefore critical for identification of both drug targets and pharmacogenetic/compound response predictors. Here we present results from a large genome-wide association study conducted in DLB. We have included only cases that have been neuropathologically diagnosed with DLB, according to the latest criteria greatly reducing the likelihood of contamination of other forms of dementia. Our cohort comprises 1,300 cases and 2500 population controls. We have performed genotyping using Illumina's HumanOmni 2.5M GWAS array plus exome content. Our results confirm that APOE ε4 is a strong genetic risk factor for DLB, and also suggest the involvement of other genes in this disease, opening novel research and therapeutic avenues into this disorder. Future plans for this cohort include whole-exome sequencing of cases.

1193M

The contribution of uncommon coding variants to risk for Alzheimer's disease, frontotemporal dementia, and progressive supranuclear palsy: an exome array study of the multi-ethnic GIFT cohort. *J.A. Chen[1], Q. Wang[1], J. Davis-Turak[1], Y. Li[1], D. Chatzopoulou[1], H. Chui[2], C. Cotman[3], C. DeCarli[4], T. Foroud[5], A. Huang[1], A. Karydas[6], E. Klein[1], W. Kukull[7], J. Lee[1], A. Levey[8], M. Mendez[9], J. Miller[10,11], W. Poon[3], J. Ringman[9], A. Rosen[8], R. Sears[1], J. Shapira[9], A. Varpetian[2], K. Wojta[1], B.L. Miller[6], D.H. Geschwind[1,9], G. Coppola[1,9].* 1) Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA; 2) Department of Neurology, University of Southern California, Los Angeles, CA; 3) Institute for Memory Impairments and Neurological Disorders, University of California, Irvine, CA; 4) Department of Neurology, University of California, Davis, CA; 5) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; 6) Memory and Aging Center, University of California, San Francisco, CA; 7) National Alzheimer's Coordinating Center, University of Washington, Seattle, WA; 8) Department of Neurology, Emory University, Atlanta, GA; 9) Department of Neurology, University of California, Los Angeles, CA; 10) Department of Nutritional Sciences, Rutgers University, New Brunswick, NJ; 11) Department of Pathology and Laboratory Medicine, University of California, Davis, CA.

Genetic contributions to neurodegenerative diseases such as Alzheimer's disease, frontotemporal dementia, and progressive supranuclear palsy have become increasingly well characterized. However, few studies have examined the effect of low-frequency coding variants on a genome-wide level. Here, we used the Illumina HumanExome BeadChip array to genotype a large number of exonic variants and other polymorphisms in a cohort of patients with dementia (224 with Alzheimer's disease, 168 with frontotemporal dementia, and 48 with progressive supranuclear palsy), and 224 non-demented controls from the Genetic Investigation in FrontoTemporal dementia and Alzheimer's disease (GIFT). An additional multi-ethnic replication cohort of 240 Alzheimer's disease patients and 240 controls was used to validate suggestive findings. Association testing, on both the variant level and the gene level, was performed. No novel loci were detected with genome-wide significance in these tests, likely due to the limited power to detect an effect from rare variants of modest effect size. However, using a gene-wise association test (SKAT), we replicated an association with the known Alzheimer's disease gene ABCA7 in several ethnicities, hinting that exonic variants within this gene modify disease risk. Additionally, two suggestive candidate genes were identified in Alzheimer's disease patients. Corroborating evidence from other exome array studies and gene expression data points toward potential involvement in Alzheimer's disease pathogenesis.

1194T

A variant in *Cadherin 1* (*CDH1*) achieves near genome-wide significance in African Americans using a liability model. *J. Mez[1], J. Chung[2], K.L. Lunetta[3], J. Haines[4], R.P. Mayeux[5,6], M.A. Pericak-Vance[7,8], G.D. Schellenberg[9], L.A. Farrer[1,2,3,10,11], Alzheimer's Disease Genetic Consortium.* 1) BU Alzheimer's Disease Center (Neurology), Boston University School of Medicine, Boston, MA; 2) Biomedical Genetics (Medicine), Boston University School of Medicine, Boston, MA; 3) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 4) Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH; 5) Taub Institute on Alzheimer's Disease and the Aging Brain (Neurology), Department of Neurology, Columbia University, NY; 6) Gertrude H. Sergievsky Center (Neurology), Columbia University, New York, NY; 7) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 8) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 9) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 10) Department of Ophthalmology, Boston University School of Medicine, Boston, MA; 11) Department of Epidemiology, Boston University School of Public Health, Boston, MA.

Late onset Alzheimer's disease (LOAD) risk is influenced by multiple known genetic, clinical and environmental factors. Including these factors as covariates in traditional genetic analyses of subjects ascertained by phenotype (case-control ascertainment) may decrease power to detect an association. However, a liability model can be used to extract risk information from known genetic and non-genetic factors without a reduction in power. A recent genome wide association study (GWAS) by the Alzheimer's Disease Genetic Consortium (ADGC) found that, in addition to the *APOEε4* allele, a variant in the *ABCA7* gene (rs115550680) was significantly associated with LOAD at the genome-wide level in African Americans. In the current study, we conducted a GWAS in African Americans, employing a liability model that included these genetic risk factors as well as age and sex. Subjects included 1,910 well-characterized African American LOAD cases and 3,829 cognitively normal African American controls from 9 datasets from the ADGC. First, we used logistic regression (or logistic generalized estimating equations (GEE) for family-based datasets) to derive a liability score (Pearson residual) for each dataset to serve as a quantitative continuous phenotype for GWAS. The model included *APOEε4* status, rs115550680 (either genotyped or imputed weight), age and sex. Then, we evaluated association of the liability score with a genome-wide set of six million imputed markers using linear regression (or linear GEE for family-based datasets) that included principal components of population substructure. Results were meta-analyzed across datasets. We obtained near genome-wide evidence of association ($p=2.21\times10^{-7}$) between the liability score and a SNP in *CDH1*, a gene which encodes a calcium dependent cell-cell adhesion glycoprotein. In the ADGC's previous African American GWAS of LOAD that used a traditional case-control model, the same *CDH1* SNP reached a p-value of 1.33×10^{-5}, two orders of magnitude larger than using the liability model. The current study suggests a liability model can improve association signals in GWAS. We plan to add more clinical risk factors to our model with the hope of further increasing our power to detect an association.

1195S

A genome wide association study on fine motor speed. *C.L. Satizabal[1,2], J.A. Smith[3], J.C. Bis[4], L. Yu[5], A. Beiser[1,6], W. Zhao[3], J.I. Rotter[7], A.S. Buchman[5,8], R. Au[1,2], S.T. Turner[9], W.T. Longstreth[10], S.L.R. Kardia[3], B.M. Psaty[11], D.A. Bennett[5,8], A.L. Fitzpatrick[11], S. Seshadri[1,2], T.H. Mosley[12], Neuro-CHARGE consortium.* 1) Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 2) Framingham Heart Study, Boston, MA, USA; 3) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA; 4) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA; 5) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA; 6) Boston University Schools of Medicine and Public Health, Boston, MA, USA; 7) Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; 8) Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA; 9) Division of Nephrology and Hypertension, Department of Internal Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA; 10) Departments of Neurology and Epidemiology, University of Washington, Seattle, WA, USA; 11) Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA; 12) Department of Medicine-Geriatrics, University of Mississippi Medical Center, Jackson, MS, USA.

Background: The Finger Tapping Test (FTT) is one of the most extensively used tasks to evaluate fine motor speed, and lower scores are often indicative of brain damage. However, genetic variants associated with fine motor functions have not been investigated in large community samples. The aim of this study is to relate genome-wide genetic variation to fine motor speed in a large sample of middle-aged and elderly individuals using FTT scores. Methods: FTT scores were defined as the average number of taps for the right and left hands over 10 or 15 seconds. We conducted two separate meta-analyses of genome-wide association studies (GWAS) of FTT scores in 7,580 individuals of European and 1,405 of African-American descent from five population-based cohorts of the Neuro-CHARGE consortium. Genotype imputation was based on the 1,000 Genomes reference panel. Results across cohorts were combined with pooled inverse variance-weighted meta-analysis methods using METAL and applying genomic control. We considered p-values of $< 5\times10^{-5}$ as findings of interest. Results: In the European sample, we identified suggestive intronic variants in ASTN1 (rs78553813, $p=1.2\times10^{-6}$); RREB1 (rs9379083, $p=1.3\times10^{-6}$); METTL21C (rs117561181, $p=1.5\times10^{-6}$); UBTD2 (rs116792690, $p=1.9\times10^{-6}$) and ALS2 (rs3731702, $p=2.2\times10^{-6}$). These genes have been associated with glial-guided neuronal migration during brain development, cell differentiation, bone-muscle functions and juvenile lateral sclerosis. Several other suggestive extragenic variants were found on chromosome 6p24.3. As for the African-American sample, we found several suggestive intronic variants encompassing the EFHD1 gene (between $1.6\times10^{-6} < p < 3.6\times10^{-6}$ for the top hits), which expression is augmented during neuronal differentiation; as well as other extragenic variants near LINC00276 (rs1878107, 2.6×10^{-7}) and TIGD2 (rs6853480, $p=2.7\times10^{-7}$). Conclusions: Although our results did not reach genome-wide significance, we found several novel suggestive variants associated with fine motor speed in European and African-American samples. We plan to further explore these findings through increased sample sizes and trans-ethic meta-analyses, which could provide novel insights into the biology of fine motor speed.

1196M

Variants near CCK receptors are associated with electrophysiological responses to pre-pulse startle stimuli in a Mexican American cohort. *T.M. Norden-Krichmar[1], I.R. Gizer[2], K.C. Wilhelmsen[3], N.J. Schork[4], C.L. Ehlers[1].* 1) Department of Molecular and Integrative Neurosciences, The Scripps Research Institute, La Jolla, CA; 2) Department of Psychological Sciences, University of Missouri, Columbia, MO; 3) Department of Genetics and Neurology, University of North Carolina, Chapel Hill, NC; 4) J. Craig Venter Institute, La Jolla, CA.

Neurophysiological measurements of the response to pre-pulse and startle stimuli have been used to study the mechanisms for a wide range of psychiatric disorders. Some Mexican American (MA) populations have a higher risk for the development of alcoholism. We have previously shown that the electrophysiological response to pre-pulse/startle responses is associated with alcohol dependence in a young adult MA population. That is, the participants diagnosed with alcohol dependence demonstrated significantly less inhibition and more facilitation to startle responses following pre-pulse stimuli. In the present study we investigated whether genetic factors associated with this phenomena could be identified. The study included 420 young adult (age 18 - 30 years) MA men (n=170) and women (n=250). Information on alcoholism and other psychiatric disorders were obtained through an interview using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA). Acoustic pre-pulse and startle stimuli were presented binaurally through headphones. Electrodes were used to record the event-related potential (ERP) response in frontal cortex to the stimuli. For all subjects, DNA was extracted from blood samples, followed by genotyping using an Affymetrix Axiom Exome1A chip. An exome-wide association analysis of the genotyping data revealed that the *CCKAR* and *CCKBR* (cholecystokinin A and B receptor) genes each had a nearby variant that showed suggestive significance ($p<=E-05$) in association with the pre-pulse response values. The rs2171755 variant (NC_000004.12:g.26502338T>C) is located 12kb upstream from the *CCKAR* gene in a MIRb class SINE repetitive element. The rs58905541 variant (NC_000011.10:g.6296157C>T) is located 24kb downstream from the *CCKBR* gene in a DNaseI hypersensitivity cluster. Significance of these variants was further validated by permutations and inclusion of the covariates age and gender. The neurotransmitter cholecystokinin (*CCK*), along with its two receptors, *CCKAR* and *CCKBR*, have been previously associated with anxiety and panic disorders, schizophrenia, and alcoholism. Additionally, the administration of CCK-related peptides has been found to influence the startle response in rats and humans. These results suggest that variants near *CCKAR* and *CCKBR* may play a role in the pre-pulse/startle response, and possibly in alcohol dependence and other behavioral disorders, in this Mexican American cohort. (Supported by AA06420, DA030976).

1197T

Large-scale genetic predictor of gene expression associated with risk of bipolar disorder. *K.P. Shah, E.R. Gamazon, N.J. Cox, H.K. Im.* Department of Medicine, University of Chicago, Chicago, IL.

Genome-wide association (GWA) studies have successfully identified thousands of loci associated with complex diseases, many of which have also been associated with gene expression phenotypes across a variety of tissues, known as expression quantitative trait loci (eQTLs). The enrichment of eQTLs in GWA studies suggests that altered gene expression plays a role in risk for a number of complex diseases. Recognizing the important role of gene expression in risk for complex diseases, we developed and applied a novel approach, PrediXcan, to predict gene expression traits in GWA studies and test these predicted gene expression traits for association with disease. We developed large-scale predictors of gene expression for all genes using gene expression data from the whole blood of control individuals as part of the GTEx Pilot Project. We then tested each of these predicted gene expression levels for association with bipolar disorder (BPD) using GWA data from the GAIN sample. We found that predicted levels of *RFNG* are associated with risk of BPD ($p\text{-value} = 2.59\times10^{-6}$) even after correcting for multiple testing ($p\text{-value}_{Bonferroni.corrected} = 4.19\times10^{-2}$). Higher predicted *RFNG* expression is associated with lower risk of BPD in this study. Our *RFNG* result replicated in the WTCCC GWA study as well (p-value = 0.00017). *RFNG*, RFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase, is part of the Notch signaling pathway and involved in neurogenesis. Previous studies have highlighted the importance of Notch signaling in psychiatric disease through GWA studies and linkage studies. Specifically, *NOTCH4* has been associated with risk of schizophrenia and mutations in *NOTCH3* have been seen in patients with CADASIL, a rare autosomal dominant disease characterized by cortical white matter death, stroke, mood disorder, migraines, and dementia. Our results suggest that *RFNG* gene expression may be relevant to the development of BPD further highlighting the importance of the Notch signaling pathway in neuropsychiatric disease risk. This study also shows the potential of our PrediXcan approach for identifying genes relevant to disease phenotypes.

1198S

eQTL analysis of a large-scale RNA-sequencing cohort of schizophrenic and normal brains. *S.K. Sieberts for the Common Mind Consortium, Swedish Schizophrenia Consortium, Schizophrenia Working Group of PGC.* Sage Bionetworks, Seattle, WA.

The most recent schizophrenia GWAS reported >100 independent associated loci, implying the disorder is highly polygenic. To better understand the pathology of neuropsychiatric disease, we formed the CommonMind Consortium (commonmind.org) to generate large-scale data (RNA-seq, ChIP-seq, DNA-seq/genotyping) from human post-mortem brain samples. Here, we focus on eQTL using RNA-seq (Illumina HiSeq2000 - paired end reads) of the dorsolateral prefrontal cortex (BA9/46) in 265 schizophrenia cases and 289 controls. Effects of known clinical (gender, age of death, medications) and technical (brain bank, post-mortem interval, RNA quality, sequencing batch) covariates, as well as hidden confounders, were removed using surrogate variable analysis (SVA). Data were normalized via voom and a linear model was applied to detect eSNPs, adjusting for genetic structure. Genotypes were assayed on the Illumina Infinium HumanOmniExpressExome8 chip and were imputed to 1000 Genomes. Preliminary differential expression analysis using linear models implemented in voom/limma identified 15.6% of all expressed genes as differentially expressed between cases and controls (FDR 5%). These genes were enriched for DNA variants associated with schizophrenia, including rare (allele frequency < 0.1%) nonsynonymous variants in Swedish case-control exome sequencing (p=0.012) and common GWAS variants (p=0.045). Preliminary eQTL analysis of the assayed genotypes (~620,000 with MAF≥0.05) identified 102,461 and 85,015 proximal eSNPs (distance < 1Mb) at an FDR of 5% in controls and patients with schizophrenia, respectively; representing 53.1% and 44.3% of expressed genes. 51,823 eSNPs were common to both groups, and 24.8% of expressed genes have at least one eSNP in common. Preliminary results show that 54 of the 108 GWAS regions from the Psychiatric Genomic Consortium contain eSNPs. Complete analysis using imputed SNPs will examine condition dependent eQTL, incorporating epigenetic information, and identify likely gene candidates in regions indicated by GWAS in hopes of contributing to the understanding of the biological mechanisms contributing to disease. This large dataset will be made public in early 2015, including a catalogue of brain-expressed genes and their isoforms in cases and controls, which will contribute to empowering the research community to make novel discoveries relating neurobiology disease risk and mechanisms in hopes of advancing novel diagnostic tools and therapies.

1199M

A Spatiotemporal Systems Biology Approach to Understanding Autism Spectrum Disorder and Schizophrenia. *A.J. Willsey[1], S.J. Sanders[1], M. Li[2], S. Dong[3], A.T. Tebbenkamp[2], R.A. Muhle[2], S.K. Reilly[2], L. Lin[2], S. Fertuzinhos[2], J.A. Miller[4], C. Bichsel[2], W. Niu[2], J. Cotney[2], X. He[5], E. Hoffman[2], L. Klei[6], J. Lei[5], W. Liu[2], L. Liu[5], C. Lu[5], Y. Zhu[2], E.S. Lein[4], L. Wei[7], J.P. Noonan[2], K. Roeder[5], B. Devlin[6], N. Sestan[2], M.W. State[1].* 1) Psychiatry, University of California, San Francisco, San Francisco, CA; 2) Yale University School of Medicine, New Haven, CT; 3) Peking University, Beijing, China; 4) Allen Institute for Brain Science, Seattle, WA; 5) Carnegie Mellon University, Pittsburgh, PA; 6) University of Pittsburgh, Pittsburgh, PA; 7) National Institute of Biological Sciences, Beijing, China.

Recent advances in genome- and exome-wide methods, including SNP genotyping arrays and high-throughput sequencing, have facilitated identification of genes associated with autism spectrum disorder (ASD) and schizophrenia (SCZ). This ever-expanding list of genes offers an unprecedented opportunity to elucidate the biology of ASD and SCZ. However, the large number of genes involved coupled with the many biological functions of each gene confound interpretation. Working on the hypothesis that the many genes involved in each disorder must disrupt a much smaller number of shared biological processes to result in a common phenotype, we previously identified the midfetal prefrontal cortex and primary motor-somatosensory cortex as a point of convergence in the expression of ASD genes in the human brain. Since this initial analysis, the number of ASD-associated genes has increased more than three-fold, and additional loci continue to be implicated in SCZ. By applying this methodology in a cross disorder analysis of ASD and SCZ we aim to identify novel components of the neuropathology and to determine whether these are shared or distinct between the disorders. Using a rich expression dataset encompassing multiple human brain regions across the timespan of human brain development, we construct co-expression networks around the most current set of 'high-confidence' ASD genes and then interrogate spatial and temporal 'windows' of brain development for enrichment of additional ASD genetic risk. We then identify genes with risk shared across both ASD and SCZ and construct co-expression networks around these genes in order to understand the similarities and differences between the spatial and temporal characteristics of these disorders. Our preliminary findings suggest that we will: 1) strengthen evidence for the spatiotemporal points previously identified as critical nexuses in ASD risk; 2) identify additional nexus points of risk in ASD; and 3) increase our understanding of the similarities and differences between ASD and SCZ, such as the specific brain regions and developmental epochs central to onset of these disorders. The work presented here represents a novel paradigm for progressing from genes to etiology by placing genetic findings in the relevant spatial, temporal, and cellular context. Moreover, identifying key differences between ASD and SCZ is crucial to improving our understanding of these disorders.

1200T

Multiple system atrophy and spinocerebellar degeneration associated with mutations in the COQ2 gene. *H. Sakamoto[1], M. Hirano[1,2], Y. Nakamkura[1], K. Saigoh[2], S. Ueno[1,2], Y. Mitsui[2], S. Kusunoki[2].* 1) Department of Neurology, Sakai Hospital Kinki University Faculty of Medicine, ≤, Osaka, Japan; 2) Department of Neurology,Kinki University Faculty of Medicine, Osakasayama, Osaka, Japan.

Multiple system atrophy (MSA) is a rapidly progressive disorder with severe motor disability within a few years, and is characterized by cerebellar ataxia, parkinsonism, and autonomic disturbances. This disease is clinically classified into two subgroups: MSA with predominant parkinsonism (MSA-P) and MSA with predominant cerebellar ataxia (MSA-C). The cause of MSA is largely unknown, but recently, mutations or sequence variations in the COQ2 gene encoding a protein for coenzyme Q10 synthesis have been identified in patients with MSA. Variations were more frequent in patients with MSA-C. In contrast, no report has described effects of COQ2 variations on clinical phenotypes in other neurodegenerative diseases, including autosomal dominant-spinocerebellar degeneration (SCD), Parkinson's disease (PD), and progressive supranuclear palsy (PSP). In this study, we sequenced the COQ2 gene in 31 Japanese patients with MSA, 20 with known causes of ataxia, 23 with PD, and 11 with PSP. We found the reported variation (p.V343A) in the COQ2 gene in five MSA (MSA-C, 3 and MSA-P, 2). The positive ratio in MSA (5/31, 16%) was larger than in controls (7/200, 3.5%, $p<0.05$), which is consistent with the reported finding. Additionally, we found this reported variation in a patient with SCA8, who had rare symptoms for Japanese patients with SCA8: parkinsonism and sleep apnea syndrome in addition to cerebellar ataxia. No patients with PD and only one patient with PSP had this variation with typical features of the disease. Notably, we found a novel missense mutation in the COQ2 gene in a patient with SCA6. This patient had rare symptoms for SCA6: major urinary incontinence and increased deep tendon reflexes, symptoms common in MSA. Our result suggested that COQ2 variations may affect phenotypes of SCD.

1201S

AUTOSOMAL DOMINANT CEREBELLAR ATAXIA AND MENTAL IMPAIRMENT WITH A NOVEL NONSENSE MUTATION OF *PRKCG*. *H. Shimazaki, J. Honda, T. Naoi, M. Namekawa.* Division of Neurology, Dept. of Internal Medicine, Jichi Medical University, Shimotsuke, Tochigi, Japan.

Backgrounds: Autosomal dominant cerebellar ataxias (ADCA) comprise clinically and genetically heterogeneous neurodegenerative disorders characterized by progressive ataxia with other neurological system disturbances. Machado-Joseph disease (MJD) / Spinocerebellar ataxia type 3 (SCA3), SCA6, SCA31, dentatorubral-pallidoluysian atrophy (DRPLA), SCA2 and SCA1 are frequent ADCAs, but other types are rare in Japan. We encountered a family with two patients of cerebellar ataxia with mental impairment. We attempt to identify the causative gene mutation of this family with ADCA. Methods: We investigated the proband with neurological examination, brain MRI, SPECT and gene analyses. Results: The neurological examination revealed cerebellar ataxia and cognitive impairment, but tremor or myoclonus were not observed. Brain MRI demonstrated cerebellar vermian atrophy. Brain SPECT showed cerebellar hypoperfusion. We could not detect CAG repeat expansions of SCA1, 2,3,6,7,8,12,17,DRPLA genes, and sequencing analyses could not reveal pathologic substitutions in the 5' UTR of the puratrophin gene. Whole-exome sequencing (WES) could identify the novel heterozygous nonsense mutation in the PRKCG gene, which is the causative gene for SCA14. This heterozygous mutation was confirmed by Sanger sequencing, and found in another patient, co-segregated within the family members and not found in controls. Conclusion: We could identify PRKCG mutation in this ADCA family. SCA14 usually showed cerebellar ataxia with tremor or myoclonus. Mental impairment is rare in SCA14. To the best of our knowledge, PRKCG gene mutations were only missense mutations in the previous literatures. This family has atypical clinical features with a novel nonsense mutation in the PRKCG gene. We could consider this nonsense mutation of PRKCG gene is associated with atypical clinical symptoms compared with typical ones of SCA14.

1202M

SPG7 mutations in a French-Canadian family affected by a recessive spastic ataxia. *M. Tetreault[1,2], S. Yang[3], K. Choquet[2,3], K. Boycott[4], J. Majewski[1,2], B. Brais[2,3], C4R Consortium.* 1) McGill Univeristy and Genome Quebec Innovation Center, Montreal, QC, Canada; 2) Department of Human Genetics, McGill University, Montreal, QC, Canada; 3) Neurogenetics of motion laboratory, Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada; 4) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON, Canada.

The current implementation of whole exome (WES) and whole genome (WGS) sequencing, in research has greatly accelerated the identification of disease-causing genes in Mendelian disorders, and more importantly in diseases that were poorly diagnosable. The diagnosis of ataxias is difficult due to rarity, clinical and genetic heterogeneity, and still today a large number of patients remain without a molecular diagnostic. We recruited a French-Canadian (FC) family (10 individuals; 4 affected cases) with an adult onset autosomal recessive spastic ataxia with dysarthria. Due to the heterogeneity and the difficulty to identify a disease gene by relying only on clinical features, we decided to send one affected individual for sequencing. WES was performed using Agilent SureSelect capture kit and Illumina HiSeq. A list of rare homozygous and compound heterozygous variants (MAF <3%; in 1000genomes) was produce. Variants that were also homozygous or compound heterozygous in our in-house exome database were excluded. A total of 22 candidate genes remained in our list, among them SPG7, encoding a mitochondrial metalloprotease. Mutations in SPG7 are associated with an autosomal recessive spastic paraplegia (SPG7). Our FC patient is compound heterozygote for two missense variants (p.A510V and p.P750L). The variant p.A510V has already been reported in the literature as a pathogenic variant and seems to be the most frequent mutation in SPG7. This variant was observed at low frequency in 1000genomes and EVS (0.1%; and 0.3%;) but with a frequency of 1%; in our in-house database. The novel variant p.P750L has never been reported in any public or our in-house databases. Both variants are predicted to be pathogenic and are present at a conserved position. Clinical similarities between our patients and reported SPG7 cases, as well as the segregation of the variants with the phenotype confirms SPG7 as the causal gene in this family. The presence of a causing variant in 1%; of the population highlight the existence of "not so rare" variants causing recessive rare diseases and this should be taken into account when searching for genetic causes.

1203T

Discovery, validation and genotyping of CNVs by analysis of genome sequence and microarray. *D. Antaki[1], M. Gujral[1], W. Brandler[1], J. Michaelson[1], D. Malhotra[1], J. Estabillo[1], C. Corsello[4], J. Sebat[1,2,3].* 1) Institute for Genomic Medicine, La Jolla, CA; 2) Department of Psychiatry, UCSD, La Jolla, CA; 3) Department of Cellular and Molecular Medicine, UCSD, La Jolla, CA; 4) Autism Discovery Institute, Rady Children's Hosptial, San Diego, CA.

Advances in Next Generation Sequencing technology have enabled detection of an unprecedented array of structural genetic variants including Copy Number Variants (CNVs). CNVs confer risk for a variety of psychiatric and neurodevelopmental disorders, including Autism Spectrum Disorder (ASD). As we begin to apply whole genome sequencing approaches in our genetic studies of ASD, sensitive and accurate methods for CNV detection are needed. We applied a combination of DNA analysis techniques to the investigation of a cohort of 161 samples (28 ASD trios and 19 discordant sibling pair quad families), including deep 40X whole genome sequencing (325 bp libraries with 100 bp paired end reads), and two independent genotyping array platforms (Affy 6.0 and Illumina 2M). Genotyping was executed via a clustering algorithm, partitioning around medoids, resulting in significant noise reduction commonly found with SNP chip hybridization assays while distinctly segregating CNVs into clusters of heterozygous or homozygous duplications or deletions. Our robust multifaceted clustering method further enabled the detection of de novo CNVs. We demonstrate the advantage of a multidimensional approach in clustering and genotyping CNVs as opposed to the standard unidimensional approach.

1204S

Meta Analysis of Case/Control Autism Exome Sequencing Data. *J.A. Kosmicki[1,2], K. Roeder[3,4], B. Devlin[5], J.D. Buxbaum[6,7,8,9,10,11], M.J. Daly[1,2,12], Autism Sequencing Consortium.* 1) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114; 2) Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA, 02142; 3) 3Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; 4) Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; 5) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA; 6) Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 7) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 8) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 9) Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 10) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 11) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 12) Harvard Medical School, Boston, Massachusetts, USA.

Autism spectrum disorders (ASDs) currently affect 1 in 66 individuals and are highly heterogeneous in their phenotypic outcome ranging from severely affected individuals with speech impairment, intellectual disability, and behavioral problems to mildly affected individuals. Our understanding of the genetic architecture of ASD has drastically improved thanks in part to large-scale exome sequencing of individuals with autism focusing on rare loss-of-function (LoF) de novo mutations. Here we continue the work, calculating case/control association statistics on all variants, not just the LoF variants, from the exome sequences of 1300 cases and 3318 controls of European descent across three repositories collected by the Autism Sequencing Consortium (ASC). The case/control results were meta-analyzed using a modified version of weighted Z-scores with inverse variance to account for differential allele frequencies in rare variants across the three population samples. Single variant tests were computed using exact tests for the rarest variants (MAF < 4.33×10-3). Rare variant burden tests were used for low frequency (< 0.001%) LoF variants in a given gene to achieve sufficient statistical power to detect association.

Four coding variants falling into four separate genes (*C4BPA*, *BPTF*, *DFNB31*, and *STAG1*) and one gene, *MTMR2*, were identified as exome-wide significant by the single variant and rare variant burden tests respectively. Of the five genes, only *C4BPA* has never been associated with ASD. These and other top results were explored further using the transmission disequilibrium test (TDT) on 1241 trios collected by the ASC which have provided to date the primary discovery analysis through patterns of de novo variation (see abstracts by De Rubeis and He). Of the four exome-wide significant single variants, only *DFNB31* shows over transmission from the TDT. Going forward, combining results from case/control and trio association studies using a unified statistic may increase our power to detect novel genetic associations.

1205M

Interstitial duplication Xp11.4 and triplication of Yq11.22 leading to disruption of TSPAN 7 and NLGN4Y in a child with autism. *W.S. Baek.* Neurology, Parkside Medical Group, Upland, CA.

Introduction TSPAN 7(tetraspanin 7) is a gene of the tetraspanin superfamily located at Xp11.4, which is highly expressed in the prefrontal cortex. TSPAN 7 mutations have been reported to be associated with non-syndromic X-linked mild to moderate mental retardation. NLGN4Y (neuroligin 4, Y-linked) is a protein-coding gene located at Yq11.22, of which missense mutations leading to loss-of-function has been reported in autism. We report an unprecedented case with a combination of both Xp11.4 duplication and Yq11.22 triplication leading to partial duplication of the TSPAN 7 gene presenting as autism. Keywords: tetrspanin 7(TSPAN 7), neuroligin 4, Y-linked (NLGN4Y), copy number variant (CNV), autism Materials and Methods: Case Report A 5 year-old Hispanic boy presented with autism. Birth and family history were unremarkable. He had speech delay and behavioral problems. He was receiving special education and speech therapy. He was not dysmorphic. Results: SNP microarray analysis detected a 1.40MB interstitial duplication of Xp11.4 and a 5.67 MB interstitial triplication of Yq11.22, which was found to be a familial inheritance. Conclusions: The interstitial duplication of Xp11.4 led to disruption of the TSPAN 7 gene. As the duplication disrupts the gene, this may result in a dominant-negative effect. Inheritance of an X chromosome copy number variant (CNV) from a clinically normal mother may yield clinical consequences in male offspring due to differences in gene dosage. The interstitial triplication of the Yq11.22 led to disruption of the NLGN4Y (neuroligin 4) gene. Missense mutations of NLGN4Y leading to loss-of-function have been reported in autism. Although direct inheritance of a Y chromosome CNV from a phenotypically normal father is generally considered not to be clinically significant, we hypothesize that this too might have had an additive effect on the proband.

1206T

TRPM1, the transient receptor potential cation channel M1, harbors rare putatively damaging missense variants disproportionately transmitted to affected sibs in schizophrenia quads. *S. Gulsuner[1], T. Walsh[1], M.-C. King[1], J.M. McClellan[2].* 1) Department of Medicine (Medical Genetics), University of Washington, Seattle, WA; 2) Department of Psychiatry, University of Washington, Seattle WA.

Schizophrenia can be caused by de novo damaging mutations, either point mutations or CNVs. We extended this model to search for rare inherited damaging mutations in schizophrenia. We screened exome sequence data of the parents of 81 quads, each including a proband with schizophrenia and a healthy sibling. For each gene, all rare putatively damaging variants were identified and aggregated gene by gene. Applying the rare variant transmission disequilibrium test with discordant siblings (RV-TDT-DS), we calculated chi-square values for transmission disequilibrium of these variants for every gene. Genome-wide, the gene TRPM1 was the most extreme outlier for disequilibrium in transmission of rare damaging variants. At TRPM1, 14 parents carried putatively damaging variants, which were transmitted to 11 of 14 affected children and to 2 of 14 unaffected siblings (X2 = 11.63, 1 df; P=0.0006). All rare putatively damaging variants were missense mutations at highly conserved residues. TRPM1 encodes the Transient Receptor Potential Cation Channel M1, which is highly expressed in retina, melanocytes, and brain. In rod bipolar cells, TPRM1 mediates synaptic transmission, specifically by depolarizing bipolar cells in response to glutamate release from photoreceptors. Recessive loss-of-function mutations in TRPM1 lead to congenital stationary night blindness. Recurrent gene-disrupting CNVs at TRPM1 have been previously identified in both schizophrenia and autism patients. TRPM1 lies in the 1.5 MB hotspot for such CNVs on 15q13.3. We suggest screening other cohorts of quads or trios with schizophrenia or autism for transmission disequilibrium at TRPM1, and conversely, evaluating the frequency of mental illness phenotypes in parents of children with congenital stationary night blindness. We also suggest evaluating the possible role of TRPM1 in synaptic transmission in brain structures.

1207S

Screening for Mutations in Non-Syndromic Autosomal Recessive Intellectual Disability Genes in Non-Consanguineous Intellectual Disability and Autism Populations. *X. Liu[1], K. Mittal[2], N. Vasli[2], A. McNaughton[1], A. Mikhailov[2], K. Roetzer[3], M. Hudson[1], C. Windpassinger[3], P. Magee[1], L. Al Ayadhi[4], W. Kaschnitz[3], E. Petek[3], D.J. Stavropoulos[5], M.J. Carter[6], M. Ayub[1], J.B. Vincent[2].* 1) Dept Psychiatry, Queen's Univ, Kingston, ON, Canada; 2) MiND LAB, CAMH, Toronto, ON, Canada; 3) Human Molecular Genetics Institute, Medical University of Graz, Graz, Austria; 4) King Saud University, Ryadh, Saudi Arabia; 5) Department of Paediatric Laboratory Medicine, Hospital for Sick Children, Toronto, ON, Canada; 6) Department of Pediatrics, Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, University of Toronto, Toronto ON Canada.

Intellectual disability (ID) and autism spectrum disorder (ASD) are believed to occur each with a prevalence of ~1% . In addition, ID is present in upto 70% of individuals with ASD, and in ID populations as many as 28% may meet criteria for ASD. Both ASD and ID are frequently the result of genetic aberrations, however only a fraction of patients currently receive a genetic diagnosis. Genes identified for ID have frequently been implicated in ASD and vice versa. Autosomal recessive (AR) mutations are the main cause of ID in populations where consanguinity rates are high, but even in outbred populations the rate may be 13-24%. Here we use a targeted next generation sequencing approach using Ampliseq primer pools and the Ion Proton platform to assess the involvement of known recessive ID genes in outbred ID and ASD populations, and comparison to some known dominant and X-linked genes. We combined this with a pooling strategy to enable mass screening. Our primer pools targeted 96 known ID and ASD genes, including 67 genes reported as non-syndromic ARID genes, 7 reported as syndromic ARID genes, 3 reported as AR ASD genes, 12 reported as autosomal dominant ASD or ID genes, and 6 reported as X-linked ID or ASD genes. To date, using pools of 20, we have screened ~1,200 ASD and 740 ID individuals from outbred populations. We included as positive controls two samples, with known homozygous base substitution or deletion. Our work flow and pipeline aimed to identify nonsense and damaging missense mutations and indels, homozygous or compound heterozygous for AR, heterozygous for dominant and X-linked recessive and hemizygous for X-dominant. Alignment and analysis was performed using Ion Torrent and Ion Reporter. The control base substitution was present in ~10% of reads for that pool, indicating good coverage and identification, however the 5bp deletion was present in only 8 out of 1300 (0.6%), suggesting the alignment for indels is inefficient. Although analysis and validation is ongoing, we have identified numerous variants of interest in NS-ARID genes such as MAN1B1, but very few in others. We have also identified variants of interest in X-linked genes PTCHD1 and MECP2. A pooling strategy for targeted gene sequencing is an efficient means of screening large numbers of patients for multiple candidate genes, however improved alignment algorithms are needed to identify indels. Some AR genes may be relatively common causes of ID and ASD in outbred populations.

1208M

Association study of *TREM2* exon 2 variants with late-onset Alzheimer's disease in Iranian elderly population. *Z. Mehrjoo[1], A. Najmabadi[2], S.S. Abedini[1], K. Kamali[3], H. Najmabadi[1], H.R. Khorram Khorshid[1].* 1) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; 2) Division of Biological Sciences, University of California, San Diego, California, USA; 3) Reproductive Biotechnology Research Center, Avicenna Research Institute, Academic Center for Education Culture and Research (ACECR), Tehran, Iran.

Dementia, a major elderly disability, is increasing all over the world, especially in developing countries. The most common form of dementia is Alzheimer's disease (AD [MIM 104300]), which is defined as a neurodegenerative disorder that affects memory, behavior and thinking ability; it then impairs basic body movement, and eventually leads to death. Recent studies on AD disclosed significant association between the rs-75932628-T variant of the *TREM2* [MIM 605086] gene and AD in European and North American populations, but not in Han Chinese individuals where no rs-75932628-T was identified in AD patients and controls. So far none of these association studies have covered Middle Eastern population. We performed this study to analyze the association between *TREM2* exon 2 variants and late-onset AD in Iran. Until now a total of 131 AD patients and 157 controls have been genotyped using PCR and Sanger sequencing. The rs-75932628-T allele frequency in this cohort of Iranian elderly population was higher than other reported populations (0.86%) but did not show a statistically significant association with AD (odds ratio [OR]: 4.8; 95% confidence interval [CI]: 0.54 to 43.6; P = 0.270). Although we identified more rare variants in AD patients compared to controls, the abundance of TREM2 rare variants do not reach a statistically significant association with AD, which could be a result of low sample size. Therefore, to continue this study and to obtain meaningful results, we are collecting and genotyping more samples.

1209T

Increased Genome-wide Burden of Rare Coding Variants in Schizophrenia. *L.M. Olde Loohuis[1], J.A.S Vorstman[2], A.P. Ori[1], K.A. Staats[1], T. Wang[1], J. DeYoung[1], R.S. Kahn[2], R.M. Cantor[1,3], R.A. Ophoff[1,2,3].* 1) Center for Neurobehavioral Genetics, University California Los Angeles, Los Angeles, CA; 2) Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 3) Department of Human Genetics, University of California Los Angeles, Los Angeles, USA.

Schizophrenia is a highly heritable polygenic disorder with suspected neurodevelopmental and immunological components. While GWAS continue to uncover common single nucleotide variants (SNVs) contributing to this phenotype, to date, these variants explain only a portion of the estimated heritability. In addition to common variants, rare variants have been found to play a role in disease risk, including de novo mutations and genomic copy number variants. However, due to their low minor allele frequencies and modest effect sizes, it is difficult to establish association with rare SNVs using standard association tests. We assessed the burden of rare SNVs in schizophrenia by analyzing the cumulative effects of coding variants occurring in cases or controls only. Our analysis is based on exome SNP array genotyping of 1,002 schizophrenia cases and 931 controls from a homogeneous population from the Netherlands. Using CONDEL, a scoring tool for estimating the functional impact of non-synonymous SNVs, we assigned a burden score to each individual by summing the scores of every variant carried. In cases versus controls, we observe a significantly increased individual genome-wide rare variant burden for deleterious non-synonymous variants (empirical P = 0.024). In addition, an increased number of genes with double hits (empirical P = 0.035), as well as splice site variants (empirical P = 0.003) were seen. The genes implicated by the increased burden of rare coding variants overlap significantly with genes from recent GWAS (empirical P = 0.033), and are enriched for those that are expressed in fetal brain and spleen. We demonstrate that non-synonymous rare variants are important in the etiology of schizophrenia. Previous reports suggest the enrichment primarily of singleton nonsense variants in schizophrenia in a set of pre-selected genes. Our results now extend these and implicate non-synonymous SNVs at a genome-wide level. This signal is enriched for genes implicated through GWAS studies of schizophrenia, implying the contribution of common and rare variants to schizophrenia from the same loci. The genes containing these rare coding variants overlap significantly with genes expressed in the fetal brain and spleen, highlighting the potential involvement of neurodevelopment and immune system in disease etiology.

1210S

Targeted sequencing of candidate genes identified in extended families with Alzheimer disease. *J. Rehker[1], R. Levy[1], R. Nesbitt[1], Q. Yi[2], B. Martin[2], D. Nickerson[2], W. Raskind[2], J. Shendure[2], Z. Brkanac[1].* 1) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 2) Department of Genome Sciences, University of Washington, Seattle, WA.

Late onset Alzheimer's disease (AD) is a most common cause of dementia. The genetics of AD is complex with contribution from rare and common variants. In addition to PSEN1, PSEN2 and APP rare variants with large effects were recently identified in TREM2 and PLD3. This indicated that additional genes might play a role in AD. To identify novel genes we have performed exome sequencing in 14 families with 4 or more affected cases that include at least one distant relative (cousin). For each family we have performed frequency filtering to identify private functional variants that most closely co-segregate with the disease. We defined private variants as not present in 1000 Genomes, ESP6500 and dbSNP132. This way we have identified 21 novel candidate genes. To obtain statistical evidence for association we performed a large gene-based case-control study. For 21 candidate genes, whose coding sequence encompass 68,518 bp, and 8 known AD genes we have performed Molecular Inversion Probes (MIP) capture and targeted sequencing of 879 familial AD cases and 478 cognitively normal elderly controls. The cases were obtained from NCRAD and NIMH collections and controls from ACT study. Variant calling and annotation were performed in the same manner as for exome sequencing with modifications due to MIP capture and multiplexing. After quality control, 839 cases and 466 controls were used in analysis. In our analysis we have identified 14 rare functional PSEN1 variants in FAD cases and none in controls (P= 0.0054, Chi-square) thus confirming the validity of our case-control approach. Our case-control analysis excluded 8 genes as candidates as they had more private variants in controls as compared to cases. For three novel genes the association with AD was nominally significant at $p<0.05$ and for 10 genes the analysis was inconclusive due to lack of power or low level of sequencing coverage. The novel gene with the strongest association has 25 rare functional variants in cases and 1 variant identified by the same criteria in controls (p=0.0007). This is highly significant but it does not reach significance level of 10-6 required for genome-wide correction. In order to increase the power of our study we have performed additional sequencing of 13 candidate genes in 384 cases and controls from NCRAD-LOAD study, and 220 cases and 150 controls from Washington University. We will report the results of our combined analysis.

1211M

Identification of Molecular Markers in Parkinson's Disease Using Next Generation Sequencing. *S.M. Sperber[1], Y. Shi[2,3], R. O'Rourke[3], K. Jones[3], E. Berry-Kravis[4], E. Spector[2,3], D. Hall[4].* 1) Genetic Testing Laboratory, Department of Genetics & Genomic Sciences, Icahn School of Medicine at Mt. Sinai, New York, NY; 2) Children's Hospital of Colorado, Aurora, CO; 3) University of Colorado AMC, Aurora, CO; 4) Rush University, Chicago, IL.

Parkinson's disease (PD) is a complex heterogenic neurodegenerative disorder with insidious onset and devastating progressive decline currently affecting over 4.5 million older Americans. Prevalence is projected to double by 2030 causing considerable economic burden. High impact genes that contribute to PD have been identified in individual families and by GWAS studies. How different genetic variants influence the PD phenotype has yet to be fully explored. A targeted test could rapidly identify variants in multiple genes simultaneously that provide risk information and may be informative for potential phenotypic categorization necessary for future therapies. Here we identified genetic variants contributing to PD in a cohort of clinically defined PD patient samples compared to genome population data using a targeted panel of genes associated with PD and next generation sequencing techniques. Ninety-nine patients meeting Gelb criteria for PD were interviewed, examined, and had blood collected for genotyping. The mean age was 68.9±10.3 years, 65% male, and age of onset of disease was 62.1±10 years. Mean UPDRS motor score was 25±11 and mean UPDRS total score was 40±16. A set of high impact genes associated with PD was exon sequenced in a targeted manner, using a PCR-based enriched panel. Sequencing the cohort's genes identified on average, greater than 300 variants per patient. One hundred and nine variants of potential clinical interest were identified. Of the 109 variants, 26 were previously identified as pathogenic, 34 were of uncertain significance, 2 were suggested to be associated with a reduced risk of PD and 47 are likely benign. The patients exhibited variant profiles that highlight the utility of genetic testing to further define the gene pathways and their associated phenotypic characteristics. The targeted nextgen panel is sensitive, cost effective and efficient. The information collected will provide a necessary link between phenotypic and genotypic data pointing to particular pathways for further investigation. Acknowledgements: Parkinson's Disease Foundation, Anti-Aging Foundation, Denver Genetic Laboratories.

1212T

Mutation in the chromatin-remodeling factor BAZ1A is associated with intellectual disability. *A. Zaghlool, J. Halvardson, J. Zhao, A. Kalushkova, K. Konska, H. Jernberg-Wiklund, A-C. Thuresson, L. Feuk.* Immunology, Genatics & Patholohy, Science for Life Laboratory, Uppsala University, Sweden.

Exome sequencing of trio samples has led to the identification of several mutations in genes involved in chromatin remodeling in syndromic forms of intellectual disability. However, in most cases functional studies to understand the biological contribution of the mutations to the disease symptoms are still lacking. Here, we used exome sequencing to identify a single non-synonymous de novo mutation in BAZ1A, encoding the ATP-utilizing chromatin assembly and remodeling factor 1 (ACF1), in a patient with severe intellectual disability, autism, profound speech impairment, seizures, hypotonia, hypoplastic finger-nails and distal phalanges. ACF1 has been previously reported to bind to the promoter of vitamin D receptor (VDR) regulated genes and suppress their expression in the absence of vitamin D. Using RNA sequencing, we found that the mutation in BAZ1A affects the expression of many genes, mainly involved vitamin D3 metabolism, synaptic function and extra cellular matrix organization. The differential expression of CYP24A1, involved in vitamin D3 metabolism, and SYNGAP1, involved in axonal formation and synaptic function, correlate with the clinical diagnosis of the patient. By chromatin immunoprecipitation analysis, we show that the differential expression of CYP24A1 is caused by reduced binding affinity of the mutated ACF1. We therefore propose that BAZ1A represents yet another chromatin remodeling gene involved in causing an intellectual disability syndrome.

1213S

Whole Exome Sequencing in Females with Autism Implicates Novel and Candidate Genes. *M.G. Butler[1], S.K. Rafi[1], H. Wang[2], D. Stephan[3], A.M. Manzardo[1].* 1) Psychiatry δ Behavioral Sciences and Pediatrics, University of Kansas Medical Center, Kansas City, KS, USA; 2) Silicon Valley Biosystems, Foster City, CA, USA; 3) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

Classical autism or autistic disorder belongs to a group of genetically heterogeneous conditions known as Autism Spectrum Disorders (ASD) with a predominant male to female (4:1) ratio. The overall occurrence rate is higher than epilepsy or Down syndrome. Although understudied in females, heritability estimates are as high as 90%. A recent review of genetic linkage data, candidate genes and genome-wide association studies along with high resolution DNA microarray and next generation sequencing have led to a compilation of clinically relevant candidate and known genes for ASD for a total of 629 genes (Butler et al., 2014). We chose to undertake whole exome sequencing of 30 well-characterized Caucasian females with autism (average age, 7.7 ± 2.6y; age range, 5 to 16y) selected from the Autism Genetic Research Exchange (AGRE) repository (www.agre.autismspeaks.org) from multiplex families having a high probability of causation due to gene disturbances. Genomic DNA (5μg) samples were used for whole exome sequencing via paired-end next generation sequencing approach using the Illumina HiSeq2000 platform with the Agilent SureSelect Human All Exon v4 - 51Mb. Primary sequence data files were aligned to the reference, variants called and functional significance of each variant was calculated to generate a rank-ordered list. The list of disease causing genes were developed with the primary selection criteria using a machine learning-derived Classification (G2M) score and adding cutoff levels for other predictive parameters (GERP2, PolyPhen2, and SIFT). Putative candidate genes and genomic variants were then subjected to further screening for biological significance. We identified between 100 and 300 genes showing genomic variants of novel or candidate genes for autism per subject using the Classification (G2M) score >0.5, increasing to >0.7 further narrowed the list to 10 to 20 genes. Seventy-eight genes were identified as meeting our selection criteria, range of 1 to 9 genes per female. Five subjects presented with disturbances of X-linked genes (*SYTL4, GPRASP2, PIR, IL1RAPL1, GABRQ*). The cadherin, protocadherin and ankyrin repeat gene families were most commonly disturbed (e.g., *CDH6, FAT2, PCDH8, CTNNA3, ANKRD11*) along with genes related to neurogenesis and neuronal migration (e.g., *SEMA3F, MIDN*) indicating the usefulness of whole exome sequencing in females with ASD from multiplex families to enable identification of known and novel gene mutations.

1214M

Novel compound heterozygous PIGT mutations caused multiple congenital anomalies-hypotonia-seizures syndrome 3. *M. Nakashima[1], H. Kashii[2], Y. Murakami[3], M. Kato[4], Y. Tsurusaki[1], N. Miyake[1], M. Kubota[2], T. Kinoshita[3], H. Saitsu[1], N. Matsumoto[1].* 1) Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Division of Neurology, National Center for Child Health and Development, Tokyo, Japan; 3) Research Institute for Microbial Diseases and World Premier International Immunology Frontier Research Center, Osaka University, Osaka, Japan; 4) Department of Pediatrics, Yamagata University Faculty of Medicine, Yamagata, Japan.

Recessive mutations in genes of the glycosylphosphatidylinositol (GPI)-anchor synthesis pathway have been demonstrated as causative of GPI deficiency disorders associated with intellectual disability, seizures, and diverse congenital anomalies. In this study, we performed whole exome sequencing in a Japanese patient with progressive encephalopathies and multiple dysmorphism with hypophosphatasia and identified novel compound heterozygous mutations, c.250G>T (p. Glu84*) and c.1342C>T (p. Arg488Trp), in PIGT encoding a subunit of the GPI transamidase complex. The surface expression of GPI-anchored proteins (GPI-APs) on patient granulocytes was lower than that of healthy controls. Transfection of the Arg488Trp mutant PIGT construct, but not the Glu84* mutant, into PIGT-deficient cells partially restored the expression of GPI-APs DAF and CD59. These results indicate that PIGT mutations caused neurological impairment and multiple congenital anomalies in this patient.

1215T

Common, low frequency, and rare coding variants in CHRNA5 contribute to nicotine dependence in European and African Americans. *E. Olfson[1], N.L. Saccone[2], E.O. Johnson[3], N. Breslau[4], D. Hatsukami[5], K. Doheny[6], L. Fox[1], S.M. Gogarten[7], K. Hetrick[6], C.C. Laurie[7], B. Marosy[6], J. Stitzel[8], J. Rice[1], A. Goate[1], L.J. Bierut[1].* 1) Department of Psychiatry, Washington University School of Medicine, St Louis, MO; 2) Department of Genetics, Washington University School of Medicine, St Louis, MO; 3) Behavioral Health Epidemiology Program, RTI International, Research Triangle Park, NC; 4) Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI; 5) Department of Psychiatry, University of Minnesota, Minneapolis, MN; 6) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 7) Department of Biostatistics, University of Washington, Seattle, WA; 8) Institute for Behavioral Genetics, University of Colorado, Boulder, CO.

Purpose: The functional nonsynonymous variant rs16969968 in the α5 nicotinic receptor subunit gene (CHRNA5) is the single strongest genetic risk factor for nicotine dependence in European Americans (MAF=0.35), and evidence suggests a similar contribution to risk in African Americans (MAF=0.06). Additional variants in CHRNA5 likely increase liability for nicotine dependence. This study examines targeted sequence data from approximately 3,000 unrelated cases and controls to assess the influence of CHRNA5 coding variation on nicotine dependence. Methods: Community-based recruitment enrolled subjects aged 25-45. Cases had a Fagerstrom Test for Nicotine Dependence score of ≥4 and controls had a score of 0 or 1. Custom next-generation sequencing with mean on-target coverage of 180X was performed on regions associated with smoking behaviors. Logistic regression was used to model case-control status with the variables sex, age, ethnicity, individual common variants (MAF≥5%), individual or collapsed low frequency variants (5%>MAF≥0.5%), and collapsed rare variants (MAF<0.5%). Results: Sequencing identified 24 nonsynonymous variants and 2 frameshift deletions of high quality in CHRNA5. The well-studied rs16969968 was the only common variant among these, and the minor allele was associated with increased risk of nicotine dependence (OR=1.28, p= 0.0007). Three low frequency nonsynonymous variants were identified, and all individually trended in the risk direction. The collapsed low frequency variant term demonstrated that presence of a minor allele was associated with nicotine dependence (OR=1.45, p=0.01). The remaining 22 coding variants were rare (each occurred in 1-4 individuals), and similarly, the collapsed term revealed a risk effect (OR=2.39, p=0.03). Nagelkerke's adjusted R^2 was used to assess the proportion of nicotine dependence variation explained by CHRNA5 variants after adjustment for sex and age. In European Americans, the well-studied rs16969968 variant gave an R^2 of 0.9%, and the addition of all CHRNA5 coding variants increased the R^2 to 2.4%. In African Americans, rs16969968 gave an R^2 of 0.2%, and adding all coding variants increased the R^2 to 0.8%. Conclusions: Our findings suggest that common, low frequency, and rare coding variants in CHRNA5 are associated with increased risk of nicotine dependence. Coding variation in CHRNA5 accounts for over 2% of the nicotine dependence variation in European Americans and close to 1% in African Americans.

1216S

Trio-based exome sequencing indicates the ion homeostasis is relevant to bipolar disorder. *N. Matoba[1,2], M. Kataoka[1,3], K. Fujii[4], T. Kato[1].* 1) Laboratory for Molecular Dynamics of Mental Disorders, RIKEN BSI, Wako, Saitama, Japan; 2) Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan; 3) Department of Child Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku,Tokyo, Japan; 4) Department of Psychiatry, Dokkyo Medical University School of Medicine, Mibu,Tochigi, Japan.

Background Bipolar disorder (BD) is a mental disorder characterized by recurrent manic and depressive episodes. Heritability of BD is reportedly around 59-87% (Smoller et al., 2003). Though recent genome wide association studies (GWAS) showed associated with several SNPs, each candidate SNP has a weak effect and can collectively explain only 38% of heritability (Lee et al., 2011). Therefore, we hypothesize that multiple rare variants may also contribute to BD.

Methods We recruited participants mainly through E-mail newsletter of Bipolar Disorder Research Network Japan. All probands and their parents were interviewed by a trained psychiatrist using M.I.N.I. In total 59 trios including 35 original trios and independent 24 trios were analyzed. From the exome sequencing data of these trios, we selected only rare variants referring to public databases such as dbSNP137. These variants were classified as “transmitted” or “un-transmitted” based on the inheritance pattern. Then we investigated the characteristics of the set of transmitted variants compared to those in un-transmitted.

Results There was no significant difference of numbers of variants between two groups. Gene ontology analysis showed significant enrichment of 57 GO terms in transmitted variants, of those 6 GOs were confirmed in independent 24 trios. These GOs included calcium ion homeostasis. In addition, this finding was not observed in trios of healthy siblings of probands with autism spectrum disorder (ASD), which showed slightly similar trends in some of those GOs.

Discussion Our finding highlights the role of multiple rare damaging variants related to ion homeostasis in the pathophysiology of BD, which has been supported by genetic association with SNPs of calcium channel genes by GWAS and elevated calcium levels in blood cells among others. The findings need to be replicated in a larger set of families including multiple ethnicities. Since we directly compared "transmitted" and "un-transmitted" variants, the result of gene enrichment analysis are robust to artifacts due to gene length or GC contents and so on. This simple but robust approach could also be utilized to other complex genetic diseases.

1217M

Genome-wide analysis of copy-number variation in Canadian children with developmental coordination disorder implicates neurodevelopmental genes. *F.P. Bernier[1,2,3], S.J. Mosca[2], L.M. Langevin[3,4], A.M. Innes[1,2,3], A.C. Lionel[5,6], C.C. Marshall[5,6], S.W. Scherer[5,6], J.S. Parboosingh[2,3,7], D. Dewey[3,4,8].* 1) Clinical Genetics, Alberta Children’s Hospital, Calgary, AB, Canada; 2) Department of Medical Genetics, University of Calgary, Alberta, Canada; 3) Alberta Children’s Hospital Research Institute, University of Calgary, Alberta, Canada; 4) Department of Pediatrics, University of Calgary, Alberta, Canada; 5) The Centre for Applied Genomics and Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 6) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 7) Genetic Laboratory Services, Alberta Health Services, Calgary, AB, Canada; 8) Department of Community Health Sciences, University of Calgary, Alberta, Canada.

Developmental coordination disorder (DCD) is a common neurodevelopmental disorder characterized by functional motor performance deficits that interfere with daily activities in children and adults. In this pilot study we performed copy number variant (CNV) analysis in 82 well-characterized Canadian children between the ages of 8 and 17 with DCD and 2,988 healthy European controls using identical 2,500,000 SNP microarrays and CNV calling algorithms. We detected a statistically significant increased rate of both rare CNVs (p=0.018) and rare, genic CNVs (p=0.009) between 500 and 1000Kb in cases compared to controls. Within genic CNVs, there was an enrichment of duplications spanning brain-expressed gene (p=0.039) and genes previously implicated in other neurodevelopmental disorders (p= 0.043). Loci previously implicated in other neurodevelopmental disorders were investigated further. Genes and loci of particular interest in this group included: GAP43, RBFOX1, PTPRN2, SHANK3, 16p11.2, and distal 22q11.2. Our findings provide compelling evidence supporting a genetic basis for DCD, and further implicate rare CNVs in the etiology of neurodevelopmental disorders. Our data also suggests that there might be shared susceptibility genes for DCD and other neurodevelopmental disorders.

1218T

Copy Number Variation in Han Chinese Individuals with Autism Spectrum Disorder. *M.J. Gazzellone[1,2], X. Zhou[1,3,4], A.C. Lionel[1,2], M. Uddin[1], B. Thiruvahindrapuram[1], W. Jia[3], S. Caihong[3], L. Shuang[3], K. Tammimies[1,5], S. Walker[1], J. Wei[1], Z. Wang[1], Z. Mingyang[3], L. Wu[3], S.W. Scherer[1,2].* 1) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 3) Department of Children’s and Adolescent Health, Public Health College of Harbin Medical University, Harbin, Heilongjiang, People’s Republic of China; 4) Heilongjiang Provincial Centre for Disease Control and Prevention, Harbin, Heilongjiang, People’s Republic of China; 5) Center of Neurodevelopmental Disorders, Department of Women’s and Children’s Health, Karolinska Institutet, Stockholm, Sweden.

Background: Autism Spectrum Disorders (ASDs) are a group of neurodevelopmental conditions with a demonstrated genetic etiology. Rare (<1% frequency) copy number variations (CNVs) account for a proportion of the genetic events involved, but the contribution of these events in non-European ASD populations remains relatively unexplored. As clinical microarray testing becomes the adopted standard of care across medical genetics labs worldwide, studies describing the genetic architecture in these other ASD populations will become even more important. Here, we report on rare CNVs detected in a cohort of individuals with ASD of Han Chinese background. **Methods:** DNA samples were obtained from 104 ASD probands and their parents who were recruited from Harbin, China. Probands were diagnosed using the Autism Behavior Checklist (ABC) and Childhood Autism Rating Scale (CARS). The DNA samples were genotyped on the Affymetrix CytoScan HD platform, a high-resolution array comprised of nearly 2.7 million probes. High-confidence CNVs were detected using multiple CNV-calling algorithms. Rare CNVs were identified by comparing data with 873 technology-matched controls from Ontario and 1,235 additional population controls of Han Chinese ethnicity. **Results:** 8.6% of the probands had at least 1 *de novo* CNV (overlapping the *GIGYF2*, *SPRY1*, 16p13.3, 16p11.2, 17p13.3-17p13.2, *DMD*, and *NAP1L6* genes/loci). Rare inherited CNVs affected other plausible neurodevelopmental candidate genes including *GRID2*, *LINGO2*, and *SLC39A12*. An Asian-specific copy number polymorphism (duplication) was also identified at the *YWHAE* locus, which has previously been described as etiologic in ASD and other developmental disorders. **Conclusions:** Our findings help define genomic features relevant to ASD in the Han Chinese and emphasize the importance of using ancestry-matched controls in medical genetic interpretations.

1219S

Next Generation Sequencing for the study of ALS and other Motor Neuron Diseases. *C. Gellera[1], B. Castellotti[1], V. Pensato[1], S. Magri[1], D. Di Bella[1], G. Lauria[2], N. Ticozzi[3], F. Taroni[1].* 1) Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS - Istituto Neurologico Carlo Besta, Milan, Italy; 2) Unit of Neurology,Fondazione IRCCS - Istituto Neurologico Carlo Besta, Milan, Italy; 3) Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milan, Italy.

ALS is an adult-onset, progressive, neurodegenerative disease caused by the selective loss of motor neurons with a large prevalence of sporadic cases (SALS). The pathogenesis is largely unknown, but genetic factors likely play a major role in the disease. Although mutations in several genes have been described, the underlying cause is still unknown in about 50% of familial ALS (FALS) and in the large majority of SALS (80%). Genetic determinants are unlikely to be explained by traditional strategies (linkage analysis or candidate gene approach). In order to better define the genetic contribution in SALS and to fill the genetic gap in FALS, we set up an integrated NGS approach as follow: 1) customized gene panel screening, 2) wide exome sequencing (WES), 3) variant filtering and validation. The panel has been designed by TruSeq Custom Amplicon Studio Design Illumina and includes 31 ALS causative genes. Validation process indicates that these genes are analyzed with a coverage of > 90%. Though still in a limited number of SALS (n=26) (selected to be negative for SOD1, FUS, TARDPB and C9ORF72 gene mutations), data analysis allowed the identification of several variants of possible pathogenic significance. After a first filtering step which excludes intronic variants and known polymorphisms with MAF > 1%, we get a report showing on average 50 variations per sample in exonic or splicing regions. In 10 cases we identified at least one variant of possible pathogenetic significance: 2 cases with a missense variants in GARS gene, 1 case with a missense mutation in VCP gene already described in association with the disease, 1 case with a missense mutation in OPTN gene, 1 case with a missense mutation in SIGMAR1 gene, 1 case with a stop mutation in NEFH gene, 1 case with a missense mutation in DCTN1 gene, 1 case with 2 missense variants in SETX gene, 1 case with a missense variant in SQSTM1 gene, 1 case with a splicing mutation in ASAH1 gene. The genetic panel designed and validated in our laboratory will be used for a high genetic definition process not only for FALS but also for SALS or other overlapping motor neuron phenotypes. WES is currently in progress for 85 FALS of our cohort in an International collaborative study (John Landers, University of Massachusetts; Vincenzo Silani, University of Milano; Cinzia Gellera, Istituto Neurologico Besta, Milano) supported by ARISLA foundation.

1220M

Exome sequencing of mesial temporal lobe epilepsy with hippocampal sclerosis in parent-offspring trios. *S.S. Cherny[1], J.K.L. Wong[1], P.C. Sham[1], L. Baum[2], P. Kwan[3].* 1) Department of Psychiatry, University of Hong Kong, Pokfulam, Hong Kong; 2) School of Pharmacy, Chinese University of Hong Kong, Shatin, Hong Kong; 3) Department of Medicine, University of Melbourne, Australia.

Mesial temporal lobe epilepsy related to hippocampal sclerosis (MTLE-HS) is recognized to be the most drug resistant type of epilepsy, and HS is the most common pathological substrate identified in pharmacoresistant temporal lobe epilepsy. The condition is characterized by relatively well-defined clinical, electrographic, radiologic, and pathologic changes. Despite attempts in finding genetic variants associated with epilepsies by genome-wide association studies on European and Chinese subjects by our team and collaborators, limited common variants reached genome-wide significance. Meanwhile, the causes of the pathological changes in MTLE-HS remain unknown, and no preventive intervention has been identified. Discovery of novel genetic markers could allow screening patients with aggressive epilepsy subtypes for early surgical or pharmaceutical intervention, and ultimately lead to genetic therapies. To further explorer the contribution of new, rare variants in MTLE-HS, we plan to sequence exomes of 30 trios of probands with their unaffected parents. So far we have sequenced 9 trios, and 10 *de novo* mutations were identified. Among these, two *de novo* mutations are within genes involved in neurological functions (*NLGN3* and *NBEAL1*) and might be involved in disease development. *NLGN3* was reported to be associated with X-linked autism in a Swedish family and given the high comorbidity of autism spectrum disorders with epilepsy, this *de novo* mutation might contribute to MTLE-HS development.

1221T

Homozygous mutation in Synaptic Vesicle Glycoprotein 2A gene results in intractable epilepsy, microcephaly, intellectual disability and growth retardation. *A. Huq, F. Serajee.* Pediatrics and Neurology, Wayne State University, Detroit, MI.

Synaptic Vesicle Glycoprotein 2A (SV2A) is the binding site of anti-epileptic drug levetiracetam and thus the only known synaptic vesicle target of central nervous system targeted treatment. To date, no pathogenic variant in SV2A gene has been identified in human. We have identified a homozygous mutation in SV2A gene in a patient with intractable epilepsy, microcephaly, intellectual disability and growth retardation. The proband is 4-year-and 6-month old girl of South Asian descent whose parents are multiply consanguineous. She was born healthy but never developed any ability to hold her head up, track or reach for objects. At 2 months of age, she developed seizures and dystonic posturing. The EEG showed multifocal spikes, giant spike and waves correlating with myoclonic seizures and diffuse attenuation superimposed by diffuse fast wave activity correlating with tonic spasms. Her seizures did not respond to antiepileptic medications including clonazepam, topiramate, and levetiracetam and ketogenic diet. The family refused any trial of Vigabatrin or ACTH. The proband also had a brother who died at 7 months of age after a similar disease course. Laboratory studies revealed normal chromosome microarray pattern, mitochondrial studies and metabolic profile. Full exome sequencing, and bioinformatics analysis suggested a homozygous arginine to glutamine mutation in amino acid position 383 (R383Q) in exon 5 of SV2A gene as the candidate mutation. Both the parents were carriers for R383Q variant. Mutations in SV2a gene are not currently known to cause any Mendelian disorder in human. The R383Q change is not observed in known healthy cohorts, exome databases and Database of Single Nucleotide Polymorphisms. Given that heterozygous parents are healthy, it is likely R383Q change acts in a recessive manner. Although SV2A influences exocytosis of neurotransmitter-containing synaptic vesicles, at a cellular level the exact point at which neurotransmitter release is affected in synapse by SV2A remains unknown. The mutation in our patient lies on the cytosolic loop before the 7th transmembrane domain of the protein and alters the second adenine binding domain in SV2A protein suggesting that the mutation may alter SV2A-mediated vesicle priming by adenine nucleotides.

1222S

Personalized medicine in the treatment of epilepsy. *R.G. Lafreniere[1], E. Rossignol[2], F.F. Hamdan[2], S. Mahmutoglu[3], S.L. Girard[1], D. Andrade[3,4], A. Dionne-Laporte[5], C. Boelman[3], M. Blazejczyk[6], O. Diallo[5], A.M. Laberge[2], M. Cadieux-Dion[1], M.P. Dubé[6], G.A. Rouleau[5], H.C. Mefford[7], B. Minassian[3], J.L. Michaud[2], P. Cossette[1], the Canadian Epilepsy Network (CENet).* 1) CHUM Research Centre, Montreal, Canada; 2) Centre Hospitalier Universitaire Sainte-Justine Research Center, Montreal, Canada; 3) The Hospital for Sick Children, Toronto, Canada; 4) Toronto Western Hospital, Toronto, Canada; 5) McGill University, Montreal, Canada; 6) Montreal Heart Institute, Montreal, Canada; 7) Department of Pediatrics, University of Washington, Seattle, WA.

Epilepsy is a common complex neurological disorder characterized by recurrent episodes of paroxysmal abnormal electrical discharges in the brain (seizure). In 65% of epileptic individuals seizures occur without detectable lesion in the brain. Intellectual disability (ID) is observed in ~25% of epileptic individuals. Although treatment is effective in the majority of cases, more than 30% of patients are resistant to anti-epileptic drugs (AEDs). In this project, we hypothesize that pharmaco-resistant epilepsies (PRE) have different etiologies than other epilepsy syndromes associated with the same clinical manifestations, and posit that genetic factors can be used to predict response to AEDs. In 2013, we united Canadian clinicians and investigators to create the Canadian Epilepsy Network (CENet). In order to identify highly penetrant genetic variants that can predict resistance to AEDs, we will conduct whole genome sequencing (WGS) on 1000 epileptic individuals with extremes of response to AED treatment, including 200 multiplex families, as well as 200 epileptic trios (patient + both parents) with severe epilepsy and ID. Based on our previous studies in patients with ID or autism spectrum disorders, we expect a high rate of de novo mutation for epilepsies associated with these co-morbidities. Potentially causative genes/variants will be validated in additional epilepsy samples available through our large collaborative network with similar but complementary projects in the US (Epi4K - 4,000 epilepsy genomes) and Europe (EpiPGX). We will also assess the clinical utility of whole genome sequencing as a diagnostic test for epilepsy in a prospective cohort of 600 epileptic individuals. So far, pre-screening of 36 known genes for epileptic encephalopathy (EE) in 95 patients led us to identify 11 causative mutations in total (11.6%). At the end of the project, we expect to identify 50 novel genes causing severe forms of epilepsy and PRE. Using this catalog of gene variants, we will develop and validate a diagnostic resource (PRE-GENE) to implement WGS as the gold-standard diagnostic test for PRE in clinical practice.

1223M

Identification of rare variants from exome sequencing in a large family with dyslexia. *A. Carrion-Castillo[1], C. Francks[1,2], B. Franke[3], S.E. Fisher[1,2].* 1) Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; 2) Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands; 3) Radboud university medical center, Radboud University, Nijmegen, Netherlands.

Dyslexia is one of the most common human neurodevelopmental disorders (5-10% prevalence), and it has a complex genetic aetiology. So far, the common disease - common variant approach has not identified many risk alleles. Most of the heritability for dyslexia therefore remains unexplained. With the goal of identifying a rare variant with an individually substantial effect on this complex trait, we analyzed a 30 member multigenerational pedigree with dyslexia, in which evidence for a susceptibility locus in Xq27.3 has previously been identified by parametric linkage analysis based on microsatellite genotyping (de Kovel et al. 2004). In the current study, we sequenced the exomes of 10 affected members of the family. After quality-based and frequency-based (MAF ≤0.01) filtering we identified 177 exonic, non-synonymous, rare variants present in at least 7 affected members and not in a set of unrelated people. Microsatellite genotypes and common SNPs called from the exome data were combined to perform nonparametric linkage analysis, which defined three major regions of interest (NPL> 2) on chromosomes 2q, 20q and Xq, and suggestive regions (NPL> 1) on 13 other chromosomes. The 177 variants identified by sequencing were then inferred in the rest of the family members by sampling inheritance vectors from a Markov chain Monte Carlo analysis of the multilocus marker data. We used the imputed allelic dosages for exomic variants to carry out family-based association analysis and to determine which sequence variants reside on the haplotypes that co-segregate with dyslexia in this family. Association and linkage results were used to prioritize a subset candidate variants for validation by Sanger sequencing in all the family members.

1224T

Analysis of major amyotrophic lateral sclerosis genes in Japan. *R. Nakamura[1], J. Sone[1], N. Atsuta[1], H. Watanabe[1], D. Yokoi[1], H. Watanabe[1], M. Ito[1], J. Senda[1], F. Tanaka[2], G. Sobue[1], the Japanese Consortium for Amyotrophic Lateral Sclerosis research (JaCALS).* 1) Department of Neurology, Nagoya University Graduate School of Medicine, Nagoya, Japan; 2) Department of Neurology, Yokohama City University Graduate School of Medicine, Yokohama, Japan.

Background Amyotrophic lateral sclerosis (ALS) is an adult-onset neurodegenerative disease characterised by progressive upper and lower motor neuron loss. Five to ten percent of ALS patients have a family history of the disease, whereas most patients are sporadic. Mutations in over 20 genes have been identified to cause familial ALS, such as SOD1, TARDBP, FUS, OPTN, ANG and C9orf72. Mutations in these genes have also been identified in some sporadic ALS patients. *Purpose* To investigate the frequency and contribution of mutations in the known ALS genes in sporadic ALS patients in Japan. *Methods* A next-generation sequencing procedure based on the custom AmpliSeq™ panel was designed for sequencing 28 ALS causative or susceptibility genes on the Ion Torrent Personal Genome Machine (PGM™). We analyzed 219 sporadic Japanese ALS patients. *Results* Mutations were identified in 19 of ALS patients. SOD1 mutations were identified in 6 patients, FUS in 3 patients, OPTN in 2 patients, DCTN1 in 2 patients, TARDBP in one patient, SETX in one patient, SPG11 in one patient, ZNF512B in one patient, PRPH in one patient, and FIG4 in one patient. Seven of those mutations were novel non-synonymous variants, and were not present in normal controls. *Conclusions* We identified that 19 sporadic ALS patients have mutations of causative or susceptibility genes of ALS in Japan. This result would help neurologists in clinical practice, especially in genetic counselling.

1225S

De Novo Mutations in Autistic Children from Multiplex Families. *C.L. Simpson[1], Y. Kim[1], C. Wassif[2], N. Hansen[3], J. Mullikin[3], E. Tierney[4], F.D. Porter[2], J.E. Bailey-Wilson[1].* 1) Computational and Statistical Genomics Branch, NHGRI, NIH, Baltimore, MD; 2) Section on Molecular Dysmorphology, National Institute of Child Health, National Institutes of Health, Bethesda, MD; 3) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) Kennedy Krieger Institute, Baltimore, MD.

The presence of excess of de novo single nucleotide variants (SNVs), insertion/deletion (INDEL) and copy number variants (CNVs) has been widely reported in the analysis of whole exome sequencing (WES) in simplex autism families. It is unclear if this is a general excess or if this increased frequency is particularly marked in gene-disrupting mutations. It is also not known if this excessive mutation rate is also observed in multiplex families. We aimed to investigate if affected probands and affected siblings in multiplex families showed excess of de novo mutation compared to unaffected children. We performed WES in 124 individuals from 26 families with 2 or 3 affected children from the Autism Genetic Resource Exchange (AGRE). Poorly performing variants were identified and removed in Golden Helix SVS. Mendelian errors in SNVs were identified using PLINK. Mendelian error rates for the offspring in each pedigree were calculated in R and compared both within and across families using Welch's T test. SNVs were classified into coding and noncoding variation and sub-classified into missense, nonsense and silent mutations. There were 126 affected and 50 unaffected offspring. After quality control, 2.7% of all variants were identified as inconsistent in at least two independent pedigrees and removed as probable sequencing artifacts before analysis leaving 959224 SNVs. Affected individuals did not show higher overall rates of Mendelian inconsistency than unaffected individuals (p=0.4). Comparing only coding SNV mutation rates did not show a significant difference between cases and controls (p=0.8) and there was also no difference when sub-classifying into missense (p=0.6), nonsense (p=0.1) or silent mutations (p=0.6). De novo mutations have been reported as important in simplex autism families and nonsense mutations are the main driver of this excess de novo variation. Overall mutation rates for our multiplex families do not exhibit excessive de novo variation in affected compared to unaffected individuals. Even focusing on only coding variation, no differences appear between cases and controls. Examining the different classes of coding variation, our multiplex families demonstrate a different pattern of de novo variants compared to simplex families. Neither missense, nonsense nor silent mutations show any observed differences between cases and controls. Analyses of additional data and extension of this analysis to INDELs and CNVs are ongoing.

1226M

Differences in Genetic Features May Explain the Discordance of Monozygotic Twins for Schizophrenia. *C. Castellani[1], J. Gui[1], M. Melka[1], R. O'Reilly[2], S. Singh[1,2].* 1) Department of Biology, The University of Western Ontario, London, Ontario, Canada; 2) Department of Psychiatry, The University of Western Ontario, London, Ontario, Canada.

Despite their presumed identity, monozygotic twins often display discordance for a number of traits and disorders including schizophrenia. Attempts to identify causal genes for this common, devastating and often familial disease have remained challenging. Interestingly, the concordance for this disease in monozygotic twins is only 50%, suggesting that traditional genetic elements alone may not account for their origin and manifestation. To this end, we have proposed that the discordance for this disease in monozygotic twins may be due to ontogenetic de novo mutations. Further, the de novo events may include copy number variations, structural variants, indels and single nucleotide variants. In order to test this hypothesis, we have analyzed DNA extracted from blood samples of two pairs of monozygotic twins discordant for Schizophrenia and their parents using Complete Genome Sequencing.

The analysis of the data has revealed that the genomic sequence of monozygotic twins harbor unique variations at the DNA level. The best explanation for some of these differences includes de novo events. Also, a number of these de novo variants present in the affected twins are not shared with their co-twin and may contribute to their disease discordance. We have confirmed a number of such using Real Time PCR technology. These include a CNV loss at 7q11.21, and CNV gains at 15q11.2 and 12q24.21. In addition, a large-scale tandem duplication spanning from 1p36.31 to 1p36.33 which overlaps many genes including the GABRD that has been previously associated with schizophrenia in humans. Further, a number of Single Nucleotide Variants and small insertions and deletions have been identified to be unique between monozygotic twins and not found in parental samples. The results are compatible with the suggestion that the genetic differences between monozygotic twins may arise during ontogeny. More importantly, depending on the gene(s) affected these differences may play a role in the discordance of twins for diseases, including schizophrenia.

1227T

Association analysis of MAPT with cerebrospinal fluid tau using targeted sequencing data in older adults with mild cognitive impairment or Alzheimer's disease. *K. Deters[1], K. Nho[1], S. Kim[1], M.W. Weiner[3], T. Foroud[1], J.Q. Trojanowski[2], L.M. Shaw[2], R.C. Green[5], A.W. Toga[4], A.J. Saykin[1].* 1) Indiana University School of Medicine, Indianapolis, IN; 2) University of Pennsylvania School of Medicine, Philadelphia, PA; 3) University of California, San Francisco, San Francisco, CA; 4) University of Southern California, Los Angeles, CA; 5) Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

The microtubule associated protein tau (*MAPT*) gene encodes the protein tau, one of the main contributors to the pathogenesis of Alzheimer's disease (AD). SNPs of *MAPT* have previously been associated with AD and cerebrospinal fluid (CSF) levels of tau. However, most studies focused on only six haplotype-tagging SNPs of *MAPT*. While SNP genotyping covers only limited portions of the whole genome, whole-genome sequencing (WGS) has been used to obtain the most comprehensive genetic variation of an individual. Thus, we identified complete common variants (minor allele frequency (MAF) > 5%) of the MAPT gene from WGS data to investigate a comprehensive genetic association to CSF tau. WGS of 762 non-Hispanic Caucasian participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI) was used to identify common variants within *MAPT* resulting in 870 SNPs after standard quality control steps. Linear regression models were then used to investigate the association of *MAPT* common variants with CSF tau covarying for age, sex, and *APOE* status. Permutation testing was used to adjust for multiple-testing comparisons. Gene-based analysis of *MAPT* was significantly associated to CSF tau (p=0.03, corrected). *MAPT* SNPs rs117199550 (p=0.018, corrected) and rs63750072 (p=0.039, corrected) were significantly associated with CSF tau and were in linkage disequilibrium (r2=0.616, D'= 0.861). Rs117199550 is intronic and rs63750072 is located in an exon excluded from the major isoforms expressed in the brain but has been implicated in Frontotemporal dementia. Our results suggest that variance within *MAPT* is associated with CSF tau levels. To our knowledge, this is the first study to perform a comprehensive association analysis of *MAPT* variants with CSF tau using WGS. Our findings also suggest that WGS might enable the discovery of new variants associated with AD-related biomarkers.

1228S

Deep whole genome sequencing reveals multiple hits in non-coding sequence of autism risk genes. *F. Hormozdiari[1], M. Duyzend[1], B. Coe[1], J. Huddleston[1], D. Hanna[1], J. Smith[1], P. Sudmant[1], D. Nickerson[1], E. Eichler[1,2].* 1) Genome Sciences, University of Washington, Seattle, WA., Select a Country; 2) Howard Hughes Medical Institute, Seattle, WA.

Exome sequencing of thousands of children with autism spectrum disorder (ASD) has led to the discovery of hundreds of candidate genes based on the excess of disruptive mutations in when compared to unaffected siblings. The combination of de novo coding mutations and known pathogenic copy number variations (CNVs), however, still only accounts for a fraction of simplex ASD cases (~35% of the affected probands). In order to assess the importance of non-coding variation, we performed deep genome sequencing (>50 fold sequence coverage) on 16 autism trios where no de novo truncating mutation had been identified based on exome sequencing. We generated a comprehensive assessment of all types of genetic variation (SNVs, indels, CNVs and mobile element insertions) using a combination of tools (GATK, FreeBayes, Pindel, CommonLaw and GenomeStrip). As expected, we observed a linear correlation between paternal age and number of de novo SNVs per proband (r = 0.49). An average of 45 de novo SNVs and 1760 inherited deletions were identified per trio. In total, we characterized 322 private deletions (>200bp) of which 128 mapped within or near genes. 25% (4/16) of trios showed a pattern of multiple de novo mutations and private deletions in the non-coding region of genes previously implicated in autism and neurodevelopmental disease. One proband, for example, carried four de novo mutations in the intronic regions of *ARID1B*, *ROBO1* (x2) and *STXBP4*. Another proband inherited a 35 kbp and 150 kbp deletion of *MBD5* and *DISC1*, respectively, in addition to a de novo mutation within *ARID1B*. A third patient had a de novo mutation in the noncoding sequence of a metabotropic glutamate receptor, *GRM3*, in addition to three inherited deletions ranging in size from 12 to 120 kbp in autism risk genes *CACNA2D4*, *ARID1B* and *SCN2A*. The convergence of multiple rare non-coding mutations of a potentially deleterious nature in probands provides support for an oligogenic model of autism where the disruption of gene regulation in multiple targets contributes to disease etiology.

1229M

Exome sequencing in extended families with age-related macular degeneration reveals enrichment of genes involved in extracellular matrix pathway. *R. Priya[1], S. Perez[1], M. Mutsuddi[2], K. Branham[3], M. Othman[3], J. Heckenlively[3], A. Swaroop[1].* 1) Neurobiology-Neurodegeneration & Repair Laboratory (N-NRL), National Eye Institute, Bethseda, MD; 2) Department of Molecular and Human Genetics, Banaras Hindu University, Varanasi, India; 3) Department of Ophthalmology, University of Michigan, Ann Arbor, MI, USA.

Age-related macular degeneration (AMD) is a major cause of visual impairment in elderly population. Genome- wide association studies have identified 20 common susceptibility loci, which explain up to half of the AMD heritability. More recently targeted sequencing has helped identify penetrant, high-risk low-frequency coding variants at some of the known AMD loci. To further understand the genetic contributions of rare exonic variants, we undertook whole exome sequencing (WES) in 19 AMD families with 2-5 affected members. Exome of 69 samples from 19 families were captured using Agilent SureSelect Human All Exon kit and sequencing data was generated on Illumina GAIIx. Reads were aligned to the human reference genome (NCBI build 37.3) using Burrows-Wheeler Aligner, variants were called using GATK and annotations were done using ANNOVAR. We identified 343 rare segregating variants (MAF ≤ 0.05%) in these families. Subjecting this list to the pathway analysis using the database for annotation, visualization and integrated discovery (DAVID) recognized genes in extracellular matrix (ECM) as top functional clustering. ECM pathway along with complement and lipid metabolisms pathways have been associated with AMD based on GWAS studies. Thus it was interesting to see this association in exome sequencing analysis. Many of these candidates were highly expressed in human fetal (Retina, RPE and Choroid), aged and AMD (Retina) transcriptome data that are relevant in AMD pathophysiology. Taken together, our data shows that exome sequencing in extended AMD families can provide important insights into the disease etiology.

1230T

Large scale whole genome sequencing of Bipolar disorder 1 cases and controls in the BRIDGES study. *L.J. Scott[1], H.M. Kang[1], A.E. Locke[1], M. Flickinger[1], C. Quick[1], J. Carlson[1], D. Absher[2], Z. Chen[1], S. Vrieze[1], J. Li[3], S. Ramos[3], B. Li[3], J.L. Sobell[4], M. Rivera[5,6], L. Legrand[7], M. Burmeister[3,8,9], W.G. Iacono[7], M. McGue[7], M.T. Pato[4], J.A. Knowles[4], G. Abecasis[1], M. McInnis[8,10], S. Zoellner[1,8], G. Breen[5,6], C.N. Pato[4], J.B. Vincent[11], S. Levy[2], R.M. Myers[2], M. Boehnke[1].* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) HusdonAlpha Institute for Biotechnology, Huntsville, AL; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 4) Department of Psychiatry and the Behavioral Sciences, Zilkha Neurogenetics Institute, University of Southern California, Los Angeles, CA; 5) Medical Research Council Social, Genetic and Developmental Psychiatry Center, Institute for Psychiatry, King's College London, London, UK; 6) National Institute for Health Research Biomedical Research Centre for Mental Health at the Maudsley and Institute of Psychiatry, King's College London, London, UK; 7) Department of Psychology, University of Minnesota, Minneapolis, MN; 8) Department of Psychiatry, University of Michigan, Ann Arbor, MI; 9) Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI; 10) University of Michigan Depression Center, Ann Arbor, MI; 11) Center for Addiction and Mental Health. Department of Psychiatry, University of Toronto, ON, Canada.

Bipolar disorder 1 (BP I) is characterized by alternating periods of mania and depression and affects > 1% of individuals worldwide. Bipolar disorder is highly heritable and a substantial fraction of the heritability appears attributable to common variants. Genetic contributions from rare (< .5% minor allele frequency (MAF)) and low (.5 ≤ MAF < 5%) frequency variants are less well defined. Exome sequencing studies are exploring the contributions of coding variation, but in non-coding regions, rare and low frequency variants have been less well queried. Our aim is to ask if there are rare and low frequency non-coding or coding variants with strong effects on BP I risk. We are performing large-scale whole-genome sequencing (coverage ~10-12 x) of blood-based DNA from >3,770 European ancestry BP I cases and unrelated controls with no history of mental illness (BRIDGES study). Our current analysis contains 2,793 individuals: 1,382 BP I cases and 1,411 controls. We identified 43.4M SNPs using the GotCloud pipeline, of which 3.7M are low frequency, 12.2M rare, non-singleton and 21.4M singletons. We have 80% power (additive model) at an alpha of 5×10^{-8} to detect a single variant with 5, 1 or .5% MAF with odds ratios (OR) > 2, 3.6 or 5.3. Likewise, we have 80% power to detect at least one out of 10 variants with MAF of 5, 1 or .5% MAF for variants with OR > 1.7, 2.6, or 3.6, respectively. Using logistic analysis of single variants with adjustment for 10 principle components, we did not observe genome-wide significance association for any variant. We also did not observe multiple testing corrected significant gene-level associations of low-frequency (MAF < 1%) protein truncating and/or protein altering variation using SKAT-O or the CMC test, or from burden testing of deleterious coding singleton variants. Our current analyses suggest that, similar to other many other common, complex diseases, there are no or few moderately rare or low-frequency variants of large effects on risk for BP1. Larger sample sizes and sequencing-based studies will be necessary more fully query the contributions of BP1 risk variants from across the allelic spectrum.

1231S

Trio-Based Whole Genome Sequence Analysis of a Cousin Pair with Refractory Anorexia Nervosa. *P. Shih[1], A. Van Zeeland[2], A. Bergen[3], T. Carland[4,5], V. Bansal[1], P. Magistretti[6], W. Berrettini[7], W. Kaye[1], N. Schork[1,5].* 1) Dep of Psychiatry, Univ California, San Diego, La Jolla, CA; 2) Cypher Genomics, La Jolla, CA; 3) SRI, Palo Alto, CA; 4) The Scripps Research Institute, La Jolla, CA; 5) J. Craig Venter Institute, La Jolla, CA; 6) Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 7) University of Pennsylvania, Philadelphia, PA.

Anorexia Nervosa (AN) has an onset during adolescence and is characterized by emaciation, fear of gaining weight despite being underweight, and has the highest mortality rate of all psychiatric illnesses. Despite the serious health and psychosocial consequences of this illness, very few treatments are effective at reversing the core symptoms of AN. AN is highly heritable and show a homogeneous clinical presentation of persistent food refusal and high anxiety traits. However, AN etiology is believed to be heterogeneous as no major susceptibility gene has been consistently replicated in multiple populations. AN symptoms and personality traits tend to be present in unaffected family members of the patients, suggesting that certain shared genetic factors within each family may contribute to unique phenotype risk of the affected. To gain insights into the role "private variants" may play in AN and to maximize genetic information from family members of AN, here we leveraged a family-based study design combined with whole genome sequencing to search for genetic variants that may influence AN risk in an affected cousin pair together with their parents. By capitalizing on the homogeneity of the disease presentation among the two cousins, who both have a diagnosis of refractory AN, we report methods by which we interrogated shared chromosomal segments transmitted to them from their common grandparents that carried likely AN-related functional variants in this family.

1232M

Exome Association Study and 2nd SNP-GWAS of Japanese Parkinson's disease. *W. Satake[1], D. Shigemizu[2], Y. Suzuki[3], K. Yamamoto[4], H. Tomiyama[5], M. Yamamoto[6], M. Murata[7], N. Hattori[5], T. Tsunoda[2], M. Kubo[8], S. Tsuji[9], Y. Nakamura[10], S. Sugano[3], T. Toda[1].* 1) Div of Neurol/Mol Brain Sci, Kobe Univ, Japan; 2) Med Informatics, Cen for Genome Med, RIKEN, Japan; 3) Dept of Med Genome Sci, the Univ of Tokyo, Japan; 4) Kurume Univ, Japan; 5) Dept of Neurol, Juntendo Univ, Japan; 6) Takamatsu Neurol Clinic, Japan; 7) Dept of Neurol, Natl Cen Hsp of Neurol and Psy, Japan; 8) Genotyping Development, Cen for Genome Med, RIKEN, Japan; 9) Dept of Neurol, The Univ of Tokyo, Japan; 10) The Univ of Chicago, USA. The Univ of Tokyo, Japan.

Parkinson's disease (PD) is a complex disorder caused by multiple genetic variants, and one of the most common neurodegenerative disorders worldwide. We previously reported a Japanese SNP-GWAS which detected 4 PD-risk loci; *PARK16*, *BST1*, *α-synuclein*, and *LRRK2* (Satake *et al*, *Nature Genet* 2009). In order to search for further PD-risks in exonic areas, we performed whole exon (exome) sequencing of 755 PD patients using Sureselect and HiSeq2500. Average depth of our data is x 126, and 94.4 % of whole exon sequence was covered by 10 x or more reads. At first, using exome sequencing data of 625 PD cases and 259 controls, we tested association between PD and exonic SNVs within the 4 PD-loci (*PARK16*, *BST1*, *α-synuclein*, and *LRRK2*) reported by our previous SNP-GWAS. Genetic variants with strong PD-risk do not exist within these 4 PD-loci, indicating that these 4 PD-loci will contribute to this disease as common variants. We will subsequently test association between whole exonic SNVs and PD to identify novel PD-genes haboring rare-variant risks. Moreover, in parallel, to identify further common variant PD-risks, we performed Japanese 2nd SNP-GWAS which expanded our previous one. Using 1,948 cases and 28,990 controls, we identified a novel susceptibility locus with $P < 5 \times 10^{-8}$. Expression level of a gene within the locus is reduced when the risk SNP exists. In a fly model, knockout of the gene worsen moter function. Our genomic and *in-vivo* model data show that this gene is a novel PD-risk.

1233T

Identifying Genetic Variants Associated with Anorexia Nervosa via Exome Sequencing. *Q. Wei[1], E. Pruett[1], R. Adan[2], R. Cone[1], B. Li[1].* 1) Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 2) Department of Neuroscience and Pharmacology, Rudolf Magnus Institute of Neuroscience, University Medical Centre, Utrecht, The Netherlands.

Anorexia Nervosa (AN) is a psychiatric illness characterized by inability to maintain a minimal normal weight and persistent fear of gaining weight. It has the highest rate of mortality among psychiatric disorders and is estimated to affect 0.5%-1% of the population. Family and twin studies indicated that AN is highly heritable, and the estimated genetic contribution ranges 56-75%. However, the genetic etiology of AN is largely unknown. Previous genome-wide association studies assaying common variants failed to reveal genome-level significance. In this study, we carried out exome sequencing to identify rare variants associated with AN. We sequenced 24 AN patients and 47 of their relatives from pedigrees with high load of AN as well as 65 unrelated cases, and obtained 178 controls from National Database for Autism Research. In our analysis we focused on genes that are expressed in the prefrontal cortex of brain and functional variants (nonsynonymous, stop and splicing) that are deleterious predicted by Polyphen-2 and SIFT, LRT, MutationTaster, or CADD. Four genes (XIRP2, EXOG, MUSK, GRM6) harboring rare deleterious variants are enriched in AN patients with $p<0.0001$ in the single variant test. Furthermore, we carried out gene-based burden tests by collapsing multiple rare variants in a gene, and identified KIAA0317 with $p < 0.001$. We are exploring different grouping and weighting schemes for gene-based burden tests as well as pathway and network-based analyses to potentially increase power.

1234S

A Population-based Approach for Detecting Rare Recessive Variation Implicates the Cholesterol Biosynthesis Gene *DHCR24* in Autism Spectrum Disorder and Intellectual Disability. *E.T. Lim[1,2], Y. Chan[1,2], S. Goetze[1], D. Spatt[2], L. Kratz[3], M.B. Johnson[1,2], M. Chahrour[1,2], J.N. Hirschhorn[1,2], S. Raychaudhuri[2,4], J.M. Silverman[5], A. Kolevzon[5], J. Buxbaum[5], F. Winston[2], R.I. Kelley[3], M.J. Daly[2,6], C.A. Walsh[1,2], T.W. Yu[1].* 1) Boston Children's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Kennedy Krieger Institute, Baltimore, MD; 4) Brigham and Women's Hospital, Boston, MA; 5) Mount Sinai School of Medicine, New York, NY; 6) Massachusetts General Hospital, Boston, MA.

We developed and applied a statistical method (RAFT for Recessive Allele Frequency-based Test) for detecting rare recessive variants to a whole-exome genotyping dataset for ~1,000 families and discovered a rare missense variant (allele frequency = 0.5%) in *DHCR24*, a cholesterol synthesis gene, that is inherited recessively in 3 children from a single European American family affected by autism spectrum disorder (ASD). Subsequently, we discovered another 2 families of Middle Eastern ancestry with 6 children affected by ASD and intellectual disability with 2 other rare recessive missense variants in *DHCR24* (RAFT $P = 7.64 \times 10^{-10}$).

Recessive deleterious mutations in *DHCR24* have been associated with the cholesterol biosynthesis disorder desmosterolosis. But other than neurodevelopmental impairment, two of the three families did not exhibit additional distinctive features to suggest that diagnosis. Cholesterol is known to be important for the formation of myelin sheaths and membrane lipid rafts, and is a signaling molecule involved in critical developmental pathways that regulates skeletal and limb development. Recessive mutations in several genes that affect the precursors for cholesterol synthesis have been implicated in various neuro-developmental diseases, such as Smith-Lemli-Opitz syndrome, desmosterolosis and lathosterolosis. *DHCR24* is also named *Seladin-1* (Selective Alzheimer's Disease Indicator-1) based on initial discoveries that the mRNA expression of the gene was found to be down-regulated in individuals with Alzheimer's Disease.

To evaluate the cholesterol synthesis functionality of the missense variants identified in ASD patients, we adapted a protocol to express human *DHCR24* mutant proteins under the control of a GAL1 promoter in S. cerevisiae (Gilk et al., 2010) and optimized a biochemical assay for measuring desmosterol to cholesterol conversion (Waterham et al., 2001) to evaluate the pathogenicity of the missense variants detected in these families with autism and intellectual disability. Having validated the approach, we are now applying it to the mutations we have observed to test if the missense variants result in reduced cholesterol synthesis.

1235M

Mutational and transcriptional analysis in Autism Spectrum Disorders support their oligogenic model disturbing common functional pathways. *M. Codina-Solà [1, 2, 3], B. Rodriguez-Santiago[3, 4], J. Santoyo-Lopez[5, 6], A. Homs[1, 2, 3], M. Rigau[1], G. Aznar Lain[7], M. del Campo[8], B. Gener[9], E. Gabau[10], MP. Botella[11], A. Gutiérrez-Arumí[1, 2, 3], G. Antiñolo[12], LA. Pérez-Jurado[1, 2, 3], I. Cuscó[1, 2, 3].* 1) Universitat Pompeu Fabra, Barcelona, Barcelona, Spain; 2) Hospital del Mar Research Institute (IMIM), Barcelona, Spain; 3) Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Spain; 4) qGenomics, Barcelona, Spain; 5) Andalusian Human Genome Sequencing Centre (CASEGH); 6) Medical Genome Project (MGP) Sevilla, Spain; 7) Pediatric Neurology. Hospital del Mar. Barcelona; 8) Hospital Universitari Vall Hebron, Barcelona, Spain; 9) Hospital Universitario Cruces. Osakidetza, Spain; 10) Hospital Universitari Parc Taulí Sabadell, Barcelona, Spain; 11) Neuropediatria. HUA (Hospital Universitario de Alava), Alava, Spain; 12) Department of Genetics, Reproduction and Fetal Medicine, Institute of Biomedicine of Seville, University Hospital Virgen del Rocío/CSIC/University of Seville, Seville, Spain.

Background: Autism spectrum disorders (ASD) are a group of neurodevelopmental disorders with high heritability. There is high genetic heterogeneity and recent genetic findings support a double or multiple-hit model. **Methods**: We have performed Whole Exome Sequencing (WES) in 36 ASD patients to identify ASD causative mutations and rearrangements along with RNAseq to study the transcriptomic consequences. Pathway enrichment studies were performed by ConsensusPathDB. **Results**: We have detected likely causative mutations with monogenic models in 6 cases: 3 de novo loss-of-function mutations (genes *SCN2A*, *MED13L* and *KCNV1*), one de novo missense (*CUL3*) and two X-linked inherited mutations (genes *MAOA*, *CDKL5*). Inherited missense mutations in candidate genes, rarely mutated in controls, including *SCN2A* and *MED13L* were also found in the ASD cohort, suggesting the presence of ASD risk variants, contributing in an additive manner to the phenotype. Transcriptomic studies validated 90% of WES variants, identified deregulation (overexpression of *SEM6AB*, *MECP2*), Allele Specific Expression (*FUS*, *EP300*) and nonsense mediated decay (*RIT1*, *ALG9*) as a consequence of rare mutations. Pathway enrichment studies of inherited rare variants reveal the presence of recurrent alterations, including the previously implicated pathways PI3K-Akt signaling and Axon Guidance. **Conclusions**: WES has proven to be a highly efficient technology to identify the molecular defect in a proportion of probably monogenic ASD forms in our cohort (16.6%), as well as to detect potential risk variants contributing to the disorder in an additive oligogenic manner. Concomitant whole blood transcriptomic data allowed us to evaluate in part the functional consequences of genetic variants. Deregulation of common functional pathways by rare inherited variants such as PI3-Akt signaling pathway and Axon guidance could guide the identification of pharmacological treatments.

1236T

Whole-exome sequencing of multiplex families identifies several rare coding variants in known and novel Late-Onset Alzheimer genes. *B.W. Kunkle[1], M.A. Kohli[1], K.L. Hamilton[1], W.R. Perry[1], R.M. Carney[1], P.L. Whitehead[1], J.R. Gilbert[1], E.R. Martin[1], G.W. Beecham[1], J.L. Haines[2], M.A. Pericak-Vance[1].* 1) Hussman Institute for Human Genetics, University of Miami, Miami, FL; 2) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH.

We performed whole-exome sequencing (WES) on 28 multi-generational, late-onset Alzheimer disease (LOAD) families to identify rare LOAD risk variants. The pedigrees chosen for WES display a dominant inheritance pattern, have an average of eight Alzheimer-affected subjects, are free of known Alzheimer disease (AD) mutations, and do not cluster for the APOEe4 allele. 1-4 affected subjects per pedigree were sequenced in 22 families. Results for these families were combined with six previously reported pedigrees (4-9 affected subjects sequenced). Variants were filtered for segregating, nonsynonymous or splice-site rare variants (MAF<0.005) assuming autosomal or X-linked dominant models. Filtered variants were examined for implication as LOAD candidate genes by comparing results across families, and by comparison to GWAS-confirmed LOAD susceptibility genes and biologically relevant KEGG AD pathway genes. Six new rare missense variants in the genes PLEKHG5, ADAMTSL4, OR2L13, ZBTB11, CCDC39, and SEMA4B segregated with disease status in two or more different families. An additional family segregated a separate rare missense variant in PLEKHG5, and another family segregated a different rare missense mutation in ADAMTSL4. Three candidate genes (ABCG2, CD163L1, DMD) from our previously reported WES received additional support, and a new variant in the previously reported TTC3 gene segregated in two different pedigrees. Several of these genes have been functionally implicated in dementia-related processes and diseases (i.e. SEMA4B regulates axonal extension and synapse development; ABCG2 is involved in amyloid-β clearance across the blood brain-barrier; TTC3 is a Down syndrome critical region protein crucial for neuron survival), supporting their potential role in LOAD pathogenesis. In addition to these candidates, filtered variants also segregated with disease in the two LOAD GWAS genes (SORL1 and EPHA7), and several KEGG AD pathway genes (LPL, CACNA1D, RYR3, GRIN2A and NCSTN), making them important candidates for causal variants. In conclusion, WES of LOAD pedigrees identified several rare missense variants co-segregating with disease in multiple families and supported the involvement of rare variants in previously reported genes (TTC3, SORL1 and EPHA7) as being involved in AD risk.

1237S

Targeted resequencing of non coding functional DNA elements in autism. *D. Malhotra[1 6], A. Watts[1], T. Chapman[1], N. Plongthongkum[2], A. Gore[2], A. Fung[2], K. Zhang[2], J. Sebat[1 3 4 5].* 1) Department of Psychiatry, University of California San Diego, La Jolla, CA, USA; 2) Department of Bioengineering, University of California San Diego, La Jolla, CA; 3) Institute of Genomic Medicine, University of California San Diego, La Jolla, CA; 4) Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA; 5) Beyster Center for Genomics of Neuropsychiatric Diseases, University of California San Diego, La Jolla, CA; 6) Neuroscience Discovery and Translational Area, Pharma Research & Early Development (pRED), F. Hoffmann-La Roche Ltd, Basel, Switzerland.

While the role of rare de novo loss of function single nucleotide variants (SNVs) and copy number variants (CNVs) in autism is now firmly established, the genetic contribution of non-coding functional DNA elements to autism risk is unknown. We developed an inexpensive and high-throughput padlock probe based assay to resequence 5 megabase (Mb) of non-coding functional genome in 500 autism quad families from Simons Simplex Collection (SSC). We designed 77,277 padlock probes to capture the target region which mainly includes a) conserved mutation hotspots (CMH) in the human genome and, 2) putative enhancer and promoter elements of 95 high-risk autism genes. We performed a series of optimization experiments to develop an assay that a) captured 99.7% of target bases, b) had high sensitivity and specificity to detect SNVs confirmed empirically by testing the assay in two hapmap trios that were whole-genome sequenced at deep coverage by the 1000 genomes project and, c) had greater than 40X coverage per base per sample in a 384 sample multiplexed Illumina HiSeq sequencing run. Our experimental approach and study design will provide insights into the role of rare variation in the non-coding functional genome in risk for autism.

1238M

Identification of Rare Variants for Bipolar Disorder by Exome Sequencing in Multiplex Families. *S. Ramdas[1], J. Li[2], AB. Ozel[2].* 1) Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Bipolar Disorder is a heritable psychiatric disease characterized by alternating manic and depressive episodes. The phenotypic and genetic complexity in this disorder has made the identification of causal genes or biological pathways a considerable challenge. In this study, we focus on 34 multigeneration multiplex BPD families from the NIMH repository. Owing to the high occurrence of BPD in these families, we hypothesize that these families represent a nearly Mendelian subset of bipolar cases and transmit one or a few high-impact coding variants that result in the disorder. We genotyped and performed exome sequencing for 90 affected individuals in 34 families representing first cousin pairs or more distant relatives, aiming to detect functionally damaging rare variants in regions shared by affected relatives. Variant filtering were applied to identify those that are bi-allelic, have a minor allele frequency of less than 1% in European samples from the Exome Sequencing Project and 1000 Genomes, and are damaging missense, nonsense, or splicing site variants, resulting in a median of 508 variants per family. As exome chip genotype data allow inference of genomic segments shared among family members identically by descent (IBD), we used Beagle to identify IBD regions shared by BPD cases within each family. This led to a further reduction in the number of candidate variants, with a median of 147 variants. We found that HLA genes HLA-DRB1 and HLA-DRB5 showed damaging mutations in 8 and 7 families respectively. These genes have been previously reported to be associated with schizophrenia. Also found in our final list of genes are glutamate receptors GRIK3 and GRIN3B, and calcium channel receptor CACNA1C, all of which are known risk loci. Validation and segregation analyses using 254 additional members of the same families are underway to further prioritize the candidate genes and pathways (This study is supported by the IMHRO - Johnson & Johnson Rising Star Translational Research Award.).

1239T

A Cohort for Researching Autism Genetics in New Zealand. *B. Swan[1], J.C. Jacobsen[1], R. Hill[2], B. Tsang[3], M. Taylor[1], D. Love[4], J. Taylor[5], R.G. Snell[1], K. Lehnert[1].* 1) Centre for Brain Research and School of Biological Sciences, The University of Auckland, Auckland, New Zealand; 2) Department of Neurology, Auckland City Hospital, Auckland, New Zealand; 3) Paediatrics & Newborn Services, Waitemata District Health Board, Auckland, New Zealand; 4) Diagnostic Genetics, LabPLUS, Auckland City Hospital, Auckland, New Zealand; 5) Genetic Health Service New Zealand, Auckland City Hospital, Auckland, New Zealand.

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that presents with impaired social capabilities and restricted behavioural traits. It has a strong genetic aetiology and is observed in communities throughout the world. We have recently established a study investigating the genetic contributions to ASD in New Zealand, and to this end have developed a unique research cohort consisting of diverse populations living in New Zealand.

Recruitment of subjects was performed through collaboration with clinicians and an online project registry (www.mindsforminds.org.nz). Since August 2013, 899 participants have either registered themselves or been registered by their care-giver (232 self-registered, 667 registered by carer). Upon registration, participants can elect to enter information detailing age, sex, ethnicity, diagnosis, clinical specialists, prior DNA testing, and other relevant diagnoses for the participant or their family. Within this database we currently have an ASD research cohort that includes 684 diagnosed with ASD or Asperger's (Asp) (492 and 192, respectively). The male to female ratio of participants diagnosed with ASD is 5:1 (2.3:1 for Asp). Self-attributed ethnic background reflects New Zealand's population: 757 European (84.2%), 122 Maori (13.6%), 25 Polynesian (2.8%), and 35 Asian (3.9%) (individuals can register as more than one ethnicity). The 2006 census national population frequencies of these ethnicities in New Zealand is European 76.8%, Maori 14.9%, Polynesian 7.2%, and Asian 9.7%. The ASD/Asp cohort includes 119 multiplex families, 320 individuals with co-occurring neurological disorders [ASD:196; Asp:124], and 78 individuals with co-occurring gastrointestinal disorders [ASD:52; Asp:26]. This database is subsequently being used to identify participants for exome and genome sequencing. We are currently consenting individuals and undertaking exome sequencing on our first group of subjects with the goal of eventually sequencing the entire cohort.

1240S

Mutations in adaptor protein AP-5 subunits lead to peripheral neuropathy, spastic paraplegia and parkinsonism with aberrant endolysosomes. *M. Madeo[1], J. Edgar[2], F. Darios[3], T.N. Jepperson[1], J. Li[4], C. Blackstone[5], M.S. Robinson[2], J. Hirst[2], M.C. Kruer[1].* 1) Sanford Children's; 2) Cambridge University; 3) INSERM; 4) Vanderbilt; 5) NIH NINDS.

AP-5 is one of five adaptor protein complexes that play important roles in intracellular trafficking. We identified 7 patients with mutations in AP-5 subunits whose clinical phenotype ranged from isolated peripheral neuropathy to spastic paraplegia with parkinsonism. Primary cultured fibroblasts from affected patients showed accumulation of enlarged endolysosomes. Patient cells also displayed autofluorescence and accumulation of lysosomal storage material by electron microscopy. These findings indicate that AP-5 mutations lead to diverse neurological phenotypes and further support links between endosomal-lysosomal dysfunction and hereditary spastic paraplegia and parkinsonism.

1241M

Targeted sequencing of African American autism spectrum disorder patients reveals loss of function variants in novel autism genes. *P. Whitehead[1], A.J. Griswold[1], D. Van Booven[1], N. Dueker[1], J.A. Rantus[1], J.M. Jaworski[1], S.H. Slifer[1], M.A. Schmidt[1,2], W. Hulme[1], I. Konidari[1], M.L. Cuccaro[1,2], E.R. Martin[1,2], J.L. Haines[3], J.R. Gilbert[1,2], J.P. Hussman[4], M.A. Pericak-Vance[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL; 3) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 4) Hussman Institute for Autism, Baltimore, MD, USA.

Whole exome sequencing studies in autism spectrum disorder (ASD) have identified no single factor accounting for the genetic risk, though several candidate genes and molecular pathways have been implicated. These studies have been performed primarily in individuals of white European ancestry. However, the incidence of ASD is similar across ethnicities (1 in 68 children in whites, 1 in 81 in blacks) begging the question whether the risk genes identified are generalizable across races and ethnicities. We are currently investigating rare ASD risk variants by integrating GWAS with targeted massively parallel sequencing of candidate regions from GWAS Noise Reduction analyses of European ASD datasets (Hussman et. al., 2011). We targeted 689 GWAS-NR associated genes and evolutionarily conserved intronic and intergenic regions. We sequenced 140 ASD cases and 180 controls of African ancestry determined by principal component analysis using Eigenstrat. We identified 10,859 coding single nucleotide variants (SNVs) in exons of targeted genes. Among these exonic mutations are 5,544 nonsynonymous variants and 93 LOF alterations (46 stop gains, 6 stop losses, and 41 splice sites). LOF variation has been previously implicated as ASD risk factors in whites so we examined the genes affected by such alterations in our black cohort. There was no overall enrichment for LOF mutations in black ASD cases ($p = 0.553$) and only a single previously identified ASD candidate gene was affected by a LOF (MYO16). A total of 29 other genes had at least one LOF in a black ASD case and not in a control. These include novel candidate genes within pathways previously implicated in ASD. A stop mutation was found in a single case in KCNIP1 (voltage-gated potassium channel interacting protein 1) which modulates GABAergic signaling in mouse models. A stop in SRGAP3 (SLIT-ROBO Rho GTPase activating protein 3) was identified in another case and the Rho-GTP signaling gene has been implicated by copy number screens in intellectual disability. Finally, two mutations in the cytoskeletal structural gene IFLTD1 (intermediate filament tail domain containing 1) were identified in two separate cases. Overall, targeted sequencing of European ASD associated genes reveals loss of function variants in novel genes in black ASD cases, but in implicated ASD pathways. This suggests that novel genes in similar underlying pathways might contribute to the genetic risk of ASD in other ethnicities.

1242T

Exome sequencing of 43 sporadic cases with an autism spectrum disorder in a local cohort of families identifies severe *de novo* variants and implicates additional genes in ASD pathogenesis. *W. Banks[1], D. Cunningham[1], E. Hansen[1], K. Ratliff-Schaub[3], E. Butter[3], D. Schulteis[4], C. Boreman[4], B. Kelly[2], P. White[2], G. Herman[1,3].* 1) The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) Center for Microbial Pathogenesis, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 3) Department of Pediatrics, Nationwide Children's Hospital, Columbus, OH; 4) Department of Pediatrics, Wright-Patterson Air Force Base, OH.

Autism spectrum disorders (ASD) are a common group of complex neurodevelopmental disorders, with over 500 CNVs and genes implicated in ASD pathogenesis to date. To identify additional loci involved in ASD, we undertook whole exome sequencing (WES) of 43 simplex trios from the Central Ohio Registry for Autism (CORA), a local cohort of over 200 families. Psychological testing records including ADOS and IQ, a detailed medical history, clinical genetic testing at a minimum including microarray and Fragile X DNA, and a 3-generation pedigree are gathered on all affected individuals. Families were chosen for WES based on a review of psychological and medical testing records, as well as the absence of ASD or related psychological disorders in the extended pedigrees. Exome captured libraries were prepared using Agilent Human All Exon V4/V5 kits and sequenced on an Illumina HiSeq2500. A locally developed analysis pipeline (Churchill) was used to identify statistically supported variants from the raw sequence data. The presence of putative *de novo* variants in the proband, and their absence in both parents, was validated via Sanger sequencing.

To evaluate the possible significance of coding variants identified, we analyzed 1) the predicted effect on protein structure by SIFT and PolyPhen2; 2) the expression profile of the affected gene in the developing CNS (BrainSpan.org); 3) phenotypes identified in mouse models; and 4) previous associations of the gene with human neurodevelopmental phenotypes, including intellectual disability (ID) and ASD. Severe *de novo* variants were confirmed in three genes (*LRP1*, *KDM5B*, and *IQSEC2*) previously found to harbor *de novo* variants in cases of ASD or ID. In addition, developmental expression patterns and/or mouse models displaying abnormal neurologic or behavioral phenotypes have been described for several other genes identified, supporting their potential involvement. Further studies in larger ASD cohorts, as well as functional studies in *in vitro* and/or *in vivo* models may further help clarify the role of these candidate loci in disease pathogenesis.

1243S

De novo and rare inherited mutations implicate the transcriptional coregulator TCF20/SPBP in autism spectrum disorder. *A.O.M. Wilkie[1,3], C. Babbs[1], D. Lloyd[1], A.T. Pagnamenta[2], S.R.F. Twigg[1], J. Green[1], S. McGowan[1], G. Mirza[2], R. Naples[1], V.P. Sharma[1,3], E.V. Volpi[2], V. Buckle[1], S.A. Wall[3], S.J.L. Knight[1], J.R. Parr[4].* 1) Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 3) Craniofacial Unit, Oxford Radcliffe Hospitals NHS Trust, John Radcliffe Hospital, Oxford, United Kingdom; 4) Institute of Neuroscience, Newcastle University, Newcastle Upon Tyne, United Kingdom.

Background: Autism spectrum disorders (ASDs) are common and have a strong genetic basis, yet the cause of ~70-80% ASDs remains unknown. By clinical cytogenetic testing we identified a family in which two brothers had ASD, mild intellectual disability and a chromosome 22 pericentric inversion, not detected in either parent, indicating *de novo* mutation with parental germinal mosaicism. We hypothesised that the rearrangement was causative of their ASD and localised the chromosome 22 breakpoints. **Methods:** The rearrangement was characterised using fluorescence in situ hybridisation, Southern blotting, inverse PCR and dideoxy-sequencing. Open reading frames and intron/exon boundaries of the two physically disrupted genes identified, *TCF20* and *TNRC6B*, were sequenced in 342 families (260 multiplex and 82 simplex) ascertained by the International Molecular Genetic Study of Autism Consortium (IMGSAC). **Results:** IMGSAC family screening identified a *de novo* missense mutation of *TCF20* in a single case and significant association of a different missense mutation of *TCF20* with ASD in 3 further families. Through exome sequencing in another project we independently identified a *de novo* frameshifting mutation of *TCF20* in a woman with ASD and moderate intellectual disability. We did not identify a significant association of *TNRC6B* mutations with ASD. **Conclusions:** *TCF20* encodes a transcriptional coregulator (also termed SPBP) that is structurally and functionally related to a single paralogous gene, *RAI1*. Notably, *RAI1* encodes the critical dosage-sensitive protein implicated in the behavioural phenotypes of the Smith-Magenis and Potocki-Lupski 17p11.2 deletion/duplication syndromes, in which ASD is frequently diagnosed. Our study provides the first evidence that mutations in *TCF20* are also associated with ASD and should encourage further investigation of the normal functions of TCF20/SPBP in neurodevelopment and the role of mutations in ASD.

1244M

***MEF2C* haploinsufficiency is a recurrent finding in patients with autism spectrum disorders.** *A. Ziegler[1,2], R. Delorme[2,3,4], P. Bonneau[1,6], A. Guilmatre[2,3], T. Bourgeron[2,3,5], D. Bonneau[1,6].* 1) Department of Biochemistry and Genetics, CHU Angers, Angers, France; 2) Human Genetics and Cognitive Functions, Institut Pasteur, Paris, France; 3) CNRS URA 2182 "Genes, synapses and cognition," Institut Pasteur, Paris, France; 4) Assistance Publique-Hôpitaux de Paris, Robert Debré Hospital, Department of Child and Adolescent Psychiatry, Paris, France; 5) University Denis Diderot Paris 7, Paris, France; 6) UMR CNRS 6214 INSERM 1083, Angers, France.

Since the first description of *MEF2C* haploinsufficiency syndrome in 2009, 42 patients have been reported either with deletion or point mutation of this gene. Intellectual disability (ID) is the core disorder of this syndrome but autistic features such as stereotypic movements and lack of social communications are commonly reported. To further assess the role of *MEF2C* in autism spectrum disorder (ASD), we looked for *MEF2C* point mutations by Sanger sequencing and for MEF2C copy number variations by SNP-array in a cohort of 195 patients with ASD and a mild to severe ID. We identified a de novo frameshift mutation leading to a premature stop codon in 2 siblings with ASD and severe intellectual disability. A maternal germinal mosaicism was confirmed based on the haplotype. A de novo *MEF2C* deletion was also found in a patient with ASD and a mild intellectual disability. In this study, we observed *MEF2C* haploinsufficiency in 1,5% (95% confidence interval: 0- 3.2%) of patients with ASD. This is the first study to look specifically at *MEF2C* in ASD. Of note, a mutation in MEF2C was also found once by whole-exome sequencing in another cohort of 175 trios with ASD. According to this notable *MEF2C* haploinsufficiency frequency in ASD, we strongly recommend to pay specific attention to this gene when performing non-targeted genetic screening in ASD with a comorbid ID even for patients without severe ID.

1245T

Assessing the role of methylation in autism brains. *S.E. Ellis[1], S. Gupta[1], A. Moes[1], A.B. West[2], D.E. Arking[1].* 1) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Department of Neurology, University of Alabama School of Medicine, Birmingham, AL.

The high prevalence of autism spectrum disorder (ASD) and very limited understanding of relevant risk factors have sparked an intense focus on elucidating the genetic basis of ASD. GWAS, exome-sequencing, and RNA-Seq have all helped the field move forward, but the vast majority of the risk remains unexplained. Given that prenatal brain development is both critical to the progression of ASD and a period in which development is sensitive to alterations in epigenetic pathways, many groups have turned to investigating the role of methylation in ASD. To date, microarray studies have looked for methylation differences between ASD brains and controls; however, these have been limited by the digital nature of methylation arrays and small sample sizes. We report results from the largest study to date, relying on next-generation methylation sequencing (reduced representation bisulfite sequencing) in 63 cortical brain samples (BA19), comprising 34 controls and 29 individuals with autism, a subset of ASD, all of whom were diagnosed using ADOS and ADI-R (among other exclusions). After extensive quality control, we performed a single site analysis across 739,980 CpGs looking for methylation differences between cases and controls. This analysis identified 28 differentially methylated CpGs ($q<0.05$), eight of which demonstrate a mean methylation difference greater than 10 percent (5 hyper- and 3 hypo-methylated in ASD). While none of these 28 sites fall in regions demonstrating differential gene expression (as measured by RNA-Seq), six are exonic, nine fall within promoters, 16 are within CpG Islands (CGI), and five are within CGI shores, suggesting a potential functional role for these individual sites. In particular, one exonic CpG site in SHANK3 demonstrates a 6.4% decrease in methylation in ASD cases. While the need to validate this finding by an independent method remains, this finding is of particular interest for future investigation as SHANK3 has frequently been reported to have a role in ASD. Finally, ongoing studies will leverage these data, along with RNA-seq carried out on these same individuals, to gain a more complete understanding of the relationship between methylation and gene expression in the prefrontal cortex with risk for ASD.

1246S

Methylation pathway and chromatin modification in autism. *M. Smith, P. Flodman, J. Bevilacqua, J-J. Gargus.* Dept Pediatrics, Univ California, Irvine, Irvine, CA.

We analyzed data from exome sequencing in 9 probands with autism. In each of the autistic probands our analyses revealed the presence of rare and low frequency potentially functionally damaging variants in genes that encode enzymes involved in synthesis of methyl residues and/ or genes that encode enzymes involved in the transfer of methyl residues to or from chromatin, (histones and DNA). A rare allele in lysine specific methyltransferase KMT2C variant rs10454320 occurred in 3 of the 9 autistic probands; a rare allele at KMT2C variant rs78352960 occurred in 2 autistic probands. A rare allele KMT2C variant at rs200804156 occurred in 1 autistic proband. It is interesting to note that one proband had 3 rare variants in KMT2C and one rare variant in KMT2D lysine specific methyltransferase rs201114196. Low frequency potentially damaging variants in the MTRR gene 5 methyltetrahydrofolate homocysteine methyltransferase reductase rs2287780 occurred in 2 autistic probands. An autistic twin pair had a rare potentially damaging variant in MTHFS 5-10 methenyltetrahydrofolate synthetase. The variants described were potentially damaging and were inherited from one or other of the parents where testing of family was possible (5 out of 7 families) and were present in heterozygous state in the probands. We identified rare potentially damaging variants in genes that encode proteins involved in chromatin modification. Rare potentially damaging variants, each identified in 1 proband occurred in KDM5A lysine specific demethylase at rs200804533, in KAT2B lysine acetyl transferase at rs148960024, in CHD8 chromatin helicase at rs2000465274 in CHD6 chromatin helicase at rs61752057. Studies by Melnyk, James and coworkers have implicated methylation dysregulation in autism PMID: 21519954. There is increasing evidence that abnormalities of chromatin modification are frequent cause of cognitive disorders in children, (Ronan and Crabtree PMID: 23568486). It is important to consider the possibility that rare or unusual variants and perhaps even common variants in a specific pathway may play roles in the etiology of autism and that de novo and inherited variants may have impact.

1247M

Mutation Screening in Saudi Parkinson's Disease Patients. *B.R. Al-Mubarak[1], A. Magrashi[1], M. AlTurki[1], S. Boholega[2], T. Alkhairallah[2], B. AlTawil[1], H. Abou Al-Shaar[1], L. Al-Jomaa[1], N. Al Tassan[1].* 1) Genetics, King Faisal Specialist Hospital, Riyadh, Saudi Arabia; 2) Neurosciences, King Faisal Specialist Hospital, Riyadh, Saudi Arabia.

Parkinson's disease (PD) is classified as the second most prevalent aging-related neurodegenerative disease after Alzheimer's, and is predicted to affect approximately 8-9 million individuals by 2030. The disease is characterized by the occurrence of four cardinal symptoms: bradykinesia, resting tremor, rigidity and postural imbalance. In most cases the disease occurs sporadically with complex etiopathology involving a combination of genetic propensity and environmental factors. However, familial cases also exist accounting for 15-20% of the patients. Mutations in *PARKIN*, *SNCA*, *DJ-1*, *PINK1*, and *LARRK2* genes have been reported in patients with different modes of inherence of PD. In this study we sought to determine the genetic basis underlying PD in 97 Saudi individuals affected with either sporadic or familial form of the disease through screening the entire ORF region in addition to exon-intron junctions of known PD-causing genes for possible disease-related variants. Direct sequencing of *PARKIN*, *SNCA*, *DJ-1*, *PINK1* and *LARRK2* genes in these samples failed to reveal pathogenic mutations with the exception of one familial case homozygous for c.938 C>T (p.Thr313Met) mutation in *PINK1*. Moreover, the mutation analysis was extended to include more recently identified PD susceptibility genes particularly, *FBXO7* , *GIGYF2* , *VPS35* , and *UCHL1* . Out of the 121 detected SNVs, only few were potential disease-causing mutations, while the majority was either present in ethnically matched normal controls or did not segregate with the mode of inheritance. The absence of pathogenic mutations in the screened genes in our cohort doesn't rule out the possibility of their involvement in the development of the disease, as gene expression maybe perturbed. However, it also suggests that other genetic factors may be implicated in these patients. These findings are in keeping with the long-held notion of the complexity and genetic heterogeneity of PD.

1248T

Genetics of dementias in a Turkish cohort. *R. Guerreiro[1,2], G. Guven[3], J. Bras[1], L. Darwent[1], J.R. Gibbs[1,2], N. Unaltuna[3], H. Gurvit[4], B. Bilgic[4], H. Hanagasi[4], M. Emre[4], A. Singleton[2], J. Hardy[1], E. Lohmann[4,5].* 1) Department of Molecular Neuroscience, UCL Institute of Neurology, University College London, Queen Square, London, UK; 2) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA; 3) Department of Genetics, Institute for Experimental Medicine, Istanbul University, Istanbul, TR; 4) Behavioral Neurology and Movement Disorders Unit, Department of Neurology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, TR; 5) Department Of Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, GER.

Turkey has a population of over 77 million, with a growth rate of 1.5% per year and an average life expectancy of 73.2 years. Large family units, high birth rates, and high rates of consanguineous marriages can often be found in this population, making genetic diseases a common health problem. In parallel, the improvement of living standards and health services will increase survival rates and, consequently, the incidence of Alzheimer's disease (AD [MIM 104300]) and other dementias. The current study is the first large molecular study of dementias in Turkey. Our main goals were: 1) to establish the frequency of mutations in the most common genes known to cause dementias (APP, PSEN1, PSEN2, MAPT, GRN and C9orf72); 2) to characterize the mutation spectrum in these genes by describing genotype-phenotype correlations; 3) to study recessive forms of dementia; 4) to identify novel genes/variants involved in different forms of dementia. We have established a cohort of 150 families presenting with different forms of dementias that we have been studying by using genome-wide genotyping techniques and next generation sequencing combined with Sanger sequencing. We established a frequency of 11.2% of mutations and variants in APP, PSEN1 and PSEN2, and of 5.2% of pathogenic mutations in MAPT, GRN and C9orf72, suggesting that mutations in these genes are not uncommon in Turkey. The application of exome sequencing to the cases without mutations in the known dementia genes revealed some unexpected results with the identification of a previously reported NOTCH3 mutation in an AD family and three homozygous TREM2 mutations in Frontotemporal dementia-like (FTD [MIM 600274]) families. NOTCH3 mutations have been previously associated with Cerebral arteriopathy, autosomal dominant, with subcortical infarcts and leukoencephalopathy (CADASIL [MIM 125310]) while TREM2 mutations are known to cause Polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy (PLOSL [MIM 221770]). Exome sequencing is a valid, rapid and cost-effective tool to identify genetic mutations in complex diseases like dementias, and is particularly useful in the study of autosomal recessive forms of disease when coupled with loss-of-heterozygosity analyses. The application of exome sequencing to neurological syndromes has revealed several unexpected phenotypes for mutations in known disease genes, that otherwise would have never been screened for, given the atypical presentation.

1249S

Mutation screening in exon 2 of synaptic gene SHANK3 in Brazilian individuals with Autism Spectrum Disorder. *D.BA. Rosan[1], P.P NASCIMENTO[1], A.L BOSSOLANI-MARTINS[2], L.A BRITO[3], M.R PASSOS-BUENO[3], A.C FETT-CONTE[4].* 1) Universidade Estadual Paulista Junior de Mesquita Filho; 2265 Rua Cristóvão Colombo, Jardim Nazareth, 15054-000, São José do Rio Preto, SP, Brazil; 2) Universidade Federal do Mato Grosso do Sul Campus Paranaiba, 800 Rua Maria Antonia - 79500-000, Paranaíba, MS, Brazil; 3) Universidade de São Paulo, instituto de Biociencias, 14 Rua do Matão, - Butantã, 05508-090, São Paulo - SP, Brazil; 4) Faculdade de Medicina de São José do Rio Preto, 5416 Av. Brigadeiro Faria Lima, Vila São Pedro, 15090-000 - São José do Rio Preto - SP - Brazil.

Austism Spectrum Disorder (ASD) is a complex behavioral disorder characterized by manifestations usually noted before the age of three and that affect the field of social communication and behavioral domain, atypical behaviors, restricted and repetitive interests, compromising the entire life of the individual. With a complex etiology, numerous genetics, epigenetics and environmental factors can explain this changes that affect 1-110 individuals with males being affected four times more than females. Large numbers of genes have been proposed and investigated in the predisposition to ASD. Many of them are expressed in the development of the adult brain, such as those responsible for encoding proteins involved in the formation and function of synapses. Among the major genes are SHANK family, which act particularly in post-synaptic region and are composed by three genes, SHANK1, SHANK2 and SHANK3 that are expressed in different regions of the brain encoding scaffolding proteins of the postsynaptic synapses. All these genes have been described as candidates for autism by the significant association of mutations with the behavioral phenotype and intellectual deficits. The SHANK3 gene, the most studied SHANK of the family, is located on 22q13.3 region and is predominantly expressed in the cerebral cortex and cerebellum. The exon 2 of SHANK3 gene has never been studied in a brazilian individuals with idiopathic ASD and this study aimed to evaluete the presence of mutations in this exon to investigate a possible association with the behavioral phenotype. We sequenced 190 Brazilian individuals and found a heterozygous alteration of A/G at the position 22: 50675139 resulting in a substitution of the amino acid threonine by an alanine at codon 42. The change was found in a 21 years old male patient with diagnosis of autism without phenotypic changes. Karyotypic investigation and molecular testing to Fragile-X syndrome were normal. We searched for this mutation in the "1000 Genomes- Deep Catalog of Human Genetic Variation" and in "NHLBI Exome Sequencing Project (ESP) Exome Variant Server" and nothing was found. Genetic studies have demonstrated the strong association of mutations in SHANK3 with susceptibility to psychiatric disorders, particularly autism, however, little is known about the neural consequences of these mutations and further investigations should be carried out to clarify the causes of this disease.

1250M

Variations in hotspot region of β-amyloid precursor protein (APP) gene in various neurological disorders from Hyderabad, a cosmopolitan city of South India. *W. Thomas[1], S. Divyakolu[1], VR. Sreekanth[2], VR. Om Sai[3], V. Nagaratna[3], Q. Hasan[1, 4], YR. Ahuja[1].* 1) Department of Genetics and Molecular Medicine, Vasavi Medical and Research Centre, Hyderabad, Andhra Pradesh, India; 2) Department of Neurology, Apollo Hospital, Jubilee Hills, Hyderabad, Andhra Pradesh, India; 3) Department of Medical Sciences, National Institute of Mentally Handicapped, Bowenpally, Secunderabad, Andhra Pradesh, India; 4) Department of Genetics and Molecular Medicine, Kamineni Hospital, LB Nagar, Hyderabad, Andhra Pradesh, India.

Background: Non-synonymous mutations/ polymorphism in amyloid precursor protein (APP) gene causes overproduction of Aβ proteins or affects its split into Aβ40 and Aβ42 peptides. Aβ42 has been considered to be a toxic peptide playing a major role in the pathogenesis of Alzheimers (AD). Similar APP plaques were observed in the brains of Down syndrome (DS) patients and high level of plasma APP was observed in patients with severe Autism spectrum Disorder (ASD). The aim of this study was to evaluate exon 16 and 17, the hotspot regions of APP gene in patients with neurobehavioral disorders like AD, DS and ASD.

Methods: A total of 75 cases were recruited in the study which included AD (n=25), DS (n=25), and ASD (n=25). Polymerase chain reaction (PCR) analysis and sequencing was carried out using exon-intron encompassing primers for the selected APP gene regions. In-silico analysis was also carried out to identify the impact of sequence variants on the protein structure.

Results: Three exonic variants, two in exon 16: V683V, H684Y and one in exon 17, H733Q were identified in sporadic AD cases. Apart from these, two intronic variants were also observed. In-silico analysis showed that H733Q mutation may affect the structure and function of APP, whereas H684Y mutation is neutral. In an ASD case, our analysis showed an intronic variation ie., an A insertion at c.1964-13_1964-12insA. In-silico analysis predicted that this variation affects the elongation feature of the transcript. None of the DS cases had any variation in this hotspot region.

Conclusion: Our data indicate that variations in the selected hotspot region of APP may play an important role in the aetiology of neurobehavioural disorders.

1251T

DYT16 revisited: exome sequencing identifies PRKRA mutations in a European dystonia family. *M. Zech[1, 2], F. Castrop[1], B. Schormair[1, 2], A. Jochim[1], T. Wieland[2], N. Gross[1], P. Lichtner[2], A. Peters[2], C. Gieger[2], T. Meitinger[1, 2], T. Strom[1, 2], K. Oexle[1], B. Haslinger[1], J. Winkelmann[1, 2, 3].* 1) Technical University of Munich, Munich, Germany, Munich, Germany; 2) Helmholtz Zentrum München, Munich, Germany; 3) Stanford University School of Medicine, Palo Alto, CA, USA.

Objective: Recessive DYT16 dystonia associated with mutations in PRKRA has until now been reported only in seven Brazilian patients. The aim of this study was to elucidate the genetic cause underlying disease in a Polish family with autosomal-recessive, early-onset generalized dystonia and slight parkinsonism, and to explore further the role of PRKRA in a dystonia series of European ancestry. Methods: We employed whole-exome sequencing in two affected siblings of the Polish family and filtered for rare homozygous and compound heterozygous variants shared by both exomes. Validation of the identified variants as well as homozygosity screening and copy number variation analysis was carried out in the two affected individuals and their healthy siblings. PRKRA was analyzed in 339 German patients with various forms of dystonia and 376 population-based controls by direct sequencing or high-resolution melting. Results: The previously described homozygous p.Pro222Leu mutation in PRKRA was found to segregate with the disease in the studied family, contained in a 1.2 Mb homozygous region identical by state with all Brazilian patients in chromosome 2q31.2. The clinical presentation with young-onset, progressive generalized dystonia and mild parkinsonism resembled the phenotype of the original DYT16 cases. PRKRA mutational screening in additional dystonia samples revealed three novel heterozygous changes (p.Thr34Ser, p.Asn102Ser, c.-14A>G), each in a single subject with focal/segmental dystonia. Conclusions: Our study provides the first independent replication of the DYT16 locus at 2q31.2 and strongly confirms the causal contribution of the PRKRA gene to DYT16. Our data suggest worldwide involvement of PRKRA in dystonia.

1252S

Rapid multiplex sequencing of genes associated with progressive neurodegenerative disorders. *M.O. Dorschner[1,2,3], M.A. Weaver[1,2,3], G.L. Carvill[5], H.C. Mefford[5], I.F. Mata[3,4], C.P. Zabetian[3,4], M. Rumbaugh[3,4], T.D. Bird[3,4], D.W. Tsuang[2,3].* 1) Department of Pathology, University of Washington, Seattle, WA; 2) Department of Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA; 3) Geriatric Research, Education, and Clinical Center, VA Puget Sound Health Care System, Seattle; 4) Department of Neurology, University of Washington, Seattle, WA; 5) Department of Pediatrics, University of Washington, Seattle, WA.

Alzheimer's disease (AD) is an insidious neurodegenerative disorder that accounts for significant morbidity and mortality. While there are multiple causes of dementia, including dementia with Lewy bodies (DLB), frontotemporal dementia (FTD) and Parkinson's disease dementia (PDD), greater than half of cases are due to Alzheimer's disease. Dementia manifests clinically with subtle and poorly recognized cognitive deficits and slowly becomes more severe and eventually incapacitating. An accurate diagnosis early in the disease process would greatly improve the ability of physicians to predict disease progression and manifestation of specific symptoms. We developed a multigene sequencing panel for several purposes: 1) to establish an assay for clinical diagnostic testing; 2) to select research subjects for discovery-based efforts prior to exome or genome sequencing, and 3) to identify rare disease-causing variants and study the type and distribution of variants with respect to phenotypic characteristics. The coding regions of these genes were captured with gap-fill molecular inversion probes (MIPs) and sequenced using Illumina technology. Sequence data were aligned with BWA and variants called with GATK. Several genes proved difficult to capture due to GC content and sequence context. To improve the capture of these genes, we made several modifications to the protocol and separated the poorly performing MIPs from those that were working efficiently. Two captures were performed for each sample and combined after amplification. With 928 MIPs we were able to obtain adequate coverage for > 95% of the target when sequenced in pools of 192 subjects. To test the ability of our panel to detect disease-causing variants we assayed a set of positive controls with 11 known mutations. 9 of the 11 known mutations were detected. The two failures were located in the *PRNP* gene, in a difficult to capture segment. The assay was tested with 53 individuals clinically diagnosed with a variety of neurodegenerative disorders for which the genetic etiology was unknown. Six individuals carried predicted pathogenic, likely pathogenic, or risk variants. The combination of molecular inversion probe enrichment and next generation sequencing provides a rapid, low cost procedure for screening large numbers of individuals for neurodegenerative disorders. We are currently expanding and deploying this assay to screen additional populations with dementia and Parkinson's disease.

1253M

Genome sequencing in X-Linked Ataxia Dementia. *J.L. Farlow[1], K. Hetrick[2], H. Ling[2], B. Craig[2], M. Farlow[1], E. Pugh[2], K. Doheny[2], T. Foroud[1].* 1) Medical and Molecular Genetics, Indiana University School of Medicine, 410 West 10th Street, HS 4000, Indianapolis, IN 46202; 2) Center for Inherited Disease Research, Johns Hopkins University, 333 Cassell Drive, Suite 2000, Baltimore, MD 21224.

X-Linked Ataxia Dementia (XLAD), also known as X-linked spinocerebellar ataxia type 4, is a neurodegenerative disorder characterized by ataxia in childhood, variable onset upper motor neuron disease, and adult-onset dementia. In the one kindred described with this syndrome, moderate phenotypic variability is observed in affected males, and carrier females show a milder phenotype that includes cognitive and motor abnormalities. Previous linkage studies demonstrated a LOD score of 5.29 in the region Xq21.33-q23. Further genotyping on two distantly related affected males and two unaffected males on an Illumina Omni1-Quad array narrowed this region to 19Mb and ruled out a large copy number variant. Whole genome sequencing (30X) was performed on an obligate carrier female and her affected son to identify the causative mutation in this region. The Agilent SureSelectXT2 Library Prep Kit and Illumina HiSeq2000 were used to generate 100bp paired end sequencing data. Paired end alignment was performed to the GRCh37 reference genome with Burrows-Wheeler Aligner, and duplicates were marked using Picard. The Genome Analysis Toolkit was used for insertion/deletion (indel) realignment, base call quality score recalibration, multi-sample variant calling (Unified Genotyper), and Variant Quality Score Recalibration. Variants were annotated by ANNOVAR and the recently developed Combined Annotation-Dependent Depletion program. Variants of interest were visually inspected using the Integrated Genomics Viewer. The average transition/transversion ratio for exonic variants and all variants was 3.10 and 2.12 respectively, and the percentage of variants found in dbSNP137 was 98.98%. There were 7,901 variants identified in the region of interest. Of these, 5,798 were heterozygous in the mother and hemizygous in the son, and 505 of these were not observed in 1000 Genomes, the Exome Sequencing Project, and dbSNP137. Variants were retained if they were in an exonic, splicing, UTR, or regulatory region. Using these criteria, 22 variants (3 single nucleotide variants and 19 indels) were retained, all of which were in UTR or regulatory regions. Genome sequencing has not identified a clearly causative exonic variant for this region. Further work remains to look for structural variants such as larger insertions or deletions and inversions. For all variants identified, challenges remain to predict the resulting functional impact and to prioritize variants for molecular characterization.

1254T

Prion disease with chronic diarrhea associated with PRNP mutation Q160X has reduced penetrance. *J.C. Fong[1], J. Bang[1], A. Legati[2], K.J. Rankin[1], N. Block[1], R. Ketelle[1], A.M. Karydas[1], G. Coppola[2], M.D. Geschwind[1].* 1) Dept of Neurology, UCSF Memory and Aging Center, San Francisco, CA; 2) Depts of Psychiatry and Neurology, UCLA, Los Angeles, CA.

Background: Prion diseases are neurodegenerative disorders occurring in genetic, sporadic, or acquired forms. Genetic prion disease is caused by mutations in the prion protein gene (PRNP)—missense, nonsense, insertions, deletions—each associated with varying symptoms, including mutations characterized by diarrhea and dysautonomia. To date, nonsense PRNP mutations have not been associated with reduced penetrance. We report a rare nonsense PRNP mutation in an index case, diagnosed initially with an unspecified dementia, and his unaffected father. Methods: The proband was evaluated clinically. Whole exome sequencing or WES (Illumina) was done on the proband and both parents. Reads were mapped to GRCh37 using BWA. Variants were called with GATK, then annotated and filtered using the Ingenuity platform (Qiagen). Results: The proband presented at age 30 with a 5-yr history of chronic diarrhea, followed by behavior, language, memory problems, and recent movement, gait, and swallowing difficulty. Symptoms began at age 25 with alternating diarrhea/constipation. At Yr1, he had trouble managing finances, academic problems, and agitation. By Yr2, he dropped out of college and was fired from his job for forgetting to complete tasks. He was fired from a 2nd job for making violent threats against coworkers. Between Yrs4-4.5, speech and language declined. By Yr5, he required help with activities of daily living. Exam at Yr5 showed severe grammar deficits with intact semantic knowledge, mild motor deficits, and gait ataxia. MRI showed profound, diffuse bilateral cortical and white matter atrophy without restricted diffusion. Paternal FHx was significant for an autosomal dominant pattern of dementia (clinically Alzheimer's or AD) preceded by diarrhea in 5 relatives, including a great-aunt (died at age 51) whose autopsy showed plaques and tangles. WES identified a PRNP Q160X mutation with methionine homozygosity at PRNP codon 129. The asymptomatic 56yo father had the same mutation with methionine-valine heterozygosity at codon 129. Conclusions: The proband's gastrointestinal symptoms at onset and slow course are distinct from most other prion diseases. Codon 129 heterozygosity could explain lack of symptoms in the father. AD pathology in the aunt does not preclude the presence of a PRNP mutation, as mutations have been associated with dual AD and prion pathology. PRNP testing should be considered in cases of unexplained diarrhea and dementia with or without positive FHx.

1255S

Rare disease allele penetrance and loss-of-function tolerance in a dominant disease gene: analysis of variation in >60,000 exomes. *E.V. Minikel[1,2,3], S.M. Vallabh[1,2], M. Lek[1,3], D.G. MacArthur[1,3], Exome Aggregation Consortium (ExAC).* 1) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Prion Alliance, Boston, MA; 3) Broad Institute, Cambridge, MA.

Sequencing of control individuals at an ever-larger scale has long promised to shed light on disease biology, but sample sizes are only now approaching the numbers required to afford insights into ultra-rare diseases and into genes with short coding sequences, where fewer variants are expected a priori. Here we demonstrate the power of massive exome data sets to address two critical questions about the biology of a dominant disease gene: the penetrance of previously reported disease mutations, and the discovery of loss-of-function (LoF) variants that demonstrate the tolerance of this gene to heterozygous inactivation. As part of the Exome Aggregation Consortium (ExAC) we aggregated exome sequence data from >60,000 individuals from >60 studies and performed joint variant calling of SNPs and indels across all samples. To demonstrate this dataset's power to illuminate rare disease biology, we analyzed rare (<0.1%) variation in the prion protein gene (PRNP). Prion diseases affect 1 person per 1 million population per year, and the PRNP coding sequence is short (253 codons). As no samples were ascertained on neurodegenerative disease phenotypes, they should offer an unbiased view of rare PRNP variation. We therefore used this dataset to answer two longstanding questions in prion disease biology: (1) are pathogenic point mutations in PRNP fully penetrant, and (2) is prion protein LoF tolerated in humans? We find that reported disease-associated missense mutations in PRNP are found in our dataset at >20 times the frequency expected based on genetic prion disease incidence, suggesting that a subset of mutations are either benign or incompletely penetrant. The pathogenic variants most common in our controls are those least frequently reported in prion disease patients, and vice versa, suggesting that these variants occupy a spectrum from mild risk factor to fully penetrant Mendelian mutation. Prion protein (PrP) knockout confers total resistance to prion disease and is associated with relatively mild phenotypes in mice, cows and goats. Reducing PrP expression is therefore considered to be a therapeutic strategy in prion disease, yet no LoF allele has ever been observed in humans. We find three heterozygous LoF alleles in our dataset, consistent with expectation, and phenotypic follow-up suggests that these individuals are healthy. This supports reduction of PrP expression as a viable therapeutic strategy in this class of diseases.

1256M

A novel insertion mutation of *MAPT* causes FTDP-17 with distinct pathology. *H. Morino[1], Y. Matsuda[1], R. Ohsawa[1], K. Hiraki[1], T. Kurashige[2], Y. Izumi[3], Y. Yamazaki[2], T. Takahashi[2], A. Takashima[4], Y. Soeda[4], T. Miyasaka[5], M. Higuchi[6], N. Sahara[6], T. Suhara[6], H. Shimada[6], H. Maruyama[2], H. Ito[7], H. Kawakami[1].* 1) Department of Epidemiology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima, Japan; 2) Department of Clinical Neuroscience & Therapeutics, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima, Japan; 3) Department of Clinical Neuroscience, Graduate School of Medicine, University of Tokushima Graduate School, Tokushima, Japan; 4) Department of Aging Neurobiology, National Center for Geriatrics and Gerontology, Ohbu, Japan; 5) Neuropathology Department of Medical Life Systems, Faculty of Life and Medical Sciences, Doshisha University, Kyoto, Japan; 6) Molecular Neuroimaging Program, National Institute of Radiological Sciences, Chiba, Japan; 7) Department of Neurology, Graduate School of Medicine, Wakayama Medical University, Wakayama, Japan.

PURPOSE: *MAPT*, *LRRK2*, *GRN*, and *C9orf72* are known as responsible genes for several neurological disorders, including progressive supranuclear palsy (PSP), Parkinson's disease (PD), and frontotemporal dementia (FTD). We conducted genetic analysis of a family with PSP, PD, and FTD over two generations in order to identify a causative gene for these neurological disorders.

METHODS: We performed high-density SNP typing on three affected and two unaffected persons, followed by exome sequencing on the three affected persons. Candidate regions were selected by linkage analysis and Homozygosity Haplotyping based on the results of the high-density SNP analysis. The variants were obtained using BWA, Samtools, Picard, and GATK from the exome sequencing data. We reduced the candidates by referring to the public variant databases, and confirmed the remaining mutations by Sanger sequencing. In addition, we screened for the mutations in other patients with similar symptoms.

RESULTS: Relatively long segments of high LOD score exist on chromosome 6, 9, 14, 17, and 20. Homozygosity Haplotyping indicated common haplotype regions of chromosome 13 and 17 in the patients, but not in the unaffected. From the result of exome sequencing two novel variants remained finally. One was located in the *microtubule-associated protein tau* (*MAPT*) gene that is known as the causative gene of frontotemporal dementia with parkinsonism on chromosome 17 (FTDP-17). The same variant was detected in two PD patients of another family, and the haplotype around the variant was identical to that of the index family. Biochemical experiment indicated that the MAPT mutation reduced the ability of microtubule polymerization, and accelerated tau aggregation. Pathological findings from autopsy of the FTD patient demonstrated notable observations.

CONCLUSION: The novel *MAPT* variant we identified presumably causes FTDP-17. The variant is located in the region in which several mutations were reported previously, but it is more unique that the variant is non-frameshift single amino acid residue insertion. In the further investigation, we will aim at revealing the pathogenic mechanism by evaluating biochemical changes by the mutation.

1257T

Strategy to discover new ALS causative genetic variant in Japanese ALS patients. *J. Sone[1], R. Nakamura[1], M. Nakatochi[2], H. Watanabe[1], N. Atsuta[1], F. Tanaka[3], G. Sobue[1]*. 1) Department of Neurology, Nagoya University Graduate School of Medicine, Nagoya, Aichi, Japan; 2) Center of advanced medicine and clinical research, Nagoya University Hospital, Nagoya, Aichi, Japan; 3) Department of Neurology and Stroke Medicine , Yokohama City University Graduate School of Medicine, Yokohama, Kanagawa, Japan.

Amyotrophic lateral sclerosis (ALS) is a progressive and fatal neurodegenerative disease characterized by progressive muscle atrophy, paralysis, and death within a few years of diagnosis. A number of hypotheses for the pathogenesis of ALS have been proposed, however, the pathogenic mechanism has not been clarified, and no effective therapy has been developed. Approximately 90% of ALS cases are sporadic (SALS) and 10% are familial ALS (FALS). Recently, many FALS causative genes have been identified with next-generation sequencing methods and reported. To clarify the pathogenesis of ALS, we are now doing a study on the genomic basis of FALS and SALS with a next generation sequencer and about 1,000 cases of DNA samples from the Japanese Consortium of Amyotrophic Lateral Sclerosis (JaCALS). At first, we established a genetic screening system of already discovered ALS causative genes with an Ion Torrent sequencer. We selected 28 ALS genes as follows, *SOD1, ALS2, SETX, SPG11, FUS, VAPB, ANG, TARDBP, FIG4, OPTN, VCP, UBQLN2, SIGMAR1, DAO, NEFH, DCTN1, TAF15, EWSR1, PRPH, GRN, CHMP2B, ZNF512B, PFN1, ATXN2, TFG, C9orf72, RNF19A* and *SQSTM1*. We designed a primer set for multiplex PCR by Ion Ampliseq Designer. The total coverage of the designed primer set was 97.7 % of all exon of the 28 ALS genes. We made a fragment library of ALS patients with the Ion Ampliseq kit, the designed primer set and analyzed sequences with Ion torrent systems. We screened the discovered ALS causative variant in 28 ALS genes, then we divided the ALS patients into two groups, the causative variant positive ALS group and the causative variant negative ALS group. Next, we engaged in exome sequencing of the causative variant negative ALS group. We are now collecting exon data and analyzing it with the dbSNP database, the 1000 genome database and the HGVD database (Human Genetic Variation Database in Japan). We will present the progress of this study.

1258S

Mutation detection in Amyotropic Lateral Sclerosis from RNAseq data. *K.A. Staats[1], A. Furterer[1], S. de Jong[1], A. Ori[1], L. Olde Loohuis[1], T. Wu[1], E.R. Dennis[1], C. Lomen-Hoerth[2], M. Wiedau-Pazos[3], R.A. Ophoff[1]*. 1) Center for Neurobehavioral Genetics, UCLA, Los Angeles, USA; 2) Department of Neurology, UCSF, San Francisco, USA; 3) Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, USA.

Amyotrophic Lateral Sclerosis (ALS) is a devastating progressive neurodegenerative disease, resulting in selective motor neuron degeneration and paralysis. Patients die approximately 3-5 years after diagnosis, as, at current, there is no medical cure available. Disease pathophysiology is multifactorial, including disease mechanisms such as excitotoxicity, RNA metabolism and aggregation, but is not yet fully understood. Approximately 10% of all ALS patients suffer from the disease with a familial history for which a genetic cause is assumed. Genetic analysis has proven fruitful in the past to further understand genes modulating the disease and increase knowledge of disease mechanisms, including the discovery of ALS genes such as *SOD1*, *TARDBP*, *FUS-TLS* and *C9orf72* (among others). Here, we performed RNA sequencing on whole blood from ALS patients not carrying the *C9orf72* repeat expansion. Besides obtaining gene expression data, we have analyzed the RNAseq data for disease causing variants in coding sequence. This was conducted with an adapted protocol of SNPiR with GATK variant calling. Variants detected were tested for validation by Sanger sequencing, including the validated pathogenic mutation I383V in *TARDBP*. This method allows for the secondary use of RNAseq data for (pathogenic) coding variant detection, though it is not suitable for all ALS mutations, as some mutations are not in transcripts expressed in whole blood. We used this method to examine known ALS genes but are now expanding the effort for detection of novel mutations.

1259M

***PRKAR1B* mutation associated with a new neurodegenerative disorder with unique pathology.** *T.H. Wong[1], W.Z. Chiu[1], G.J. Breedveld[2], K.W. Li[3], J.M.H. Verkerk[4], D. Hondius[3,5], R.K. Hukema[2], H. Seelaar[1], P. Frick[6], L.A. Severijnen[2], G.J. Lammers[7], J.H.G. Lebbink[8], S.G. van Duinen[9], W. Kamphorst[5], J.M. Rozemuller[5], E.B. Bakker[10], M. Neumann[6,11], R. Willemsen[2], V. Bonifati[2], A.B. Smit[3], J.C. van Swieten[1,12,13], Netherlands Brain Bank, The International Parkinsonism Genetic Networks.* 1) Neurology, Erasmus Medical Center, Rotterdam, Netherlands; 2) Clinical Genetics, Erasmus Medical Center, The Netherlands; 3) Molecular and Cellular Neurobiology, Centre for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University, Amsterdam, The Netherlands; 4) Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; 5) Pathology, VU University Medical Centre, Amsterdam, The Netherlands; 6) DZNE, German Centre for Neurodegenerative Disease, Tübingen, Germany; 7) Neurology, Leiden University Medical Centre, Leiden, The Netherlands; 8) Cell Biology and Genetics and Radiation Oncology, Erasmus Medical Center, Rotterdam, The Netherlands; 9) Pathology, Leiden University Medical Centre, Leiden, The Netherlands; 10) Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands; 11) Neuropathology, University of Tübingen, Tübingen, Germany; 12) Alzheimer Centre, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; 13) Neurology, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands.

Pathological accumulation of intermediate filaments can be observed in neurodegenerative disorders, such as Alzheimer's disease [MIM 104300], frontotemporal dementia [MIM 600274] and Parkinson's disease [MIM 168600], and is also characteristic of neuronal intermediate filament inclusion disease. Intermediate filaments type IV include three neurofilament proteins (light, medium and heavy molecular weight neurofilament subunits) and a-internexin. The phosphorylation of intermediate filament proteins contributes to axonal growth, and is regulated by protein kinase A. Here we describe a family with a novel late-onset neurodegenerative disorder presenting with dementia and/or parkinsonism in 12 affected individuals. The disorder is characterized by a unique neuropathological phenotype displaying abundant neuronal inclusions by haematoxylin and eosin staining throughout the brain with immunoreactivity for intermediate filaments. Combining linkage analysis, exome sequencing and proteomics analysis, we identified a heterozygous c.149T>G (p.Leu50Arg) missense mutation in the gene encoding the protein kinase A type I-beta regulatory subunit (*PRKAR1B* [OMIM 176911]). The pathogenicity of the mutation is supported by segregation in the family, absence in variant databases, and the specific accumulation of PRKAR1B in the inclusions in our cases associated with a specific biochemical pattern of PRKAR1B. Screening of *PRKAR1B* in 138 patients with Parkinson's disease and 56 patients with frontotemporal dementia did not identify additional novel pathogenic mutations. Our findings link a pathogenic *PRKAR1B* mutation to a novel hereditary neurodegenerative disorder and suggest an altered protein kinase A function through a reduced binding of the regulatory subunit to the A-kinase anchoring protein and the catalytic subunit of protein kinase A, which might result in subcellular dislocalization of the catalytic subunit and hyperphosphorylation of intermediate filaments.

1260T

Transcriptome sequencing in bipolar disorder identifies a global downregulation in the anterior cingulate and dysregulation of G protein-coupled receptors. *C. Cruceanu[1,2,4], P. Pavlidis[3], S. Rojic[3], P.P. Tan[3], J.P. Lopez[1,2], S.G. Torres-Platas[2,4], G.A. Rouleau[1,4], G. Turecki[1,2].* 1) Human Genetics, McGill Univ - Douglas Hosp Res Inst, Montreal, QC., Quebec, Canada; 2) McGill Group for Suicide Studies & Douglas Research Institute, Montreal, QC, Canada; 3) Centre for High-Throughput Biology and Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada; 4) Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada.

Objective: Bipolar disorder (BD) is a complex mental illness characterized by episodes of mania and depression. Gene expression changes and alterations in regulatory mechanisms have been associated with BD, mainly through candidate gene and a limited number of global microarray expression studies in postmortem brain, but questions remain about isoform-specific dysregulation of known genes as well as the plethora of non-coding transcripts whose importance has been demonstrated recently in the brain but not characterized for bipolar disorder. Method: RNA sequencing (RNA-seq) is a powerful technique that captures the complexity of gene expression, and greatly improves upon previous approaches in both accuracy and level of information. We used RNAseq in tissue from the anterior cingulate gyrus from 13 BD cases and 13 matched controls. We sequenced an average of 180M paired-end reads per subject, and obtained information from ~22,000 transcribed coding and non-coding RNA species. Results: We computed differential expression between cases and controls using the HTSeq pipeline and detected 10 differentially expressed transcripts at a false discovery rate of <5%, replicated these findings using Cufflinks, and validated them by quantitative real-time PCR. Among the most significant results, we observed genes coding for Class A G protein-coupled receptors (GPCRs), SSTR2 (somatostatin receptor 2), CHRM2 (cholinergic receptor, muscarinic 2) and RXFP1 (relaxin/insulin-like family peptide receptor 1). Interestingly, a gene ontology analysis of the entire set of differentially expressed genes pointed to an overrepresentation of genes involved in GPCR regulation. We further followed-up the top genes by querying the effect of treatment with mood stabilizers commonly prescribed in BD through an in vitro study, and found evidence that these drugs affect expression of some of these genes. Conclusion: By using total transcriptome analysis with RNASeq in the post-mortem BD brain, we identified a global downregulation in the ACC, and interesting profile of GPCR dysregulation, and pointed to several new BD genes. Furthermore, we characterized the non-coding transcriptome in BD and possibly identified the first long intergenic non-coding RNAs in BD. Our findings have important implications in regards to fine-tuning our understanding of the dysregulated BD brain as well as for identifying potential new drug target genes or pathways.

1261S

Targeted-resequencing gene panels for the genetic diagnosis of spinocerebellar ataxia and spastic paraplegia in Italian patients. *D. Di Bella[1], S. Magri[1], E. Sarto[1], S.M. Caldarazzo[1], M. Plumari[1], S. Baratta[1], L. Nanetti[1], C. Mariotti[1], C. Gellera[1], P. Bauer[2], F. Taroni[1].* 1) Unit of Genetics of Neurodeg and Metabolic disease, IRCCS Istituto Neurologico C Besta, Milan, Italy; 2) Institute of Medical Genetics, University of Tübingen, Tübingen, Germany.

Spinocerebellar ataxias and hereditary spastic paraplegias (HSP) are genetically highly heterogeneous group of neurological disorders involving both central and peripheral nervous system. More than 50 genes have already been identified for each disease group, but >50% of the patients remain undiagnosed. With the advent of next-gen sequencing, virtually all known disease genes can be tested at once, hugely increasing the expected diagnostic yield. We defined a comprehensive disease gene list and developed 2 different TruSeqCustomAmplicon (TSCA, Illumina) exon enrichment assays for 54 HSP genes and 76 genes for both dominant (SCA) and recessive (SCAR) spinocerebellar ataxias. Aim of the study was to analyze all the patients referred to our laboratory and negative for the common genes. Our data indicate that we could sequence 90% of the coding regions of these genes at more than 20X coverage. Mean coverage was 400-600 reads. A standard bioinformatic pipeline for mapping and annotation yielded a total of 200-400 variants in our disease genes per subject, which can be reduced to 5-10 with a filtration strategy. We included in the study 60 patients (9 AD, 15 AR, 36 S) with ataxia and 37 patients with sporadic spastic paraplegia. Overall, pathogenic mutations were identified in ~20% of patients negative for the common disease genes. In particular, in patients with ataxia, we identified mutations in 2 challenging genes (SYNE1 and SACS), which are extremely difficult to be studied by conventional sequencing because of their length. Moreover, we identified mutations in extremely rare SCA genes (SCA5 and SCA14) in 2 large families. As regards HSP, we could identify mutations in the large SPG11 gene and in rarer AR genes (GBA2, SPG39). Different missense variants of uncertain pathogenicity were identified in genes responsible for dominant ataxia and several heterozygous mutations were identified in recessive genes. All high-quality variants were confirmed by Sanger sequencing indicating reliability of this approach. Further analyses are required for the validation of uncertain variants and the assessment of the presence of large in/del mutations in recessive genes in which heterozygous mutations have been identified. In conclusion, our data confirm the need for multiplexed gene panel approach for the detection of the molecular causes of neurodegenerative diseases with high genetic heterogeneity. (E-Rare grant to PB and FT; Italian Ministry of Health grant to FT).

1262M

Deconstructing obsessive-compulsive disorder (OCD) by whole exome sequencing and integration of clinical endpoints and cognitive domains. *L. Domenech[1], K. Rabionet[1], G. Escaramis[1], D. Trujillano[1], S. Ossowski[1], A. Carracedo[2], P. Alonso[3], X. Estivill[1].* 1) Bioinformatics and Genomics Program, Centre for Genomic Regulation, Barcelona, Spain; 2) Santiago de Compostela University, Spain; 3) Hospital Universitary Bellvitge, IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain.

Obsessive-compulsive disorder (OCD) is a neuropsychiatric condition that affects 1-3% of the population worldwide. Genome-wide association studies for elucidating the genetic basis of OCD have given extremely limited results. We are taking a genome-centric approach to exhaustively explore the genome of selected OCD cases and to evaluate new aspects of the disease biology. As a first step in an integrative approach to dissect the biology of OCD we have performed whole-exome sequencing (WES) of 200 cases of OCD that underwent exhaustive clinical, neuro-psychological, and neuro-imaging assessment, and received standardized treatment. Control samples for WES have been the 1000 genome project European subset of WES data, the European-American data of the exome variant server (EVS), and the in-house data from over 500 WES of Spanish non-psychiatric subjects. We have implemented a Unified Mixed-Effect Models for Rare Variant Association to perform the analysis of this set of WES data. This approach can evaluate the association to disease of both rare and common variants detected by WES, implementing a combination of a Burden Test and a Sequence Kernel Association Test (SKAT), weighted by the variants' effects and frequencies. We are also developing a pathway-based approach, which will work on protein-protein interaction sub-networks (PPI-SNs), build by integrating knowledge from PPIs and identified recurrently mutated genes. In comparison to rare exome variant association studies on pathways we expect this new approach to more specifically identify protein interactions, regulatory mechanisms or protein complexes involved in OCD development. After pathway analysis and protein-protein interaction (PPI) approaches we have found an accumulation of mutations in epigenetic and neurodegenerative-related genes. Strong evidence of association of several genes in these pathways will be obtained throughout replication in well-characterized cohorts of OCD in the framework of the International OCD consortium and functional analysis of the identified variants. Supported by the Spanish Ministry of Economy and Competitiveness. LD is supported by a Severo Ochoa fellowship.

1263T

Targeted sequencing of a visual migraine aura locus on chromosome 9q22. *M.E. Hiekkala[1], M.A. Kaunisto[1,2], V. Artto[3], S. Paavonen[3], E. Hämäläinen[2], M. Färkkilä[3], A. Palotie[2,4], M. Kallela[3], M. Wessman[1,2].* 1) Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland; 2) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 3) Dept. of Neurology, Helsinki University Central Hospital, Helsinki, Finland; 4) The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

Migraine is a common and disabling neurological disorder characterized by recurrent multiphase attacks. The two main phases of a migraine attack are the aura phase and the following headache. In most cases aura is visual but it can also be sensory, motor or speech disturbance. We have identified a migraine visual aura locus in Finnish families on 9q21-q22 (Tikka-Kleemola et al. 2010). The same region has also shown linkage to occipitotemporal lobe epilepsy and migraine with aura in a Belgian family (Deprez et al. 2007). Further evidence for a migraine locus on 9q22 was reported in the genetic isolate of Norfolk Island (Cox et al. 2012).

In order to identify susceptibility variants on 9q22 we decided to sequence a 10 Mb large region showing a significant evidence of linkage to migraine visual aura in our previous microsatellite and SNP genotyping analyses. DNA samples of 85 individuals belonging to 13 families - having multiple migraine visual aura patients - were sequenced using NGS targeted sequencing techniques (SeqCap EZ Choise XL Nimblegen, HiSeq Illumina). The mean bait coverage of the targeted area of all sequenced samples was 109. Thus far data from three families, altogether 19 individuals has been analyzed. These analyses highlighted 93 potentially functional, low frequency or rare (MAF <5%) exonic variants that occurred in at least two affected offspring. These variants were located in 51 genes, many of which are involved in functions of membrane transport, neural development or epigenetic regulation. One of the functionally interesting candidate genes, *HIATL1*, harbors a previously unknown missense variant predicted to be deleterious. The variant was identified in one of the families segregating perfectly. It exists in the Finnish exome dataset Sequencing Initiative Suomi (SISu; http://www.sisuproject.fi/) with a frequency of 0.0011 but not in any other public dataset. HIATL1 belongs to the superfamily of solute carriers but its specific biological function is not well studied. It is highly expressed in the skeletal muscle but also in the central nervous system and especially in hippocampus. After analyzing data from all the families the most interesting variants will be genotyped in other visual migraine aura sample sets.

1264S

Exome sequencing of familial agenesis of corpus callosum cases. *L. Jouan[1], B. Oulad Amar Ben Cheikh[1], H. Daoud[1], A. Dionne-Laporte[1], S. Dobrzeniecka[1], D. Spiegelman[1], D. Rochefort[1], P. Hince[1], A. Szuto[1], P.A. Dion[1,2], E. Sherr[5], H. Théoret[3,4], G.A. Rouleau[1].* 1) Montreal Neurological Institute and Hospital, Department of Neurology and Neurosurgery, McGill University, Montréal, QC H3A 1A1, Canada; 2) Department of Pathology and Cell Biology, Faculty of Medicine, Université de Montréal; 3) Department of Psychology, Université de Montréal, QC H3C 3J7, Canada; 4) CHU Sainte-Justine Research Center, Montreal, QC H3T1C5, Canada; 5) Department of Neurology, University of California, San Francisco, California, United States of America.

Agenesis of the corpus callosum (ACC), the largest connective structure in the brain which assures transfer of information between the two cerebral hemispheres, is a common brain malformation which can be observed either as an isolated condition or as a manifestation of numerous congenital syndromes. Its estimated prevalence is 3 to 7 per 1000 births but it has also been shown that 3-5% of individuals with neurodevelopmental disorders have ACC. The exact prevalence of asymptomatic ACC is unknown but it has been estimated to be around 0.5/10000. While accumulating evidence suggests an important contribution of genetics to the etiology of ACC, causative genes remain largely unknown. To identify novel ACC genes, we performed exome sequencing in a family with three children with ACC and no other malformation. Since our pedigree clearly suggested a recessive mode of inheritance, we prioritized genes with either homozygous or compound heterozygous (CH) mutations shared by the three affected patients. We identified rare CH mutations in the *CDK5RAP2* gene (Cyclin dependent kinase 5 regulatory subunit-associated protein 2), c.G280C (p.G94R) and c.A3695G (p.N1232S) inherited from the unaffected mother and father, respectively. Interestingly, *CDK5RAP2* is also known as *MCPH3*, a causative gene for autosomal recessive primary microcephaly in which the reduced brain size is believed to result from asymmetric division of neuronal progenitor cells causing a reduced number of neurons. To validate the role of the *CDK5RAP2* gene in ACC, we used the Access Array System (Fluidigm) combined with next generation sequencing to screen 288 additional patients with ACC as well as 288 unaffected controls. We identified one additional patient with 2 rare potential compound heterozygous mutations in *CDK5RAP2*, therefore reinforcing the role of this gene in isolated ACC. This is the first description of *CDK5RAP2* recessive mutations in patients with isolated ACC. Altogether, our findings unravel an unprecedented role of *CDK5RAP2* in cerebral cortex development and offer novel insights into the physiopathology of ACC.

1265M

Rare Alleles Altering Schizophrenia Risk Occur in Exons and Noncoding Functional Sequences. *E.K. Loken[1, 2, 3], S.-A. Bacanu[1, 2], D. Walsh[4], F.A. O'Neill[5], K.S. Kendler[1, 2], B.P. Riley[1, 2].* 1) Virginia Institute for Psychiatric Genetics, Virginia Commonwealth University, Richmond, VA; 2) Dept. of Psychiatry, Virginia Commonwealth University, Richmond, VA; 3) Clinical and Translational Research Center, Virginia Commonwealth University, Richmond, VA; 4) Health Research Board, Dublin, Ireland; 5) Queens University Belfast, Belfast, Northern Ireland.

Background: Schizophrenia demonstrates high heritability, in part accounted for by common simple nucleotide variants (cSNV), rare copy number variants (CNV) and most recently rare exome variants (rSNV). Although heritability explained by rSNV and CNV is small compared to that explained by cSNV, rSNV in functional sequences may identify specific disease mechanisms. However, current exome methods do not capture a large proportion of potentially functional bases where rare variation may impact disease risk: as much as two-thirds of conserved sequences lie outside the exome in noncoding regions of cross-species evolutionary constraint. Methods: We reasoned that the candidate loci from the Psychiatric Genomics Consortium Phase 1 (PGC1) schizophrenia study represent good target loci to test for the impact of rare SNVs in non-coding constrained regions. We developed custom reagents to capture mammalian constrained non-coding regions, exons and 5'- and 3'-untranslated regions (UTRs) in the 12 PGC1 loci for pooled sequencing in 912 cases and 936 controls. Results: Our design contains substantially more highly conserved bases (46,412 vs. 31,609) and variants (390 vs. 193) in noncoding as compared to coding targets. Using C-alpha to detect excess variance due to aggregate risk increasing or decreasing rSNV effects, we identified signals attributable to alleles with MAF < 0.1% in both coding sequences and in functional noncoding sequences, including variants within ENCODE transcription factor binding sites, DNase hypersensitive regions, and histone modification sites in neuronal cell lines. We also observed significant excess risk-altering variation in the CUB domain of CSMD1, a gene expressed in the developing central nervous system. Discussion: These results support the hypothesis that common and rare variants in the same loci contribute to schizophrenia risk, but highlight the need to expand capture strategies in order to detect trait-relevant sequence variation in a broader set of functional sequences.

1266T

A targeted-resequencing approach for the genetic diagnosis of inherited peripheral neuropathies in Italian patients. *S. Magri[1], D. Di Bella[1], M. Milani[1], S.M. Caldarazzo[1], P. Saveri[2], I. Moroni[3], D. Pareyson[2], F. Taroni[1].* 1) Unit of Genetics of Neurodegenerative and Metabolic Disease, IRCCS Istituto Neurologico Carlo Besta, Milan, MI, Italy; 2) Neurology Unit 8, IRCCS Istituto Neurologico Carlo Besta, Milan, MI, Italy; 3) Child Neurology Unit IRCCS Istituto Neurologico Carlo Besta, Milan, MI, Italy.

Inherited peripheral neuropathies are a complex group of genetically heterogeneous disorders caused by mutations in more than 70 genes that lead to the distal degeneration of peripheral nerves. They are divided into 3 groups according to the phenotype and the neuropathological changes: 1) Charcot-Marie-Tooth (CMT) disease characterized by motor and sensory neuropathy, subdivided into demyelinating (CMT1-4) and axonal (CMT2) forms; 2) distal hereditary motor neuropathies (dHMN), which have only motor involvement; 3) hereditary sensory neuropathies (HSN), which involve sensory nerves. These disorders not only represent a phenotypic continuum, but also show a genetic overlap and inheritance pattern variability. For these reasons, efficient genetic diagnosis would require a comprehensive and systematic approach such as that provided by NGS technology. We developed and validated two customized gene panels using TruSeqCustomAmplicon (Illumina) technique: one included 53 genes for CMT2 and dHMN, while the second one included 54 genes for CMT1-4 and HSN. We analyzed 78 patients previously screened for the CMT1A deletion/duplication and for ~4-5 genes selected according to the phenotype. We identified ~300 variants per patient, which are reduced to 2-8 by filtering. All variants were subsequently imported into a local database created specifically for the purpose of: 1) filtering identified variants according to quality criteria, functional consequence, and frequency in variant databases; 2) annotating the frequency of polymorphic variants in the Italian population; 3) collecting data to allow an easy and quick query also by other laboratories. Coverage analysis revealed that more than 90% of the target region has coverage greater than 20X and a mean coverage of about 400X. The molecular cause of the disease was identified in 24/78 patients (30.8%). Moreover, likely pathogenic/uncertain variants were identified in 18% of patients and one heterozygous mutation in recessive genes was identified in 22% of patients. In these cases, further analyses are required to validate variants and to assess the possible presence of large in/dels. In conclusion, this approach allowed us to increased by 10-fold the number of genes analyzed, significantly expanding the diagnostic yield. Our data confirm the usefulness and time/cost effectiveness of NGS-based gene panels for the genetic diagnosis of hereditary neuropathies. (Supported by Italian Ministry of Health grant to FT).

1267S

Molecular Studies of mTOR and Tau pathways in Focal Cortical Dysplasia. *M.G. Mazutti[1], F.R. Torres[1], P.A.O. Ribeiro[1], S.H. Avansini[1], R. Secolin[1], B. Carvalho[1], M.G. Borges[1], F. Rogério[2], L.S. Queiroz[2], A.C. Coan[3], H. Tedeschi[3], E.P.L. Oliveira[3], F. Cendes[3], I. Lopes-Cendes[1], The Brazilian Institute of Neuroscience and Neurotechnology (BRAINN).* 1) Department of Medical Genetics; School of Medical Sciences, University of Campinas UNICAMP , Campinas, Brazil; 2) Department of Anatomical Pathology; School of Medical Sciences, University of Campinas UNICAMP , Campinas, Brazil; 3) Department of Neurology; School of Medical Sciences, University of Campinas UNICAMP , Campinas, Brazil.

Focal cortical dysplasia (FCD) is a sub-type of malformation of the cerebral cortex which is a frequent cause of intractable focal seizures, requiring surgical treatment. FCD presents cortical architecture abnormalities also observed in tuberous sclerosis (TS) and hemimegalencephaly (HME). In addition, FCD, TS and HME show aberrant expression of genes belonging to the mTOR signaling pathway. Potential involvement of Tau pathway was also reported in FCD. Therefore, the similarity in histological features as well as abnormal gene expression pattern suggests that pathogenic mechanisms could be common to these three disorders. Recently, somatic mosaic mutations have been indentified in patients with TS and HME. Therefore, the objective of this work is to investigate whether somatic mosaic mutations in genes belonging to the mTOR and Tau pathways are present in the central nervous system of patients with FCD. NGS was performed in genomic DNA extracted from brain tissue resected by surgery (BTRS) and peripheral blood of patients with FCD. We performed exome capture with Nextera® Expanded Kit (Illumina®) and NGS on a Hiseq 2500 bench top sequencing machine. A bioinformatics pipeline was applied, using filters to variants present only in brain tissue. To date, BTRS and blood samples of four patients with FCD were sequenced and a total of 749 and 91 variants were identified in genes belonging to the mTOR and Tau pathways, respectively. Among these variants we found 107 and 12 mutations in a mosaic state (present only in the BTRS samples), respectively; including 77 variants not described in databases of human mutations. Genes disrupted by mutations code for proteins involved in regulation of cell growth and cellular processes such as proliferation, differentiation and development, as well as genes already implicated in other cerebral cortical malformations. Our preliminary results confirm the presence of mosaic mutations in mTOR and Tau pathways in FCD. However, it is still not clear whether these mosaic mutations are a causative factor in FCD or a consequence of mutations in another gene whose disruption causes a cumulative mutational burden by dysregulation of mitotic checkpoints. Additional experiments, including high-deep NGS, will be carried out in order to answer these questions. Supported by CEPID-FAPESP.

1268M

Evidence for association of CDH26 with Autism Spectrum Disorders. *F. Mentch[1], R. Golhar[1], K. Wang[1,2], J. Bradfield[1], S. Murray[3], H. Hakonarson[1,4].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) Department of Pathology, University of California San Diego, La Jolla, CA; 4) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA.

Background: Autism spectrum disorders (ASDs) have been associated with genetic variation near neuronal cell-adhesion molecules (NCAMs). Apart from association through common GWAS variants, this category of molecules has been shown to be disrupted due to rare copy number variation (CNVs) in a subset of ASD cases. Here, we investigated if rare single nucleotide variants (SNVs) in neuronal cell adhesion genes are associated with ASD. Methods: We performed exome capture and targeted sequencing of 99 cell adhesion genes, including neurexins, neuroligins, semaphorins, cadherins, protocadherins and other ASD candidate genes. A total of 225 ASD cases were sequenced on Illumina HiSeq. Variants were annotated using ANNOVAR then filtered for potential pathogenic impact. Candidate variants were subsequently tested for association in 3,100 ASD cases and unrelated 2,992 controls (966 cases, 776 controls on Veracode; 2,134 cases, 2,216 controls on Goldengate) previously registered in the de-identified biorepository at the Center for Applied Genomics (CAG) at The Children's Hospital of Philadelphia (CHOP). EIGENSTRAT and R::KNN1 were used to control for population stratification and PLINK for Chi-Squared test for association. Results: Of 292 annotated SNPs from the targeted sequencing studies and typed in 3,100 ASD cases and 2,992 controls, three missense mutations in CDH26, not reported previously, were found to be associated with ASD ($0.0001 <= P <= 0.007$), with the top SNP surviving Bonferroni corrections for multiple testing. Of those, (1) rs11086690 (MAF 0.007, OR 2.06, SIFT 0.18) is located at the end of the Ca2+ binding site, which mediates cell-cell contact, with (2) rs28409250 (MAF 0.007, OR 1.94, SIFT 0.72) residing close to a Ca2+ binding site, whereas (3) rs41310817 (MAF 0.009, OR 1.83, SIFT 0.3) is not part of any protein domain. rs11086690 is in strong linkage disequilibrium (LD) with rs28409250 ($r2=0.99$), while rs41310817 ($r2=0.57$), and rs28409250 and rs41310817 ($r2=0.56$) is similar. Conclusions: These associations demonstrate that both common and rare variants in CDH26 associate with ASDs, suggesting they may play a role in the pathogenesis of the disease.

1269T

Searching for a common founder - exome sequencing of sporadic early-onset Parkinson's disease in Norway. *A.H. Rengmark[1], L. Pihlstrøm[1], Z. Iqbal[1], M.D. Vigeland[2], M. Toft[1].* 1) Department of Neurology, Oslo University Hospital, Oslo, Oslo, Norway; 2) Department of Medical Genetics, Oslo University Hospital, Oslo, Norway.

Age is a major risk factor for developing Parkinson's disease (PD), and <1 % of patients develop disease before the age of 40. Mutations in known autosomal recessive genes are normally found in up to 10 % of patients with early-onset PD (EOPD). The Norwegian population is relatively homogeneous and of limited size. Common founder mutations have therefore been identified in a number of Mendelian disorders in the past. We hypothesized that a subset of sporadic EOPD in Norway might be caused by one or several founder mutations inherited as a recessive trait. 53 unrelated Norwegian EOPD patients were included in the study (mean age 37.1 ± 4.1 SD, range = 26-47 years). Library preparation and exome sequencing was performed by commercial providers, using Illumina and Solid sequencing platforms. Bioinformatics analyses were based on BWA (alignment) and GATK (processing and variant calling). To identify single nucleotide variants (SNVs) shared by ≥2 patients we used FILTUS, a new bioinformatic tool developed at our institution. Non-synonymous coding homozygous SNVs with minor allele frequency ≤0.02 in 1000genomes and EPS5400 databases were further filtered by removing: (1) populations-specific variants found in our in-house database, (2) sequencing platform specific variants and (3) SNVs located in highly polymorphic genetic regions. All remaining SNVs were verified by Sanger-sequencing. We also performed genotyping with the Infinium HumanExomeCore bead chip (Illumina) to identify runs of homozygosity (ROH). Our analyses have identified 12 genes containing homozygous SNVs shared by 2-5 patients. This includes variants in the MYO7B (myosin VIIB) and MYADML2 (myeloid-associated differentiation marker-like 2) genes. 23 homozygous SNVs were identified within ROH, including a mutation in the PRR12 gene that was found in 2 different patients. The function of this gene is not well characterised, but has been reported to be expressed in the brain. All identified homozygous SNVs are currently being genotyped in larger patient series and candidate genes will be sequenced in additional EOPD patients. In addition, analyses of compound heterozygous variants and indels are on-going. Our study will indicate if exome sequencing of sporadic patients with an extreme phenotype (early disease onset) can be used to identify new autosomal recessive genes contributing to complex disorders.

1270S

Whole exome sequencing identifies MEOX2 as a candidate genetic factor in posterior cortical atrophy. *E.C. Schulte[1, 7], R. Perneczky[2], A. Kurz[2], J. Diehl-Schmid[2], D. Rujescu[3], M. Hüll[4], A. Peters[5], C. Gieger[6], T. Meitinger[7, 8], H. Klünemann[9], J. Winkelmann[1, 7, 8, 10]*. 1) Neurologische Klinik und Poliklinik, Technische Universität München, Munich, Germany; 2) Klinik für Psychiatrie, Technische Universität München, Munich, Germany; 3) Klinik für Psychiatrie, Friedrich Schiller Universität Jena, Jena, Germany; 4) Universitätsklinik für Psychiatrie und Psychotherapie, Albert Ludwigs Universität Freiburg, Freiburg, Germany; 5) Institute for Epidemiology II, Helmholtz Zentrum München, Munich, Germany; 6) Institute for Genetic Epidemiology, Helmholtz Zentrum München, Munich, Germany; 7) Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany; 8) Institut für Humangenetik, Technische Universität München, Munich, Germany; 9) Psychiatrische Klinik, Universität Regensburg, Regensburg, Germany; 10) Department of Neurology and Neurosciences, Stanford University, Palo Alto, CA, USA.

Background: Posterior cortical atrophy (PCA) is a rare neurodegnerative condition presenting with an impairment in visual processing skills which shares clinical characteristics with Alzheimer's disease (AD). The causes of PCA are not fully understood. Here, we performed whole exome sequencing in nine individuals with PCA in search for rare genetic variants that contribute to this phenotype. Methods: Whole exomes of 8 individuals with PCA and one trio were sequenced. Identified variants were filtered under autosomal dominant (auto dom), autosomal recessive (auto rec) and de novo models. Burden tests (CAST) were performed for all candidate genes using data from 3891 in-house exomes and 4300 NHLBI ESP exomes as controls. Subsequently, the coding regions of the most plausible biological candidate, *MEOX2*, were screened for variants by high-resolution melting curve analysis in 1271 individuals with AD and 3426 general population controls. Results: We identified de novo missense variants in *KIRREL3*, *MEOX2* and *CACNA1S*. Joint analysis of all 9 PCA cases yielded 44 and 37 genes harboring rare variants (MAF $\leq$ 0.1% for dom and $\leq$ 2.5% for rec alleles) in at least 30% of cases under an auto dom and rec model, respectivley. None of these overlapped with any of 13 genes in which the affected son of the trio held homozygous or compound heterozygous variants. None of the variants occurred in more than a single case. Burden testing identified 7 auto dom—*ANKR6/LYRM2, KEL, NAA16, NYX, PPP1R3E, SIGIRR*—and 3 auto rec —*AFAP1L2, ALPK2, IFNA10*—candidate genes in which rare variants were associated with the PCA phenotype at exome-wide significance. As *MEOX2* has been linked to AD, we screened the coding regions of this gene for additional rare variants in a large AD sample. Rare nonsynonymous variants in MEOX2 showed a trend towards being more common in individuals with AD compared to controls (38 of 1271 AD vs. 72 of 3426 controls; p=0.07, χ^2 test). The de novo variant was not found again. Conclusions: We were unable to causally link rare coding variants in any gene to the PCA phenotype but created the first bias-free short-list for potential candiate genes for PCA. A de novo variant in the homeobox gene *MEOX2* seemed especially interesting because individuals with extreme AD phenotypes have been shown to harbor copy number variants overlapping *MEOX2* and *Meox2*$^{-/-}$ mice develop an AD-like phenotype.

1271M

Exome sequencing identifies a novel missense mutation in MFN2 in familial dysautonomia. *Z. Wei[1], D. Li[2], E. McCormick[3,4], RM. Chiavacci[2], L. Tian[2], MJ. Falk[3,4], H. Hakonarson[2,3,5]*. 1) Department of Computer Science, New Jersey Institute of Technology, Newark, New Jersey, USA; 2) The Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA; 3) Division of Human Genetics Department of Pediatrics, The Children's Hospital of Philadelphia and The Perelman School of Medicine, Philadelphia, Pennsylvania, 19104, USA; 4) Division of Child Development and Metabolic Disease, Department of Pediatrics, The Children's Hospital of Philadelphia and The Perelman School of Medicine, Philadelphia, Pennsylvania, 19104, USA; 5) Division of Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

Whole exome sequencing (WES) analysis identified a novel heterozygous c.1699A>G (p.M567V) missense mutation in MFN2 that segregated with familial dysautonomia in which the teenage daughter has history of headaches, poor coordination and unawareness of her body position, extreme fatigue and weakness, and the mother has a similar history of muscle fatigue/pain/heaviness, frequent headaches, intermittent confusion episodes and non-refreshing sleep. Extensive clinical diagnostic genetic testing and whole mitochondrial genome sequencing were unrevealing before the research-based WES analysis. We examined missense, nonsense, splice-altering, and coding indels matching the dominant inheritance model. Results were filtered to exclude synonymous variants, variants with minor allele frequency (MAF) greater than 0.5%, and variants previously identified in controls by our in-house exome variant database. Relevant candidates were taken forward for manual curation. MFN2, encodes a mitochondrial membrane protein that participates in mitochondrial fusion and contributes to the maintenance and operation of the mitochondrial network, was identified as the most likely candidate. Mutations in this gene cause hereditary motor and sensory neuropathy and Charcot-Marie-Tooth disease type 2A, which are both disorders of the peripheral nervous system. We are currently conducting follow-up functional studies to elucidate the mechanisms in mitochondrial fusion defect that will be presented together.

1272T

Inherited and de novo Transposable Elements in schizophrenia. *F. Macciardi[1,7,8,9], G. Guffanti[2], P. DeCrescenzo[2], C. Pato[3], H. Mangalam[4], A. Rodriguez[5], R. Madduri[5], S. Gaudi[6].* 1) Dept of Psychiatry & Human Behavior, University of California, Irvine, Irvine, CA; 2) Department of Psychiatry Division of Epidemiology & Division of Child and Adolescent Psychiatry Columbia University/NYSPI New York, NY; 3) Department of Psychiatry and the Behavioral Sciences Keck School of Medicine, USC, Los Angeles, CA; 4) OIT, University of California, Irvine, Irvine, CA; 5) MCS, Argonne National Laboratory Computation Institute, University of Chicago; 6) Istituto Superiore di Sanita', Rome, Italy; 7) Center for Autism Research and Treatment (CART), University of California, Irvine, California; 8) Center for Epigenetics and Metabolism, University of California, Irvine, California; 9) Department of Pharmacological and Biomolecular Sciences, University of Milan, Milan, Italy.

Transposable Elements (TEs) accumulated over evolutionary time making up between 1/2 and 2/3 of the human genome. LINE1s are particularly active in creating new retro transposition events, several of which can be highly polymorphic and not easy to detect. The extent of transposon-mediated mutagenesis has been mostly investigated using sophisticated sequencing library design to specifically capture TE insertions. Recently, two computational methods, Retroseq and Tangram, have been developed to allow detecting TE insertions using traditional next-generation sequencing. We developed a workflow to detect TE insertions using either one of the two algorithms, compare the calls made by each single algorithm and retain a list of TE insertions identified by both algorithms, perform local de novo alignment (Velvet) of the regions harboring the new insertion sites (computational validation) and assign TE family and type to the de novo assembled sequence using RepeatMasker. We analyzed the whole genome sequences from six members of a family in which schizophrenia (SZ) segregates. This family is part of the Genomic Psychiatric Cohort (GPC) sample (Pato et al, 2013). Our goals are (1) to identify de novo TE insertions that may represent potential risk factors to develop SZ and (2) to estimate the sensitivity of the method to detect and computationally validate new TE insertions. We identified 29 new L1 insertions and we replicated 2 previously described L1HS originally identified in a re-analysis of the 1,000 Genomes Project trios (Rouchka et al, 2010), but not yet reported in Repbase to our best knowledge. RepeatMasker classifies all the 29 insertions as either L1HS and/or L1P1 with the only exception of 1 L1PA2, confirming that these are mostly evolutionary young insertions. More than half (55%) of these insertions are located within known GENCODE v7 annotated genes or lincRNAs, supporting the impact of young insertions on gene function and regulatory elements. None of the insertions is de novo, but the pattern of inheritance from parents to offspring suggests that all insertions are germline. At least 3 of them, all mapping to known genes, are transmitted uniquely from the SZ affected mother to two SZ affected children. This is the second study that evaluates the pattern of inheritance of new TE insertions other than the 1000 genomes Project and the first one that looks at transmission disequilibrium of germline TEs in a SZ affected family.

1273S

Evidence for differential X chromosome gene expression in children with sex chromosome aneuploidies. *D. Hong[1,3], X. Zhang[1,2], X. Zhu[1,2], A. Reiss[1,3], A. Urban[1,2].* 1) Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA; 2) Genetics, Stanford University, Stanford, CA; 3) Center for Interdisciplinary Brain Sciences Research, Stanford University, Stanford, CA.

Introduction: Sex differences in neurodevelopment have been well-documented. Prior research in this area has primarily focused on the influence of sex hormones, though there is increasing evidence that differential gene expression from the sex chromosomes plays an equally important role. Sex chromosome aneuploidies, such as Turner syndrome (45X) and Klinefelter syndrome (47XXY), provide a unique model in which to investigate these effects. Analysis of gene expression patterns may help elucidate evidence for sex-chromosome associated dosage effects. Methods: Using peripheral whole blood, we undertook deep RNA-Seq to characterize expression profiles from a cohort of four individuals - a twin pair of a 9-year old female with TS and her 9-year old typically developing sister, as well as a 10-year old male with KS and an 8-year old male control. Sequencing was completed using Illumina HiSeq with ~50M 2×100nt paired-end reads per proband. Overall transcript levels from the X chromosome were quantified and compared to overall transcripts from all autosomes. Results: Despite the small sample size, the ratio of X chromosome activity relative to autosomes suggests that global transcription activity of the X chromosome is dose-dependent, as both the typically developing female twin (46XX) and the male with Klinefelter syndrome (47XXY) had greater than 10-fold higher X:autosome ratios compared to the female twin with Turner syndrome (45X) and the typically developing male control (46XY). Conclusion: Previous studies characterizing sex chromosome function suggest that dosage compensation mechanisms compensate for dose imbalance between males and females, as well as between the sex chromosomes and autosomes. This includes the existence of homologous pseudoautosomal regions on the sex chromosomes and upregulation of the active X chromosome in females and the single X chromosome in males to match the expression output of autosomes genome-wide. However, these mechanisms are insufficient to explain the notable phenotypic variance associated with sex chromosome aneuploidies. Here we provide novel evidence for dosage-related transcriptome sequencing patterns that directly correlate with the number of X chromosomes in a cohort of children with sex chromosome aneuploidies.

1274M

Subcortical band heterotopia (double cortex syndrome) not associated with DCX or LIS1 gene mutations. *E. Andermann[1, 5, 6, 8, 10], D. Amrom[1, 5, 6, 10], F. Dubeau[2, 4, 5, 6, 10], D. Melançon[3, 5, 7, 10], D. Tampieri[3, 5, 7, 10], F. Andermann[4, 5, 6, 9, 10], B. Dobyns[11].* 1) Neurogenetics Unit; 2) Seizure Service; 3) Department of Neuroradiology; 4) Seizure Clinic; 5) Montreal Neurological Hospital and Institute; 6) Department of Neurology & Neurosurgery; 7) Department of Radiology; 8) Department of Human Genetics; 9) Department of Pediatrics; 10) McGill University, Montreal, Quebec, Canada; 11) Departments of Pediatrics and Neurology, University of Washington; and Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA.

Subcortical band heterotopia (SBH), also called double cortex syndrome, is a malformation of brain cortical development due to deficient neuronal migration. It is classically due to either DCX or LIS1 gene mutations, resulting in a predominantly anterior or posterior distribution of the SBH respectively. However, a number of patients do not harbor point mutations or deletions/duplications of either of these genes, suggesting even greater genetic heterogeneity. We reviewed our database of adult epilepsy patients with SBH in whom no DCX or LIS1 mutation was found. Fifteen patients presenting with epilepsy, with or without associated cognitive delay, had SBH detected on brain MRI for which no DCX or LIS1 mutation was found. They had no associated pachygyria. All patients were sporadic. They were evaluated with a karyotype and DCX and/or LIS1 mutation analysis, depending on the pattern of distribution of their SBH. A few patients had both DCX/LIS1 sequencing and deletion/duplication analysis when their SBH was diffuse. In one woman, the karyotype showed a rearrangement of chromosome 9 suggesting a paracentric inversion of the q13->q22.3 region. Karyotype in both parents was normal, indicating a de novo chromosomal rearrangement in the proband. This patient had a history of infantile spasms, morning myoclonic seizures and pervasive developmental delay. She presented a thick and diffuse SBH. DCX/LIS1 sequencing and deletion/duplication analysis were negative. A CGH microarray with 135000 oligonucleotides was added and did not reveal any deletion or duplication. A higher resolution CGH microarray performed in order to detect a possible microdeletion or duplication was inconclusive for gene identification. Genetic evaluation of SBH deserves DCX and/or LIS1 sequencing and deletion/duplication analysis, but does not always reveal the genetic cause of this brain malformation. Karyotype analysis and/or CGH array may be of interest in these unsolved cases. Whole exome sequencing should be carried out to elucidate the genetic cause and provide genetic counseling for these families.

1275T

The astrocytic transporter Slc7a10 (Asc-1) is required for glycinergic inhibitory function. *J.T. Ehmsen[1], Y. Liu[2], Y. Wang[2], J.D. Rothstein[1], S.H. Snyder[1], M.P. Mattson[2], A. Hoke[1].* 1) Department of Neurology, Johns Hopkins, Baltimore, MD; 2) Laboratory of Neurosciences, National Institute of Aging Intramural Research Program, Baltimore, MD.

Slc7a10 (Asc-1) is a sodium-independent neutral amino acid transporter known to be the primary mediator of D-serine transport in the brain. Slc7a10 transports a number of additional amino acids including glycine, L-alanine, L-serine, and L-cysteine, as well as their D-enantiomers. We find that Slc7a10 is enriched in cerebellar Bergmann glia and within a subset of astrocytes of the caudal brain and spinal cord, in a distribution corresponding to high densities of glycinergic inhibitory synapses. Accordingly, we find that spinal cord glycine levels are significantly reduced in Slc7a10-null mice and that spontaneous glycinergic postsynaptic currents in motor neurons of mice lacking Slc7a10 show substantially diminished amplitude, identifying the likely etiology of sustained myoclonus and early postnatal lethality previously described for these animals. These observations establish a critical role for astrocytic Slc7a10 in glycinergic inhibitory function in the central nervous system, and implicate SLC7A10 as a candidate gene in human hyperekplexia and stiff person syndrome.

1276S

Temporal mRNA expression profile of cyclooxiganase-2a and cyclooxiganase-2b genes in adult and larvae zebrafish brain after pentylenetetrazole-induced seizure. *H.M. Gomide, P.G Barbalho, D.M Nakata, I.L Cendes, C.V Maurer-Morelli.* Medical Genetics, State University of Campinas, Capinas, Sao Paulo, Brazil.

Introduction: Although zebrafish have become a model for seizure studies, there are no studies about neuroinflammatory response after seizure in this model; therefore, we investigated the temporal mRNA expression profile of ciclooxigenase-2 genes after pentylenetetrazole (PTZ)-induced seizure in adult and immature zebrafish brain. Methods: All experiments were approved by Animal Ethical Committee/UNICAMP (#3098-1). Seven days post-fertilization (dpf) larvae and adult zebrafish were separated in Seizure (SG) and Control (CG) groups (n=5 each group. Larvae sample was composed by pooling 20 heads, and adult sample by pooling two brain). Animals from SG were exposed to PTZ 15mM, and animals from CG were handled in PTZ-free water. At 0.05h, 1h and 6h after seizure, animals were anesthetized and their heads/brains collected for RT-qPCR that was carried out in triplicates with ef1α as endogenous controls using TaqMan System. The relative quantification was calculated by the equation RQ=2-$\Delta\Delta$CT. Statistical analyses were performed by Mann-Whitney test, using the GraphPad Prism with $p < 0.05$. Results: for seven dpf larvae, only the cox-2b mRNA was increased at 0.05h and 1h after PTZ-induced seizure when compared to CG. The mean ± SEM data were: (i) cox2a: CG0.05h 1.2±0.06 vs SG0.05h 1.3±0.12 (p = 0.42); CG1h 1.32±0.5 vs SG1h 2.33±0.52 (p=0.055); CG6h 1.08±0.11 vs SG6h 1.04±0.06 (p=0.34); (ii) cox-2b: CG0.05h 0.93±0.02 vs SG0.05h 1.73±0.18 (p= 0.004); CG1h 1.43±0.4 vs SG1h 2.58±0.23 (p =0.047); CG6h 1.08 ±0.11 VS SG6h 1.05±0.13 (p=2.7). For adult zebrafish, the cox-2b mRNA levels were increased only at 6h after PTZ-induced seizure compared to CG. The mean ± SEM data were: (v) cox-2a: CG0.05h 0.74±0.08 vs SG0.05h ± (p = 0.42); CG1h 0.66±0.12 vs SG1h 0.62±0.06 (p =0.34); CG6h 1.08±0.2 vs SG6h 0.87±0.22 (p=0.07); (vi) cox-2b: CG0.05h 1.04±0.09 vs SG0.05h 1.3±0.2 (p=0.2); CG1h 1.03±0.06 vs SG1h 1.21±0.13 (p=0.2); CG6h 0.98±0.03 vs SG6h 1.21±0.08 (p=0.04). Conclusion: This is the first study investigating the temporal mRNA expression profile of cox2a and cox2b genes in adult and larvae zebrafish brain after seizures. Our results showed that both genes have different mRNA expression response after PTZ-induced seizure, suggesting that zebrafish cox-2b is more similar than is cox-2a from mammalian Cox-2. This study provides evidence that this little fish is a valuable model for further studies of inflammation and seizures. Support: FAPESP and CNPq.

1277M

Profiling gene expression in CFW mouse brains to refine our understanding of the genetic architecture of behavioral traits. *S. Gopalakrishnan[1], P. Carbonetto[1], N.M. Gonzales[1], E.H. Leung[1], J. Park[1], E. Aryee[1], J. Davis[3], C.C. Parker[2], A.A. Palmer[1].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Dept. of Psychology and Program in Neuroscience, Middlebury College, Middlebury, VT; 3) Stanford School of Medicine, Department of Genetics, Stanford, California.

Translating results from genome wide association studies to elucidate the underlying biology of complex human diseases and traits remains a massive challenge. It is now widely appreciated that variation in gene expression plays an important role in the genetic architecture of most complex traits. However, it is not always feasible to access the relevant tissues in human subjects to measure gene expression. Here we demonstrate the use of an outbred (CFW) mouse population to model neuropsychiatric traits for methamphetamine response sensitivity, conditioned fear behavior and prepulse inhibition. We complement this with RNA-Seq assays to quantify gene expression in three brain regions that are relevant to these traits, the hippocampus (n = 80), striatum (n = 55) and pre-frontal cortex (n = 54). The CFW mouse stock offers several important advantages over other mouse populations for mapping, especially the resolution of QTL regions. We mapped QTLs and eQTLs in these animals using genotypes obtained from a modified Genotyping-by-Sequencing protocol and a linear mixed model. Each QTL region contains many fewer genes as compared with traditional mouse populations; however, in many cases the QTLs still implicate dozens of genes. In these cases, gene expression data can be useful to prioritize candidate genes within QTL regions. In each tissue, we identified expression QTLs (eQTL) using two approaches: a linear mixed model approach to assess support for associations, and a two-sample paired t-test to identify allele-specific expression. Using eqtlbma (Flutre et al.) to model shared eQTLs across tissues, we find that a high proportion of eQTLs (approximately 70%) are common to the three brain tissues, with a small set of tissue specific eQTLs. Within each behavioral QTL, we prioritize genes for which there is an eQTL under this QTL peak, i.e. genes with expression affected by a variant associated with behavior are ranked higher than genes without such an eQTL. Preliminary results suggest that gene expression is an important mediator of behavior, with more than half of the QTL regions containing at least one variant that is also an eQTL.

1278T

Contiguous deletion of CADPS2 and GRM8 associates with severe autism spectrum disorder. *C. Hatano[1], T. Yokoi[1], K. Wakui[2], K. Enomoto[1], Y. Kuroda[1], I. Ohashi[1], R. Kosaki[3], K. Kurosawa[1].* 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Division of Medical Genetics, National Center for Child Health and Development, Tokyo, Japan.

Autism spectrum disorders (ASDs) are complex neurobehavioral conditions and known to have multi-hit etiology models. Some copy number variations (CNVs) in human genes, in both cases of rare inherited and de novo, contribute to genetic weakness in ASD patients. The identification of those genes seems to support proving the clinical states. We report on a female patient with severe autism, associated with contiguous deletion of CADPS2 and GRM8 at 7q31. She was born to nonconsanguineous parents at 41 weeks of normal pregnancy. The parents were healthy and their family history was unremarkable. Her birth weight was 3328 g (+0.82 SD), length 49.4 cm (+0.47 SD), and the head circumference 34.5 cm (+1.14 SD). She was first evaluated in neonate because of feeding difficulty and developmental delay (DD). She needed gavage feeding for the first three months. The standard karyotyping revealed an apparently balanced translocation, 46,XX,t(7;10)(q31.3;q23.2). Her height was 95 cm (-3.3 SD), weight 13.6kg (-1.8 SD) at 5 7/12 years of age. She crawled at 3 years old and was able to walk with help at 5 years of age. She threatened self-injury. Together with severe intellectual disability (ID) and behavioral abnormalities as ASD, we performed cytogenetic microarray analysis, which revealed a de novo 4.3Mb deletion at 7q31, extending from position 122,292,088-126,610,540 (hg19). The deleted interval included CADPS2 (calcium-dependent activator of secretion) and GRM8 (glutamate receptor, metabotropic 8). Haploinsufficiency of CADPS2 is known to cause mild intellectual delay. While GRM8 is one of the CGH-specific CNVs that can drive attention deficit hyperactivity disorder (ADHD) etiology. However, more severe intellectual disability and autism phenotype observed in our patient is inconsistent with a dosage effects of the two genes. This combinatorial suppression of the 2 genes attribute to the etiology. These results suggest that synergy of haploinsufficiency of contiguous genes contributing neuronal development and functional network of neural activity has a crucial role in the severe phenotype of ASD and DD/ID.

1279S

Genome-wide gene expression analysis of identical twins discordant for autism spectrum disorder. *A. Saffari[1], C.C.Y. Wong[2], M. Arno[3], A. Ronald[1,2], L.C. Schalkwyk[2], J. Mill[2,4], R. Plomin[2], F. Dudbridge[5], E.L. Meaburn[1,2].* 1) Department of Psychological Sciences ,Birkbeck, University of London, London, UK; 2) King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, De Crespigny Park, London, UK; 3) Genomics Centre, School of Biomedical and Health Sciences, King's College London, London, UK; 4) University of Exeter Medical School, Exeter University, St Luke's Campus, Exeter, UK; 5) Faculty of Epidemiology and Population Health,London School of Hygiene and Tropical Medicine, University of London, London, UK.

Autism spectrum disorder (ASD) defines a collection of complex neurodevelopmental disorders, typified by impairments in social interaction and language. Multiple lines of evidence suggest a strong genetic basis to ASD. However, identical twins do not show complete concordance indicating that epigenetic and environmental (i.e. non-genetic) factors may also contribute to ASD risk.

This study aims to investigate the biological basis of non-genetic contributions to ASD risk by characterizing gene expression differences in ASD-discordant identical twins. Our sample consists of 53 sets of identical twins concordant (N = 6) and discordant (N = 36) for ASD and ASD traits, and concordant controls (N = 11) that were derived from the longitudinal, population-based Twins Early Development Study (Haworth, Davis, & Plomin, 2012). Genome-wide expression profiling of whole blood samples was initially performed using the Affymetrix Human Gene array, which was subsequently followed up with RNA-Sequencing (RNA-Seq) using the Illumina TruSeq kit, performed in order to gain a more detailed, higher-resolution profile of transcription. Whilst the array-based analysis failed to identify differentially expressed genes associated with the disorder that reached genome-wide significance, pathway-based analysis identified several pathways previously associated with ASD, these included nerve growth factor (NGF) signaling and glutathione-mediated detoxification pathways. Our current efforts are focused on analysis of the RNA-Seq dataset, investigating both differential expression and differential splicing.

To our knowledge, this study represents the largest of its kind to systematically investigate non-genetically mediated gene expression differences in ASD-discordant twins, the identification of which will help to shed light on the biological pathways dysregulated in ASD. DNA methylation (Illumina 27k Human Methylation array) and DNA variation (Illumina Human OmniExpressExome) data are also available for the same sample, and future work will be focused on the integration of these diverse and multi-dimensional datasets to help tease apart the relationship between genotype, epigenotype and phenotype.

1280M

A mouse model of 2q23.1 deletion syndrome implicates MBD5 in neuronal development. *J. Young[1], V. Camarena[1], L. Cao[1], C. Abad[1], Y. Toledo[1], M. Araki[2], K. Araki[2], K. Walz[1].* 1) Hussman Inst Human Genomics, Univ Miami, Miami, FL; 2) Kumamoto University, Kumamoto, Japan.

Individuals with 2q23.1 deletion syndrome have a range of well-defined behavioral abnormalities that include intellectual disability, motor delay, behavioral abnormalities and a distinctive craniofacial phenotype. Common to these patients is a partial or total deletion of a single gene, methyl-CpG binding domain protein 5 (*MBD5*), suggesting that haploinsufficiency of this gene is responsible for the phenotype. To confirm this hypothesis and to examine the role of MBD5 in vivo we have generated and characterized an *Mbd5* gene trap mouse model. Our analysis of heterozygous mutant mice revealed a striking similarity in the behavioral manifestations of mice compared with human 2q23.1 microdeletion carriers including abnormal behavior, cognitive impairment and motor and craniofacial abnormalities. In addition, cortical neuronal cultures allowed the detection of a deficiency in neurite outgrowth. Our study indicates that the $Mbd5^{+/GT}$ mouse model recapitulates most of the hallmark phenotypes observed in 2q23.1 deletion carriers. These findings support the causal role of MBD5 in 2q23.1 microdeletion syndrome and suggest a role for MBD5 in neuronal processes. The $Mbd5^{+/GT}$ mouse model will advance our understanding of the abnormal brain development underlying the emergence of 2q23.1 deletion-associated behavioral and cognitive symptoms.

1281T

A key role for TDP2 in neuronal development and maintenance. *J.H.M. Schuurs-Hoeijmakers[1,2,12], F. Gómez-Herreros[3,12], M. McCormack[4,12], M.T. Greally[5], S. Rulten[3], R. Romero-Granados[6], T.J. Counihan[7], E. Chaila[8], J. Conroy[9], S. Ennis[9], N. Delanty[4,8], F. Cortés-Ledesma[6], A.P.M. de Brouwer[1,2], G.L. Cavalleri[4], S.F. El-Khamisy[10,11], B.B.A. de Vries[1,2], K.W. Caldecott[3].* 1) Department of Human Genetics, Radboud University Medical Centre, Nijmegen, Gelderland, Netherlands; 2) Department of Cognitive Neurosciences, Donders Institute for Brain Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, The Netherlands; 3) Genome Damage and Stability Centre, School of Biological Sciences, University of Sussex, Sussex, UK; 4) Molecular and Cellular Therapeutics, The Royal College of Surgeons in Ireland, Dublin, Ireland; 5) National Centre for Medical Genetics, Our Lady's Children's Hospital, Crumlin, Dublin, Ireland; 6) Centro Andaluz de Biología Molecular y Medicina Regenerativa (CABIMER), Departamento de Genética, CSIC (Centro Superior de Investigaciones Científicas)-Universidad de Sevilla, Sevilla, Spain; 7) Department of Neurology, University Hospital Galway, Galway, Ireland; 8) Division of Neurology, Beaumont Hospital, Dublin, Ireland; 9) School of Medicine and Medical Science, University College Dublin, Dublin, Ireland; 10) Kreb's Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, UK; 11) Center of Genomics, Helmy Institute, Zewail City of Science and Technology, Giza, Egypt; 12) These authors contributed equally to this work.

Patients with heritable DNA repair deficiency syndromes often have increased susceptibility to cancer and pronounced neurological dysfunction. By using exome sequencing, we identified a homozygous truncating mutation, c.425+1GT>A, in the DNA repair gene *TDP2* in three brothers from a consanguineous family, as well as a second homozygous truncating mutation, c.413_414delinsAA, in an additional unrelated individual. The affected individuals were characterized by early appearance of seizures and were all intellectually disabled. They also developed progressive ataxia that eventually resulted in confinement to a wheelchair. This clinical picture indicates an important neurodevelopmental role for Tyrosyl DNA phosphodiesterase-2 (TDP2) and suggests that this enzyme is continuously required for neural maintenance.

TDP2 is an enzyme that repairs 'abortive' Topoisomerase II (TOP2)-induced DSBs. TOP2 removes torsional stress from DNA and facilitates gene transcription by introducing transient DNA double-strand breaks (DSBs). Such DSBs are normally rejoined by TOP2 but on occasion can become abortive and remain unsealed. Lymphoblastoid cells from affected individuals that lack TDP2 activity are indeed hypersensitive to TOP2-induced DSBs. Furthermore, loss of TDP2 inhibits TOP2-dependent gene transcription in cultured human cells and in mouse post-mitotic neurons. We also show that TDP2 is required for normal levels of many gene transcripts in developing mouse brain, including numerous gene transcripts associated with neurological function and/or disease, and for normal interneuron density in mouse cerebellum.

In summary, we have identified inactivating mutations in TDP2 in individuals with intellectual disability, seizures and progressive ataxia. Moreover, we show that TDP2 protects gene transcription from endogenous abortive TOP2 activity, including the transcription of many genes involved in neurological development or function. Collectively, these data identify TDP2-dependent DNA break repair as a critical guardian of gene expression and highlight abortive TOP2 activity as a threat to normal neuronal development and maintenance.

1282S

Diagnostic assessment using next generation sequencing in extremely heterogeneous neurodegenerative disorders, hereditary ataxia and spastic paraplegia. *Z. Iqbal[1], L. Pihlstrøm[1], A. H. Rengmark[1], S. Pilar Henriksen[1], S. L. K. Rydning[1,2], I. M. Wedding[1], J. A. Koht[3], C. Tallaksen[1,2], M. Toft[1].* 1) Department of Neurology, Oslo University Hospital, Oslo, Norway; 2) Faculty of Medicine, University of Oslo, Oslo, Norway; 3) Department of Neurology, Vestre Viken Hospital Trust, Drammen, Norway.

Hereditary ataxias (HA) and hereditary spastic paraplegias (HSP) are very heterogeneous and rare neurodegenerative disorders. Because of the extreme clinical and genetic heterogeneity of these conditions, the establishment of molecular diagnosis has been challenging, both regarding the cost and labor-intensive experimental designs. Recently, after the advent of next generation sequencing techniques, there are several opportunities to develop a single inexpensive test and easy workflows in different diagnostic settings. Here, we have ascertained 119 clinically well-characterized patients from all-over Norway, with HA (n=61), HSP (n=44), and a mixed phenotype of HA/HSP (n=14) patients. Polyglutamine expansions had been excluded in the affected probands with ataxia phenotypes. We have followed a cost-effective, fast, and robust procedure by using a Haloplex target enrichment kit (Agilent technologies, Santa Clara, CA), in order to screen coding sequences of 91 genes implicated in these conditions, of which 89 well known and two candidate genes were included in the design. DNA from ten individuals was pooled and a single barcode was assigned, and subsequently all 12 pools were sequenced by using a single lane of Illumina HiSeq, which resulted in high quality data with an average 98.95% bases covered >80x in the targeted intervals. Our bioinformatics pipeline and criteria to assign plausible pathogenicity selected about 50 candidate variants for further validation and confirmation. We identified known pathogenic mutations and novel potential variants in several genes, including *ANO10*, *BEAN1*, *BSCL2*, *CYP2U1*, *GRID2*, *KCND3*, *KIF1A*, *KIF5A*, *PRKCG*, *REEP1*, *SPG7*, *SPG11*, and *SYNE1*. After the ongoing validation analysis, we will be able to establish the diagnosis in a certain subset of patients, in both HA and HSP. With this approach, it will be possible to estimate mutation frequencies in the involved genes. Moreover, in a heterogeneous group of disorders, such diagnostic testing will reduce the total duration of eventual correct diagnosis after the first appearance of the patient to the clinic. In conclusion, our data will have extended implications in the field of clinical genetics and development of diagnostic procedures.

1283M

Identifying biomarkers in chronic neuropathic pain. *P.C McHugh[1], D.P Finn[2], R. Airley[1], D.A Buckley[1].* 1) Pharmacy, University of Huddersfield, Huddersfield, WE., United Kingdom; 2) Pharmacology & Therapeutics, National University of Ireland, Galway, Ireland.

Chronic pain, including neuropathic pain, is poorly understood. It is estimated that almost 1 in 5 adults experience moderate-severe pain and often receive inadequate treatment. With a growing ageing population, this is fast becoming a significant healthcare issue that requires urgent attention. Blood biomarkers can provide us with a tool for early diagnosis and in turn facilitate improved treatment strategies. We wish to identify pain-related molecules, systems or pathways (biomarkers) in human plasma with the aim of improving our understanding, diagnosis and treatment of neuropathic pain. To identify novel biomarkers that can discern neuropathic pain patients from healthy controls we are using a neuropathic pain-control pilot sample (n= 20. For this we are employing a combination of quantitative PCR, PCR arrays and Affymetrix array systems to analyse gene expression (RNA), and HPLC/Mass Spectrometry for plasma analyses. Our preliminary data identifies several molecules that can differentiate between neuropathic pain and healthy controls. To validate our findings we are currently developing two independent neuropathic pain cohorts (case-control); one at NUI Galway and the other through the Leeds Pallium Research Group. These molecules could be developed into a diagnostic test that can differentiate neuropathic pain subtypes, allowing clinicians to better treat this condition.

1284T

Genetic diagnosis of neurological diseases using NGS - Report of 48 cases. *D. Garcia, A. Romera-Lopez, D. Valero-Hervas, V. Felipe, G. Hernandez, C. Collado, R. Rodriguez-De Pablo, A. Arilla-Codoner, J.C. Trivino, S. Zuniga-Trejos, G. Marco, D. Cantalapiedra, L. Perez-Carbonero, C. Buades, C. Camprubi, R. Minambres, S. Santillan.* Sistemas Genomicos, Paterna, Valencia, Spain.

The diagnosis of neurological diseases is often not straightforward, mainly due to the presence of nonspecific overlapping clinical symptoms and of genetic heterogeneity. A global approach using high throughput technologies such as NGS could be of great assistance to improve their diagnosis. We present our experience in the analysis of neurological diseases using a 200-gene NGS targeted resequencing panel. Samples from 48 patients were submitted to our laboratory for genetic testing. Clinical indications included Spastic Paraplegia (17), Motor/Sensitive Hereditary Neuropathy (12), Ataxia (6), Parkinson (4), Miastenia (3), Joubert Syndrome (2), Frontotemporal Dementia (2), Hyperkalemic Periodic Paralysis (1) and SMA (1). Coding exons and splice-site regions of the genes associated with each pathology were analysed. Enrichment and sequencing were carried out using SureSelect Enrichment System (Agilent) and SOLiD 5500/MiSeq (Life Technologies/ Illumina). Mean depth was established at 200x. In 7 out of 48 samples (14%), a genetic cause that confirmed the clinical diagnosis was found. Taking into account all the samples, 63 variants (confirmed by Sanger sequencing) were identified and classified into pathogenic (7), probably pathogenic (2), unknown (53) and probably benign (1). According to their effect, missense (29), synonymous (21), splice-site (8), frameshift (3) and nonsense (2) mutations were found. In 21 out of 48 samples, no variants (pathogenic/unknown) were identified. Despite being moderate, 14% of genetically diagnosed patients suggests that NGS is effective for the diagnosis of neurological diseases. A better understanding of unknown significance variants, that would allow classifying them into SNPs or pathogenic mutations, could significantly increase the diagnosis rate.

1285S

Transcriptome analysis of distinct mouse strains reveals kinesin light chain-1 splicing as an amyloid beta pathology modifier in Alzheimer's disease: A mouse-to-human translational approach to complex diseases. *T. Morihara[1], M. Sato[1], M. Silverman[3], H. Akatsu[2], N. Hayashi[1], M. Yokokoji[1], K. Yanagida[1], R. Hashimoto[1], H. Yamamori[1], T. Tsunoda[4], Y. Yamaguchi[5], K. Kamino[1], M. Takeda[1].* 1) Psychiatry, Osaka University, Suita, Osaka, Japan; 2) Choju Medical Institute, Fukushimura Hospital, Toyohashi-shi, Aichi, Japan; 3) Department of Biological Sciences, Simon Fraser University, Burnaby, Canada; 4) Center for Genomic Medicine, RIKEN, Yokohama, Japan; 5) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Sendai, Japan.

[Background] Amyloid beta (Aβ) is the central pathology of Alzheimer's disease (AD). The genes controlling Aβ accumulation in the brain of sporadic AD are largely unknown. Human genetic studies require huge research resources and do not reveal the role of the identified genes. To overcome the difficulties of human genetic study, we combined distinct mouse strains with transcriptomics to directly identify Aβ modifier genes.

[Methods and Results] We have observed that APP Tg mice with DBA/2 genetic backgrounds have significantly lower levels (-74.7 to 57.7%, p<0.0001-0.0002) of Aβ compared with SJL and C57BL/6 mice (Morihara PNAS 2014). These data suggested that some gene(s) in DBA/2 drastically suppress Aβ accumulation. To avoid detecting secondary affected genes by Abeta, we used non-Tg mice in the absence of Aβ accumulation and performed the first transcriptome analysis to select candidate genes differentially expressed in DBA/2 mice. The 2nd transcriptome analysis using APP-Tg mice with mixed genetic backgrounds from these three strains revealed kinesin light chain-1 (Klc1) as an Aβ modifier. The following QPCR measurement confirmed the array data and suggested that Klc1 splice variant E (Klc1E) control the accumulation levels of Aβ in mouse brain. The expression levels of Klc1E were significantly correlated with the levels of Aβ (R2= 0.21 to 0.39, p<0.0001-0.0002) in these mice. We expanded mouse strain analysis by adding two mouse strains. Three strains with higher Aβ accumulation shared common Klc1 allele and had higher expression levels of Klc1E. In contrast two strains with lower Aβ shared another common Klc1 allele and had lower levels of Klc1E. In human, KLC1E levels in brain (AD: n=10. Cont: n=14) and lymphocyte (AD: n=47, Cont: n=17) were significantly higher in AD patients compared to controls. Functional analysis using neuroblastoma cells showed that knockdown of KLC1E decrease the production of Aβ (-39.3 to -44.7%, p<0.0001). Taken together, mouse transcript, mouse genetic, human transcript and functional analysis strongly suggested that KLC1E control Aβ pathology in AD.

[Conclusion] The expression levels of KLC1E control Aβ pathology. The unique combination of distinct mouse strains and model mice with transcriptomics is useful for the study of genetic mechanisms of complex diseases.

1286M

Circadian Network and Autism: Role of the JARID1 Genes. *Z. Talebizadeh[1], A. Shah[1], D. Kalinowska[2], L. DiTacchio[2].* 1) Pediatrics, Children's Mercy Hospital, and University of Missouri-Kansas City School of Medicine, Kansas City, MO; 2) University of Kansas Medical Center, Kansas City, KS.

The circadian clock coordinates diverse cellular processes and functional outcomes, including behavior and cognition. Abnormalities in the clock genes may have a role in autism, but the underlying mechanism remains unknown. Members of the JARID1 gene family (JARID1a-d), histone demethylases, have been shown to be involved in the circadian molecular machinery in a recent work of our collaborator. JARID1a activates CLOCK-BMAL1, whereas JARID1b and JARID1c act as repressors. These opposing roles suggest that the optimal ratio among JARID1 isoforms is vital in maintaining the proper function of the circadian system. Imbalance in these isoform ratios may contribute to the etiology of diseases. Furthermore, miR132 orchestrates translational control of the circadian clock by targeting chromatin remodeling genes, including MECP2 and JARID1a. The splicing profiles of JARID1, and factors that regulate the expression of such isoforms, are not fully known. Our hypothesis is that JARID1 mis-splicing may present in at least a subset of subjects with autism. Recent findings have shown cross links between DNA methylation and gene regulatory processes, including alternative splicing and miRNA. Therefore, we used data from a DNA methylation marker for subject stratifications. We evaluated the expression level of multiple JARID1 alternative splicing transcripts in autistic subjects stratified based on the absence or presence of a given DNA methylation-related marker (DM) in lymphoblastoid cell line-derived RNAs, using Exon array profiling, TaqMan assays, followed by DNA sequencing. A distinct pattern was detected in the expression level of alternatively spliced JARID1 isoforms for autistic subjects with DM compared to those without DM and controls. Additional experiments, including miRNA mimics, are underway to further characterize the role of miR132 in regulating JARID1a by finding which isoforms(s) show miR132-dependent expression. This is the first study to evaluate a clock gene in autism, at the alternative splicing level in conjunction with DM markers. Our study indicates that (1) JARID1 undergoes complex splicing resulting in multiple splice variants, and (2) subject stratification using DM may assist with finding a more homogeneous subset to better understand the underlying mechanisms involved in autism. Funding: This study was supported by Patton Trust-KCALSI grant.

1287T

RNA-sequencing and gene co-expression analysis identifies novel genes and pathways in bipolar disorder. *N. Akula[1], J.R Wendland[1], K.H Choi[2], B.K Lipska[3], J.E Kleinman[4], F.J McMahon[1].* 1) Human Genetics Branch, National Institute of Mental Health (NIMH), National Institutes of Health (NIH), Bethesda, MD 20892; 2) Dept of Psychiatry and Program in Neuroscience, Uniformed Services University of the Health Sciences, Center for the Study of Traumatic Stress, Bethesda, MD 20814; 3) Human Brain Collection Core, NIMH/NIH, Bethesda, MD 20892; 4) Lieber Institute for Brain Development, Baltimore, MD 21205.

Bipolar disorder (BP) is a highly heritable psychiatric illness affecting ~2% of the population. Genome-wide association studies (GWAS) have consistently identified several common variants associated with BP. Less consistent results have emerged from gene expression studies, partly due to limited sample sizes. Weighted gene co-expression network analysis (WGCNA) may be a more powerful approach, since it exploits patterns of co-expression among genes, but the approach has not yet been applied to RNA-seq data, where gene expression can be measured precisely over a wide dynamic range. We generated deep RNA-seq data from post-mortem dorsolateral prefrontal cortex obtained from a total of 11 BP cases and 11 age and sex-matched controls, sequenced in two batches (Akula et al, 2014). WGCNA identified 40 coexpression modules, of which 9 were preserved between the batches (Zsummary > 10). Six of the preserved modules were significantly associated with BP (meta-p value < 0.05). Functional analysis in DAVID of genes within these 6 preserved modules revealed significant enrichment for several GO terms, including synapse, neuron projection, postsynaptic density, ion transport, and ATP binding, among others. Five of these modules were also significantly enriched for differentially expressed genes identified in Akula et al, 2014 (hypergeometric pvalue < 0.05), and 3 modules were significantly enriched (hypergeometric p-value < 0.05) for genes associated with BP in a previous GWAS (Chen et al 2013). Thus two modules were preserved between batches, associated with BP, enriched for genes known to be differentially expressed in BP, and enriched for prior BP GWAS hits. Some of the co-expressing hub genes in these modules include DCC, CAMK2D, KCNQ3 and NPAS3, which have been previously implicated in BP and related disorders. These results demonstrate that WGCNA can be applied to RNA-seq data with good results. Applied to BP, WGCNA produced results that were broadly consistent with prior findings and implicated additional genes and pathways that may contribute to risk for BP. This approach may help unify GWAS and gene expression results to suggest biological hypotheses deserving of additional study.

1288S

GluD1 is over-expressed in iPSC-derived FOXG1 neurons: a potential common therapeutic target for Rett syndrome. *S. Amabile[1], T. Patriarchi[1,2], A. Bartolini[1], M.G. Lolli[1], D. Yasui[3], C. Lo Rizzo[1], F. Ariani[1,4], F. Mari[1,4], M.A. Mencarelli[1,4], J.W. Hell[2], I. Meloni[1], A. Renieri[1,4].* 1) Medical Genetics, Univerity of Siena, Siena, IT., Si, Italy; 2) Department of Pharmacology, School of Medicine, University of California Davis, Davis, CA, United States; 3) Department of Medical Microbiology and Immunology, University of California Davis, Davis, CA, United States; 4) Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy.

Rett syndrome is due to de novo mutations in MECP2, CDKL5 or FOXG1 genes. In spite of their involvement in the same disease, a functional interaction between the three genes has not been proven and disease mechanisms remain elusive. MeCP2 and FoxG1 are transcriptional regulators; CDKL5 encodes for a kinase protein involved in multiple cellular processes including gene expression. We hypothesized that mutations in the three different genes may lead to similar phenotypes by deregulating expression of common genes. To verify this hypothesis, we have used a human model based on induced pluripotent stem (iPS) cells-derived neurons from MECP2, CDKL5 and FOXG1 patients. We previously demonstrated by quantitative RT-PCR and protein analysis an over-expression of GRID1 in both iPSC-derived MECP2 and CDKL5 neuronal precursors and neurons (Livide Eur J Hum Genet 2014). We demonstrate here that GRID1 is overexpressed also in iPSC-derived FOXG1 neuronal precursors and neurons. GRID1 encodes GluD1, a member of the delta family of ionotropic glutamate receptors that is located on the post-synaptic site and induces inhibitory or excitatory presynaptic differentiation depending on the specific brain region. GluD1 overexpression may thus favor either glutamatergic or GABAergic maturation in different brain areas and neuronal subtypes. We therefore explored the expression of excitatory/inhibitory markers. We indeed found the expected imbalance but in opposite directions. Quantitative RT-PCR on iPSCs-derived neurons mutated in CDKL5 demonstrated that the glutammatergic markers VGLUT1 and VGLUT2 are over-expressed while the GABAergic markers GAD1 and GAD2 are down-regulated. On the contrary, iPSCs-derived FOXG1 neurons demonstrated a down-regulations of the excitatory markers and an up-regulation inhibitory markers. The different imbalance found is consistent with the specific endophenotype of the patients from whom the iPSCs are derived: the FOXG1 patient does not manifest epilepsy while the CDKL5 patient has a severe epilepsy not controlled by therapy. In conclusion, at present GluD1 overexpression is the only common proved altered marker during neuronal differentiation of iPSC derived from patients mutated in MECP2, CDKL5 and FOXG1. Modulation of the basic defect by inhibition of GLUD1 could be thus explored as a therapeutic tool in both typical Rett and its variants.

1289M

Phenotypic, molecular, functional and structural analysis of new DCX and LIS1 mutations causing the subcortical band heterotopia/lissencephaly spectrum. *D.R. Amrom[1, 3, 4, 8], G. Brouhard[6, 8], S. Bechstedt[6, 8], K. Topropova[9], D. Melancon[2, 3, 7, 8], D. Tampieri[2, 3, 7, 8], S. Reck-Peterson[9], E. Andermann[1, 3, 4, 5, 8].* 1) Neurogenetics Unit; 2) Department of Neuroradiology; 3) Montreal Neurological Hospital and Institute; 4) Department of Neurology & Neurosurgery; 5) Department of Human Genetics; 6) Department of Biology; 7) Department of Radiology; 8) McGill University, Montreal, QC, Canada; 9) Department of Cell Biology, Harvard Medical School, Boston, MA, US.

Subcortical band heterotopia (SBH), also called double cortex syndrome, and lissencephaly (LIS) are part of a spectrum of malformations of cortical development due to deficient neuronal migration, referred to as the SBH/LIS spectrum. Most patients have either DCX or LIS1 gene mutations, associated with predominant anterior or posterior distribution of SBH, respectively. We report the phenotypic, molecular, functional, and structural analysis of new DCX and LIS1 mutations causing SBH/LIS. Patient 1 is a 46-year-old woman of French-Canadian ancestry who had onset of epilepsy at 3 months of age. She had multiple types of seizures and a diagnosis of Lennox-Gastaut syndrome refractory to antiepileptic medication. Brain MRI showed double cortex predominating in the frontal regions. DCX sequencing showed a c.578delA variant. The parents were not available for genetic testing. Using in vitro assays, we measured the ability of recombinant variant DCX protein to interact with microtubules. The c.578delA variant was found to be defective in promoting microtubule nucleation and polymerization, and showed impaired cooperative binding to microtubules. Patient 2 is a 28-year-old man of British ancestry. At 6 months of age he presented mild developmental delay and his first seizures. He had multiple types of seizures, most of them generalized, that were refractory to antiepileptic medication. He has pseudobulbar signs and severe developmental delay. Brain MRI showed a predominantly posterior lissencephaly associated with partial callosal agenesis and a cavum septum pellucidum, as well as diffuse cerebellar atrophy. LIS1 sequencing revealed duplication of five nucleotides in exon 8 (c.728_732dupATCAA). The parents declined genetic testing. Bioinformatic analysis of the mutant sequence and mapping onto the LIS1 structure showed that the change introduces a five residue stretch of altered sequence followed by a premature stop codon at residue 250, early in the 4th WD repeat of the Lis1 beta propeller. We report two new pathogenic variants causing severe phenotypes of the SBH/LIS spectrum. Our functional analyses show that the DCX variant disrupts microtubule binding as well as the cooperative interaction between DCX molecules. Based on our structural interpretation of the LIS1 variant, it is very likely that the LIS1 protein does not fold properly, is unable to bind dynein, and is likely targeted for degradation in cells.

1290T

***De novo TBR1* mutations in sporadic autism disrupt protein functions.** *P. Deriziotis[1], B.J. O'Roak[2,3], S.A. Graham[1], S. Busquets-Estruch[1], D. Dimitropoulou[1], R.A. Bernier[4], J. Gerdts[4], J. Shendure[2], E.E. Eichler[2,5], S.E. Fisher[1,6].* 1) Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; 2) Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA; 3) Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, Oregon, USA; 4) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA; 5) Howard Hughes Medical Institute, Seattle, Washington, USA; 6) Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands.

Next-generation sequencing recently revealed that recurrent disruptive mutations in six genes - *CHD8*, *DYRK1A*, *GRIN2B*, *PTEN*, *TBR1*, and *TBL1XR1* - may account for 1% of sporadic autism cases. *TBR1* is of particular interest because it encodes a neuron-specific transcription factor of the T-box family, with established roles in patterning of the central nervous system, including regulation of neuronal identities during cortical development. The recurrence of *de novo TBR1* mutations in sporadic autism spectrum disorders suggests that the identified mutations are likely to be pathogenic. Nonetheless, functional experiments in model systems are essential to determine the precise effect of mutations on protein function and provide insight into the molecular mechanisms of the disorder. Here we report the first functional analyses of *TBR1* variants identified in sporadic autism. *De novo* truncating (p.A136PfsX80 and p.S351X) and missense (p.K228E and p.N374H) mutations disrupt multiple aspects of TBR1 function, including subcellular localization, interactions with co-regulators and transcriptional repression. Missense mutations (p.Q178E, p.V356M, p.Q418R and p.P542R) inherited from unaffected parents did not disturb function in our assays. Our findings support the hypothesis that *de novo* mutations in sporadic autism have severe functional consequences and highlight the power of coupling novel genetic data to empirical assays of protein function in order to illuminate crucial molecular networks involved in complex neurodevelopmental disorders such as autism.

1291S

Analysis of actin cytoskeleton dynamics in stem cells from autistic patients. *K. Griesi-Oliveira[1,2], M. Suzuki[1], D.Y. Sunaga[1], A.L. Sertie[2], M.R. Passos-Bueno[1].* 1) Biosciences Institute - University of Sao Paulo, Sao Paulo, SP, Brazil; 2) Instituto Israelita de Ensino e Pesquisa - Hospital Israelita Albert Einstein.

Autism spectrum disorders (ASD) are are genetically heterogeneous diseases, which has been made difficult the understanding of their etiology. However, several genes implicated in the etiology of ASD are part of common molecular pathways, indicating these different genetic alterations may cause similar effects during neurogenesis. One of such common mechanisms may be the cytoskeleton dinamics regulation, which is essential for organization of dendritic spines and axonal growth and guidance. However, the relationship between these mechanisms and the etiology of ASD has been poorly explored in the literature, specially in their functional aspects. Previous results from our group have suggested a disturbance in cytoskeleton dynamics in cells from autistic patients. Thus, here we aimed to investigate the actin cytoskeleton dinamics regulation in stem cells from human exfoliated deciduous teeth (SHED) from ASD idiopathic patients. First, we incubate the cells of 13 patients and 8 controls for 24h using Rho kinase inhibitor (ROCKi) in order to depolymerize the microfilaments. To induce actin reconstruction, ROCKi was washed out and cells were treated with drugs that activate specifically Rac2, Cdc42 or RhoA. The percentage of cells presenting actin filaments were counted at 15, 30, 45 and 60 minutes after drugs application. Our results showed that 5 out of the 13 patients presented a significant lower percentage of cells with recovered actin filaments at all time points compared to controls when treated with Rac2 or cdc42 activator, and among these, 2 patients respond abnormally to RhoA activation (unpaired t-test $p<0.05$). We also treated the cells with a drug that directly activates the three RhoGTPases, making them constitutively activated. Upon this treatment, 4 patients responded abnormally compared to controls. Interestingly, these are the same patients that had responded abnormally to the specific activactors. We tested all the five patients that presented abnormal actin reconstitution for expression of Rac2, cdc42 and RhoA and we found that one of them has lower expression of these three RhoGTPases. These results suggest that, at least for a group of ASD patients, the dynamics of actin polymerization is slower compared to control individuals and that for one of them this could be due to lower expression of RhoGTPases. We believe that this study can contribute for the understanding of the common molecular mechanisms involved in ASD etiology.

1292M

Transcription and methylation reveals microglia related and non-coding RNA networks specifically altered in Dementia with Lewy Bodies. *C. Humphries[1,2], M.A. Kohli[1], P. Whitehead[1], G. Beecham[1], D.C. Mash[3], M.A. Pericak-Vance[1], J. Gilbert[1].* 1) Hussman Inst Human Genomics, Univ Miami, Miami, FL; 2) John T. MacDonald Department of Human Genetics, Univ Miami, Miami, FL; 3) Department of Neurology, Univ Miami, Miami, FL.

Dementia with Lewy Bodies (DLB) is the second most common dementia. To understand the molecular processes in DLB, we investigated transcription and DNA methylation in human post-mortem brain tissue. Transcription was examined using total RNA-seq permitting the identification and quantitation of known genes and non-coding RNAs. DNA methylation was examined using Illumina's Infinium HumanMethylation450 BeadChip. Neuropathological specimens were from the temporal poles (BA 38) of ten age, sex and race-matched cases each of DLB, late-onset Alzheimer's disease (LOAD) and normal controls. RNA was extracted using Qiagen's miRNeasy kit and libraries prepared with Epicentre's Script-Seq protocol. Samples were run on Illumina's HiSeq2000, generating 40-65 million reads per library. The program DESeq2 was used to examine transcriptional differences between Gencode genes. Transcriptional analysis between DLB and Controls revealed 4,770 genes out of a total of 30,929 genes (Gencode) to have expression differences ($p<0.05$). Of these, 469 were significant after controlling for multiple testing using False Discovery Rate (FDR<0.05). Subsequent comparison of these 469 DLB genes to LOAD revealed seven non-coding RNAs specifically altered. To find cellular processes disrupted in DLB, we performed network analysis using weighted gene co-expression network analysis (WGCNA). We found the 4,770 genes with expression differences between DLB and CON formed 13 networks. Eight networks were highly correlated to both DLB and LOAD. The other five networks were specific to DLB. Interestingly, the networks that correlated to both DLB and LOAD were primarily made up of protein coding genes. Non-coding RNAs primarily made up the networks specific to DLB. Using DAVID, we found the gene networks correlated to both DLB and LOAD are involved in acetylation. Genes in networks specific to DLB are mainly involved in mesoderm development ($p=2.2E-2$). Genes involved in mesoderm development give rise to the hematopoietic stem cells that generate the immune cells, microglia, that are enriched in DLB brains compared to LOAD and control brains. Assessment of DNA methylation within these networks revealed that DLB samples have significant hypermethylation. The study of this network of genes with altered expression and methylation specific to LOAD may offer a fruitful approach to advancing our understanding the role that non-coding RNAs and genes play in the etiology of Dementia with Lewy Bodies.

1293T

Loss-of-function mutations of progranulin (PGRN) in siblings with familial FTLD. *E. Vitale[1], S. Napoletano[1], S. PappatÃ [3], A. Postiglione[4], M.T. Gentile[5], L. Colucci-D'Amato[5], P. Sorrentino[2], G. Milan[2].* 1) Institute of Protein Biochemistry (IBP), CNR-National Research Council, Pozzuoli (NA), Napoli, Italy; 2) Geriatric Center "Frullone" ASL Napoli 1 - Naples-Italy; 3) Institute of Bioimaging and Biostructures , CNR, Naples-Italy; 4) University of Naples "Federico II", Naples-Italy; 5) Experimental Neuropatology, SUN Caserta-Italy.

Mutations in the progranulin (PGRN) gene, located on chromosome 17q21.32, have been linked to familial Fronto-Temporal Lobar Degeneration (FTLD) dementia. Loss-of-function mutations of PGRN represent a frequent cause of middle-age cognitive impairment characterized by selective neurodegeneration of the prefrontal and anterotemporal cortices and in a few select cases, have been identified as the cause of primary progressive aphasia (PPA) in women. Although it is well known that PGRN mutations induce haploinsufficiency of this protein, the pathogenic mechanisms by which PGRN deficiency leads to brain dysfunction and damage remain unclear. We analyzed the PGRN genes in two siblings ascertained as having FTLD, III-2H72, a male and III-3H73, a female. The two affected siblings exhibited FTLD according to the Lund-Manchester criteria. To carry out the analysis, we designed primers to amplify all exons and intron-exon boundaries of the PGRN gene and sequenced the amplified fragments using the Sanger method. We found a g.101349_101355delCTGCTGT deletion in exon 6 CDS. The mutation we identified results in a premature stop codon with a frameshift error and mRNA non-sense mediated decay. To exclude the possibility that this was a common mutation, we design an ARMS-PCR test for this mutation and check it in about 250 healthy controls matched by age, sex and geographic regions. We found no deletions carried by any of these individuals, providing specificity to the association we identified between the disease and the deletion. A subsequent CGH assay performed on the DNA from the two patients demonstrated no additional evident genomic alterations. Differential gene expression measurements revealed a significant downregulation of PGRN in the two individuals carrying the deletion when compared to controls. In addition, the correlation between PGRN mRNA expression levels and cognitive impairment was statistically significant in FTLD patients. Furthermore, PGRN protein analyses demonstrated a PGRN protein deficiency, corroborating the defective PGRN gene expression. These results strongly suggest that PGRN could be the causative disease gene in this familial FTLD.

1294S

Significant Enrichment of Disease-specific Polymorphisms surrounding microRNAs suggests further involvement in Schizophrenia and Bipolar Disorder. *V. Williamson, M. Mamdani, G. McMichael, S. Bacanu, V. Vladimirov.* Psychiatry, Virginia InstPsychiatric & Behavioral Genetics, Richmond, VA.

Schizophrenia (SZ) and bipolar disorder (BD) are debilitating neuropsychiatric disorders with substantial impact on the quality of human life. Since 2006, numerous genome-wide association studies (GWAS) have been performed to identify loci associated with SZ and BD disease risk. In one of the largest GWAS to date, performed by the Psychiatric Genetic Consortium (PGC), the most significant association signal originated from a microRNA locus (hsa-miR-137) which has been previously implicated in adult neurogenesis and neuronal maturation. Moreover, bioinformatically predicted and experimentally validated gene targets of hsa-miR-137 were also associated with SZ and BD (at genome-wide significance), suggesting a novel, miRNA-mediated mechanism of disease etiology. These results, coupled with the increasingly recognized neurodevelopmental role of miRNAs, have prompted us to investigate the genetic enrichment of disease-associated polymorphisms affecting miRNA genes. Using complementary statistical approaches (Simes and the Sum of Squares (SST) tests), we have detected robust, significant disease enrichment of 1367 LD-independent (r2 >0.8) SZ and BD related polymorphisms surrounding (+/- 1MB) miRNAs in the most recent follow-up PGC GWAS findings. In order to identify the putative regulatory relationships between these enriched variants and their cognate miRNAs' functions, we performed expression quantitative trait locus (eQTL) analysis in postmortem brains (prefrontal cortex, Brodman Area 46) from the Stanley Medical Research Institute (N=27 SZ, 29 BD, 22 Controls). We detected 23 significant cis and 9 trans eQTLs associated with our miRNAs (False Discovery Rate <0.05). In our study, the most significant eQTL, rs3733047 (A/G, Chr3: 52871929), was associated with hsa-miR-135a expression, which demonstrated a clear temporal expression during normal brain development in a recent study. Rs3733047 exerted allelic effects on hsa-miR-135a expression by creating an additional putative binding site for the transcription factor, GATA2. GATA2 regulates postmitotic neuronal differentiation and its expression is significantly downregulated (mRNA and protein) in postmortem prefrontal cortex of SZ subjects. Taken together, the significant disease enrichment of variants surrounding miRNAs and the subsequent identification of miRNA eQTLs suggest a larger role for miRNA dysregulation in the etiology of Schizophrenia and Bipolar Disorder.

1295M

The transcriptome of 16p11.2 syndrome patients uncovers a link between autism and ciliopathies. *A. Reymond[1], E. Migliavacca[1,2], C. Golzio[3], K. MÃnnik[1], I. Blumenthal[4], E.C. Oh[3], L. Harewood[1], J. Kosmicki[5,6], M.N. Loviglio[1], L. Hippolyte[7], A.M. Maillard[7], M.M. van Haelst[8], J. Andrieux[9], J.F. Gusella[4,10], M.J. Daly[5,6], J.S. Beckmann[2,7,11], S. Jacquemont[7], M.E. Talkowski[4,10], N. Katsanis[3], The 16p11.2 European Consortium.* 1) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland; 3) Center for Human Disease Modeling and Department of Cell biology, Duke University, Durham, North Carolina 27710, USA; 4) Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 5) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114; 6) Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA; 7) Service of Medical Genetics, Lausanne University Hospital (CHUV), Lausanne, Switzerland; 8) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Netherlands; 9) Institut de Génétique Médicale, CHRU de Lille - Hôpital Jeanne de Flandre, Lille, France; 10) Departments of Genetics and Neurology, Harvard Medical School; 11) Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland.

Hemizygosity of the 16p11.2 ~600 kb BP4-BP5 region (29.5 to 30.1Mb) is one of the most frequent known genetic etiology of autism spectrum disorder (ASD). It is also associated with a highly penetrant form of obesity and a significant increase in head circumference. Mirror phenotypes are observed in carriers of the reciprocal duplication, who present a high risk of being underweight, microcephalic and/or developing schizophrenia. We profiled the transcriptome of individuals carrying reciprocal CNVs in 16p11.2 and analyzed the data using a gene dosage model. The genome-wide transcript perturbations correlated with clinical endophenotypes of this CNV and were enriched for genes associated with ASD. We also uncovered a significant correlation between 16p11.2 copy number changes and expression levels of genes mutated in ciliopathies. This result was replicated in orthologous mouse models, raising the possibility that ciliary dysfunction might contribute to 16p11.2 pathologies. In support of this correlation, we found structural ciliary defects specifically in the CA1 hippocampal region of a mouse model duplicated for the 16p11.2 orthologous region. Moreover, using an established zebrafish model we show strong genetic interaction between KCTD13, a key driver of the mirrored neuroanatomical phenotypes of the 16p11.2 CNV and ciliopathy-causing genes. Strikingly, overexpression of BBS7 rescues the head size and neurogenesis defect of kctd13 morphants, while suppression or overexpression of CEP290 and KCTD13 rescues reciprocally KCTD13-induced phenotypes. Our findings are likely relevant in humans as a patient carrying a de novo 16p11.2 deletion and a heterozygous null CEP290 allele was negative for the majority of 16p11.2-associated pathologies. Taken together, our data suggest that dysregulation of ciliopathy genes, and hence ciliary dysfunction, likely contributes to the clinical phenotypes of this CNV, and possibly to ASD, and that transcripts overexpressed by 16p11.2 deletions or duplications might represent therapeutic targets.

1296T

Transcriptome Profiling and Behavioral Analysis of a VIPR2-CNV Mouse Model of Schizophrenia. *T. Chapman[1], A. Noor[1], A. Halberstadt[1], D. Malhotra[1], M. Geyer[1], J. Sebat[1,2,3].* 1) Department of Psychiatry; 2) Department of Cellular and Molecular Medicine; 3) Institute of Genomic Medicine, University of California San Diego, La Jolla, CA.

Rare copy number variants (CNVs) are risk factors for schizophrenia and have the potential to reveal new biological mechanisms and pharmacological targets. We previously demonstrated the genetic association of schizophrenia with a novel microduplication on chromosome 7q36.3, centered on the gene VIPR2. This gene encodes a class II GPCR (VPAC2) that is coupled to cAMP signaling and is expressed throughout the brain. To elucidate possible effects of increased VIPR2 dosage on brain and behavior, we utilized a transgenic mouse model that contains multiple copies of a YAC-inserted human VIPR2 and its surrounding genomic region. We determined by qPCR that human and endogenous mouse VIPR2 share a similar pattern of gene expression in brain, with low to moderate expression in frontal cortex and hippocampus and higher expression in thalamus. Because increased VIPR2-cAMP signaling could alter CREB transcription, we analyzed the effect of VIPR2 overexpression on transcriptome profiles from four brain regions using RNA sequencing. In parallel, VIPR2 transgenic mice were characterized on a number of behavioral tasks with translational relevance to schizophrenia, including T-maze, fear conditioning and extinction, prepulse inhibition, and locomotor assays.

1297S

Gene expression and neuronal morphology in differentiating human induced Pluripotent Stem Cells (iPSCs) from individuals with chromosome 15q11.2 deletions. *D.K. DAS[1,2], K. Chowdari[1], C. Celik[3], L. D'Aiuto[1], W. Joel[1], A. Ghosh-Bhattacharjee[1], V.L. Nimgaonkar[1].* 1) Department of Psychiatry, WPIC, University of Pittsburgh, Pittsburgh, PA. United States; 2) Genetic Research Centre, National Institute for Research in Reproductive Health, JM Street, Mumbai, India; 3) GATA School of Medicine, 06018 Ankara, Turkey.

Background: Copy number variants at 15q11.2 are associated with intellectual disability, autism and schizophrenia (SZ). This region, spanning BP1-BP2 encodes *CYFIP*, *NIPA1*, *NIPA2* and *TUBGCP5*. From animal models, all four genes are important for brain morphogenesis. As these genes have not been investigated in human neurodevelopment, we have analyzed gene expression and neuronal morphology in differentiating iPSCs from individuals with 15q11.2 deletion. **Methods**: The proband and his mother, both bearing a deletion at chr 15q11.2 were compared with an unrelated control without deletion. The proband was also diagnosed with SZ, but not his mother. The deletions were identified using qPCR and array-CGH. Fibroblasts derived from skin biopsies were reprogrammed with Sendai viral vector expressing transcription factors Sox2, c-Myc, Klf4 and Oct4. Following quality control, the iPSCs were differentiated into neural rosettes and neuronal progenitor cells (NPCs) and then to glutamatergic neurons. Gene expression was analyzed using qPCR and normalized against beta actin. To evaluate dendritic spine morphology, iPSCs derived neurons were transfected with pEGFP vector and immunostained with human anti-CYFIP antibody **Results**: In fibroblasts, reduced expression of all 4 genes was noted (FC: 0.2-0.3). The expression levels were generally higher in iPSCs, however, fold changes are still reduced for these genes: *CYFIP* (0.78), *NIPA2* (0.67), *TUBGC5* (0.59). Expression increased further in NPCs, but the expression of *NIPA1* was remarkably variable. Since, CYFIP interacts with FMRP, expression of *FMRP* gene was analyzed. *NIPA1* and *CYFIP* expressions are correlated with the expression of *FMRP* across all three cell types (Pearson's correlations: 0.95 and 0.81; p = 0.004 and 0.051, respectively). Further, immunostaining indicates that CYFIP is localized to dendritic spines, concordant with animal model data. Dendritic spine density is being quantified using confocal images of pEGFP transfected and immunostained neurons **Conclusions**: Reduced expression of *CYFIP*, *NIPA2* and *TUBGC5* (but not *NIPA1*) was observed in all cell types examined. Thus, there may be compensatory expression changes in *NIPA1*. The expression of *NIPA1* and *CYFIP* is correlated with *FMRP*, suggesting regulatory mechanisms. Consistent with animal models, CYFIP may be related to dendritic spine complexity. We are currently relating the function of these four genes in relation to neuronal differentiation and morphology.

1298M

Spatio-temporal 16p11.2 Protein Network Implicates Cortical Late Mid-fetal Brain Development and RhoA Pathway in Psychiatric Diseases. *L.M. Iakoucheva[1], G.N. Lin[1], R. Corominas[1], I. Lemmens[2], X. Yang[3], J. Tavernier[2], D.E. Hill[3], M. Vidal[3], J. Sebat[4,1].* 1) Department of Psychiatry, University of California San Diego, La Jolla, CA, USA; 2) Department of Medical Protein Research, VIB, and Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University, Belgium; 3) Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, MA, USA; 4) Beyster Center for Genomics of Psychiatric Diseases, University of California San Diego, La Jolla, CA, USA.

Psychiatric disorders autism and schizophrenia have a strong genetic component, and copy number variants (CNVs) are firmly implicated. Recurrent deletions and duplications of chromosome 16p11.2 confer high risk for both diseases, but the pathways disrupted by this CNV are poorly defined. Here we investigate the dynamics of 16p11.2 network by integrating high-confident physical interactions of 16p11.2 proteins with spatio-temporal gene expression from developing human brain. We observe profound changes in protein interaction networks throughout different stages of brain development and/or in different brain regions. We identify late mid-fetal period of cortical brain development as most critical for establishing connectivity of 16p11.2 proteins with their partners. Furthermore, our results implicate the regulation of CUL3/KCTD13/RhoA pathway in layer four of inner cortical plate as crucial for controlling brain size and connectivity, and its dysregulation as potential determinant of the 16p11.2 CNV phenotypes.

1299T

Transcriptome signature of schizophrenia-associated rare copy number variants (CNVs) in lymphoblastoid cell lines (LCLs). *W. Moy[1], J. Duan[1,2], E.I. Drigalenko[3], J. Freda[1], . MGS[4], H.H. Goring[3], A.R. Sanders[1,2], P.V. Gejman[1,2].* 1) Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, IL; 2) Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL; 3) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 4) Molecular Genetics of Schizophrenia (MGS) Collaboration.

Schizophrenia (SZ) has been shown to be significantly associated with multiple rare CNVs that often encompass more >1 gene, making uncertain which CNV-genes and downstream pathways are relevant to pathogenesis. We carried out RNAseq (>9M depth, 1×50bp) transcriptome analyses to identify expression signature profiles associated with each SZ-associated CNV in LCLs from 27 subjects of European ancestry carrying SZ-associated CNVs (with at least 5 subjects for each CNV), namely 22q11.21_del (n= 13), 16p11.2_dup (n=9), and 15q13.3_del (n=5), and 657 subjects (265 SZ cases and 392 controls) without these CNVs. A total of 10,813 genes had detectable expression with a mean log2RPKM>1. We controlled analyses for sequencing batch, sex, age, affection status, Epstein-Barr viral load, growth rate, energy status and genotypic ancestral PCs. Within each CNV region, we found >90% genes showing expression associated (FDR<5%) with CNV status, consistent with expected gene dosages. Outside the CNV regions, we found variable numbers of genes with expression associated (FDR<5%) with CNVs (N): 22qdel (516), 16pdup(1), 15qdel (12), representing the downstream genes affected by the CNV. Gene set enrichment analysis (GSEA) of the 516 genes with expression associated with 22qdel showed an enrichment (FDR<5%) of gene ontology (GO) terms related to chromatin modification, translation elongation, and histone modification. Given the small number of genes associated with other CNVs, we further explored the GSEA for a larger set of genes associated with CNV status at nominal p<0.05, but with a fold difference of expression >1.2 or <0.8 (~2SD). We observed enrichment (P<0.05) of pathways and GO terms relevant to brain disorders, such as cognition, neuron development and differentiation, and regulation of neurotransmitter levels for genes with expression associated with 22qdel and 16dup. To identify the primary CNV gene/s possibly causing the downstream effects on brain-related gene set/pathway, we further examined the connectivity between CNV genes and the downstream genes of each CNV-induced brain-related gene sets by DAPPLE, and identified sialophorin (SPN) and mediator complex cubunit 15 (MED15) as possible "driver" genes at 16pdup and 22qdel, respectively. Our transcritpome analysis in LCLs of SZ-CNV carriers thus suggested plausible gene pathways relevant to SZ for each CNV region.

1300S

Functional analysis of *GRIN2A* mutations in childhood epileptic encephalopathies. *L. Addis[1, 2], L.R. Vidler[2], D. Ursu[2], D.A. Collier[2], D.K. Pal[1].* 1) Department of Clinical Neuroscience, Institute of Psychiatry, King's College London, London, U.K; 2) Neuroscience Discovery, Eli Lilly Research Centre, Windlesham, Surrey, U.K.

Epileptic encephalopathies are severe brain disorders characterized by seizures and abundant epileptiform activity which contribute to cognitive and behavioural impairments. Landau-Kleffner syndrome (LKS) and continuous spikes and waves during slow-wave sleep (CSWS) are closely related encephalopathies with regression in language and global cognitive skills respectively. They show electroclinical overlap with Rolandic epilepsy (RE), the most frequent childhood focal epilepsy, forming a clinical spectrum of epileptic, cognitive, language and behavioural disorders. Recently it was discovered that around 20% of cases in this spectrum are caused by mutations in the NMDA glutamate receptor *GRIN2A*. Here we set out to determine the disease mechanism of five missense *GRIN2A* mutations recently described in these patients. Mutations spread across the gene were chosen: P79R, C231Y, C436R, E714K and N976S, and were inserted into *GRIN2A* cDNA. HEK cells were transiently transfected with the mutant *GRIN2A*, and standard *GRIN1* constructs allowing formation of heteromers. 1) Western blotting of total protein lysates revealed that all mutations caused a decrease in GRIN2A protein levels, with levels reduced to around 65% for P79R, 40% of non-mutant levels for C231Y, E714K and N976S and to 10% for C436R. The loss of the disulphide-bond of the cysteine residue in C436R is hypothesized to drastically destabilise the protein and cause its degradation. The amino acid changes of C231Y, E714K and N976S are also predicted to be destabilising. 2) Single-cell calcium imaging was used to assay glutamate binding and NMDAR function. Mutations P79R and C231Y decreased the glutamate potency compared to non-mutated construct by increasing the half maximally effective concentration of agonist (EC50) by 3 and 5 times respectively, meaning the mutant receptors can only be activated by higher concentrations of the agonist. There was no deleterious effect from E714K, predicted to be near the glutamate binding pocket, or from the intracellular N976S mutation. C436R showed no glutamate response, which is most likely due to the very low levels of expressed protein. 3) Cell surface expression of the mutant proteins using fluorescent imaging in transfected HEK also revealed alterations. Taken together, these data suggest that mutations across *GRIN2A* affect the expression and function of the receptor in different ways, with the end result of altered NMDA receptor currents and neuronal excitability.

1301M

Transcriptional regulation at the *TREM* gene cluster in AD brains. *M.M. Carrasquillo[1], A. Allen[1], J. Burgess[1], M.L. Kachadoorian[1], S. Aryal[1], F. Zou[1], H.S. Chai[2], C.S. Younkin[1], J.E. Crook[3], V.S. Pankratz[2], A.A. Nair[2], S. Middha[2], S. Maharjan[2], T. Nguyen[1], L. Ma[1], S.J. Lincoln[1], K.G. Malphrus[1], G.D. Bisceglio[1], R.C. Petersen[4], N.R. Graff-Radford[5], D.W. Dickson[1], S.G. Younkin[1], N. Ertekin-Taner[1,5].* 1) Department of Neuroscience, Mayo Clinic, Jacksonville, FL; 2) Department of Biostatistics, Mayo Clinic, Rochester, MN; 3) Department of Biostatistics, Mayo Clinic, Jacksonville, FL; 4) Department of Neurology, Mayo Clinic, Rochester, MN.; 5) Department of Neurology, Mayo Clinic, Jacksonville, FL.

Association of missense variants in *TREM2* and *TREML2* with susceptibility of late-onset Alzheimer's disease (LOAD) has been replicated in several large case-control series including our own Mayo series (N=10,000). More recently, higher blood *TREM2* mRNA and TREM2 protein levels were reported by others to correlate strongly with AD status. In this study, we evaluate the association of brain specific transcripts encoded by the *TREM* gene cluster on chromosome 6 (*TREM1*, *TREML1*, *TREM2*, *TREML2*, *TREML4*) with variants at this locus and their effect on risk of LOAD. Utilizing Illumina DASL array data from our published brain expression genome-wide association study (eGWAS), expression levels in ADs (197 cerebellum, 202 temporal cortex) were compared to levels in non-AD (177 cerebellum, 197 temporal cortex) via multivariable linear regression, adjusting for age-at-death, sex, *APOE*-ε4, RIN, and transcript levels of 5 CNS cell-type specific genes. RNAseq data from 96 LOAD brains has been obtained and will be employed for characterization of the mechanism of transcriptional changes of *TREM* genes. For validation, *TREM* transcripts detected via RNAseq will be measured individually using isoform-specific TaqMan® gene expression assays in additional samples. The associations of gene isoforms with cisSNPs will be tested using multivariable linear regression analyses controlling for appropriate covariates. *TREML1* and *TREML4* were significantly higher in the ADs compared to the non-ADs (beta=0.085, p=0.048 and beta=0.048, p=0.048, respectively) in our eGWAS of temporal cortex tissue. Furthermore, in preliminary blood DASL array data from 7 ADs and 26 clinically, non-demented controls from the Mayo Clinic Study of Aging, *TREML1* also has higher expression in the ADs compared to controls (beta= 0.38, p=0.014). eQTL analysis from our eGWAS revealed significant association of *TREML1* temporal cortex expression levels with a putatively regulatory variant located between the *TREM2* and *TREML2* genes. This variant, which has a Regulome score = 2, is also significantly associated with AD status. Thus, *TREM* family genes appear to have differential expression in brain and blood. Completion of RNAseq and isoform specific expression analysis is expected to fine tune our understanding of these gene expression changes in this gene family. eQTL study of these transcript levels may uncover novel AD risk variants at the *TREM* gene cluster that have an effect on transcriptional regulation.

1302T

Polymorphism in the miRNA-433 binding site of FGF20 is a strong risk factor for Parkinson's disease in Iranian population. *S. Abtahi, H. Darvish, A. Movaffagh, R. Dastmalchi, B. Emamalizadeh.* Medical genetics, Shahid beheshti university of medical sciences, Tehran, Iran.

DNA variations at the fibroblast growth factor 20 gene have been reported to associate with Parkinson's disease (PD). The rs12720208, a functional SNP located in the 3' UTR region of the gene, was first reported as a risk factor for the PD. Some other studies tried to replicate the result in different populations through case-control studies. However, to our knowledge, no significant association between the rs12720208 and PD was reported. In this study, we genotyped the rs2720208 SNP in 520 PD patients and 520 healthy controls from Iran. Significant differences was found in allele and genotype frequencies between patients and controls (Fisher exact $p<1\times10-6$). The results suggest that the rs12720208 (C/T) polymorphism is a strong risk factor for PD in Iranian population.

1303S

Integration of miRNA-mRNA networks to elucidate the complexity of psychiatric disorders. *C. Chen[1], L. Cheng[2], C. Zhang[3], J.A. Badner[4], E.S. Gershon[4], C. Liu[1,5].* 1) State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, China; 2) Northwestern University, Chicago, IL, USA; 3) The University of Chicago, Chicago, IL, USA; 4) Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, Chicago, IL, USA; 5) Psychiatry department, University of Illinois at Chicago, Chicago, IL, USA.

MicroRNAs (miRNAs) are small, non-coding, endogenous RNAs involved in regulating gene expression and protein translation. One single miRNA can target multiple mRNAs and a single mRNA can also be targeted by multiple miRNAs. We consider that miRNA-mRNA clusters with statistically significant associations can explore potentially regulatory mechanism and, therefore, of biological interest. In this study, we collected 89 parietal cortex samples from Stanley Medical Research Institute(SMRI), and after quality control, each sample has 420 miRNA, 19,984 mRNA and more than 1,000,000 SNPs screened. We first constructed scale-free networks including both miRNA and mRNA, and found one module exhibited differential expression between controls and psychotic patients. In this module, mir-320e acted as one of hub nodes. Quantitative Trait Locus(QTL) result indicated mir-320e was regulated by genetic variants. Another hub gene, PDLIM5, was validated by five miRNA binding prediction software. To further investigate the causal relationship between PDLIM5 and mir-320e, we applied Network Edge Orienting (NEO) and found mir-320e regulates PDLIM5. In summary, we detected one classic regulation pathway: Genotype ->mir.320e -> PDLIM5 -> gene module -> psychotics, which can be partially explain the etiology of psychiatric disease.

1304M

Disruptions to the miRNA regulatory pathway may cause an increased rate of schizophrenia in individuals with 22q11.2DS. *W. Manley, S. Ryan, S. Siecinski, L.M. Brzustowicz.* Genetics, Rutgers University, Piscataway, NJ.

Schizophrenia is a complex and poorly understood disease caused by the interplay of many environmental and genetic factors. The 22q11.2 deletion syndrome (DS) is a disorder that is caused by the microdeletion of part of chromosome 22 leading to a 25-fold greater chance of developing schizophrenia in affected individuals versus the general population. The missing 22q11.2 region contains DGCR8, which is required for the initial step of miRNA biogenesis. However, the 22q11.2 deletion itself is not directly the cause of schizophrenia, since 75% of individuals with this deletion do not develop the disease. We hypothesize that the 22q11.2 deletion increases the risk of schizophrenia through alterations in miRNA regulatory networks via depletion of several miRNAs. This may serve as a protective buffer against the accumulation of deleterious mutations at other schizophrenia risk loci. We have derived human neural stem cells (NSCs) from individuals with the 22q11.2 deletion. The NSCs were generated from iPSCs using Neural Induction Media (Life Technologies). In order to ensure the presence of the 22q11.2 deletion, a FISH probe for 22q11.2 (TUPLE) was used (Cell Line Genetics). Also, the levels of DGCR8 gene expression were quantified using the 7900HT Fast Real-Time PCR System (Applied Biosystems) along with Taqman Gene Expression Assays to ensure DGCR8 reduction in 22q11.2 DS NSCs relative to otherwise healthy control NSCs. We have begun to characterize disruptions to the miRNA regulatory network in the NSC lines using Taqman Array Human Microarray Cards Version 3.0. Here we will present the miRNAs that we have identified to be differentially expressed in otherwise healthy control NSCs versus 22q11.2 DS NSCs. Our preliminary data shows that of the 377 tested miRNAs, there are nearly 30% that have altered expression levels greater than 2 fold or less than .5 fold. We predict that these miRNAs could be involved in the elevated risk of developing schizophrenia in individuals with 22q11.2 deletion syndrome versus the general population. Future studies will focus on determining the biological targets of these miRNAs using miRNA target prediction software.

1305T

RNA-seq analysis reveals potential link between mammalian mitochondrial fatty acid synthesis (mtFAS II), RNA processing, and neurodegeneration. *S.L. Mitchell[1], A. Parl[1], S.D. Turner[2], D.C. Crawford[1], D.G. Murdock[1].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA.

Mutations in ACSF3, the gene encoding the first enzyme of the mitochondrial fatty acid synthesis (mtFAS II) pathway, cause combined malonic and methylmalonic aciduria (CMAMMA). Patients diagnosed with CMAMMA in adulthood display psychiatric and neurological symptoms similar to those observed in Alzheimer disease. The mechanism by which defects in mtFAS II result in these phenotypes is not understood. We recently demonstrated a link between mtFAS II function and PPAR-activated gene transcription. To further investigate the potential role of mtFAS II in transcriptional regulation we knocked down the pathway in HeLa cells and performed RNA-seq. HeLa cells were transfected with either a control siRNA (n=3) or siRNA to NDUFAB1 (n=3) to knockdown expression of acyl-carrier protein (ACP1), which is necessary for mtFAS II function. Single-end read 50 base pair sequencing was completed on the Illumina Hi-Seq 2000. After sequence alignment and assembly, differential gene expression analysis was performed. A total of 3,197 transcripts were significantly differentially expressed after false discovery rate (FDR) correction for multiple testing. Of these, 364 transcripts exhibited an increase or decrease in expression greater than or equal to 1.5-fold, including 303 protein-coding transcripts, 35 non-coding RNA (ncRNA) transcripts, and 31 pseudogenes. Many of the ncRNAs were among the transcripts with the highest fold changes. Remarkably, most exhibited decreased expression and are involved in RNA processing, including RPPH1 and RMRP whose expression was reduced by approximately 90% and five members of the C/D box family of small nucleolar RNAs whose reductions in expression ranged from 83% to 90%. Significant decreases in key ncRNAs responsible for RNA processing are consistent with observations that deficiencies in mtFAS II result in mitochondrial tRNA processing defects in yeast. We describe the first evidence from mammalian cells linking the mtFAS II and RNA processing pathways. These data suggest nuclear RNA processing may also be affected by a loss of mtFAS II function. Defects in RNA processing have been identified as a possible mechanism underlying neurodegeneration in diseases including ALS, Huntington, and Prion diseases. These three diseases were also identified in pathway analysis of our gene expression data. Results from this study point to a possible link between mtFAS II dysfunction, RNA processing, and neurodegeneration.

1306S

The Genetic Factors and Molecular Mechanisms Underlying Lewy Body Pathology in Alzheimer's Disease. *O. Chiba-Falek, C. Linnertz, M.W. Lutz, J. Byrne, J.F. Ervin, C. He, J. Allen, N.R. Miller, K.A. Welsh-Bohmer, A.D. Roses.* Duke University School of Medicine, Department of Neurology, Durham, NC, USA.

The genetic basis and molecular mechanisms that lead to Lewy Body (LB) pathology in 15-20% of Alzheimer disease cases (LBV/AD) are largely unknown. Parkinson's disease (PD), the prototype of LB spectrum disorders has been studied most extensively. Genetic studies including large GWAS implicated Alpha-synuclein (SNCA) and Leucine-rich repeat kinase2 (LRRK2) in the pathogenesis of PD. Furthermore, aggregates of insoluble alpha-synuclein protein are the major component of LBs and strong evidence demonstrated the importance of SNCA over expression in PD etiology. Mitochondrial dysfunction has been implicated in PD and Bender et al have recently showed that the translocase of the outer mitochondrial membrane (Tom40) encoded by TOMM40 gene mediates mitochondrial dysfunction induced by accumulation of alpha-synuclein protein in PD. The broader role of SNCA, LRRK2 and TOMM40 genes in other LB disorders is yet to be studied, as well as the molecular genetic interplay among these genes that predispose LB pathology. Using 107 LBV/AD cases and over 400 AD control subjects we tested the associations of variants in these three loci with LB pathology in AD and identified variants, SNPs and short sequence repeat (SSR), within SNCA and TOMM40 that were significantly associated with increased risk for LB pathology. When the analyses were stratified by LRRK2 genotype, the associations with SNCA variants became stronger. Next, we investigated the expression regulation of SNCA, LRRK2 and TOMM40 in relation to LB pathology. We detected significant differences in transcript levels of each of these genes in the brain cortical tissues from autopsy-confirmed LBV/AD cases compared with AD controls; expression levels of SNCA, LRRK2 and TOMM40- mRNA were increased as a function of LB pathology. Furthermore, we tested the cis-associations of the LB-risk variants with their nearby genes and identified cis-variants that modulate expression level of these genes in human brain. Our findings demonstrated that SNCA, LRRK2 and TOMM40 contribute to LB pathology in AD patients and suggested that expression regulation of these genes may be a molecular basis underlying the observed LB associations.

1307M

REPS1 is a novel gene of Neurodegeneration with Brain Iron Accumulation. *A.B. Drecourt, N. Boddaert, I. Desguerre, D. Chretien, A. Muunich, A. Rotig.* Inserm U1163 Imagine Institute, Paris, Paris, France.

Neurodegeneration with brain iron accumulation (NBIA) encompasses a group of rare neurogenerative disorders with different clinical, brain MRI and molecular features, underlined by progressive extrapyramidal dysfunction and iron accumulation in the brain. To date, mutations in PANK2, PLA2G6, FA2H, ATP13A2, C2orf37, CP, FTL and WDR45 have been associated with NBIA. Nevertheless a large number of individuals have "idiopathic NBIA", with unknown etiology. Exome sequencing of two sisters with NBIA identified two heterozygous mutations in the REPS1 gene. REPS1 is involved in endocytosis and the two mutations (p.Ala113Glu and p.Val78Leu) affect its EH1 domain that interacts with RAB11FIP2. Western Blot analysis detected a low level of REPS1 in patient's fibroblasts. The function of REPS1 in iron metabolism is unknown, but it was shown that RABB11-FIP2 functions in transferrin recycling. We investigated the iron metabolism and oxidative stress in patient's fibroblasts. Patient fibroblasts exhibited a dramatic iron overload, measured by a colorimetric ferrozine-based assay. Consistently steady-state levels of ferritin, iron responsive protein (IRP1) and SOD2 were increased whereas aconitase activity was decreased. This indicates that REPS1 mutations induce deregulation of iron metabolism. Over-expression of the wild-type REPS1 cDNA in patient's cells reduces the iron overload of these cells that display almost normal iron content. Our experiments demonstrate that REPS1 is a new gene of NBIA. Improvement in our understanding of the biochemistry and pathophysiology of this form of NBIA will help develop novel therapeutics for this neurological condition.

1308T

Impaired Function is a Common Feature of Neuropathy- Associated GARS Mutations. *L.B. Griffin[1,2], R. Sakaguchi[3], D. McGuigan[3], M.A. Gonzalez[4], C. Searby[5], S. Züchner[4], Y.M. Hou[3], A. Antonellis[1,6,7].* 1) Cellular and Molecular Biology Program, University of Michigan Medical School, Ann Arbor, MI; 2) Medical Scientist Training Program, University of Michigan Medical School, Ann Arbor, MI; 3) Department of Biochemistry and Molecular Biochemistry, Thomas Jefferson University, Philadelphia, PA; 4) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 5) Department of Pediatrics, Division of Medical Genetics, University of Iowa, Iowa City, IA; 6) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI; 7) Department of Neurology, University of Michigan Medical School, Ann Arbor, MI.

Charcot-Marie-Tooth (CMT) disease is the most common inherited peripheral neuropathy, affecting one in 2,500 individuals worldwide. Six aminoacyl-tRNA synthetase (ARS) genes have been implicated in CMT disease with an axonal pathology. ARS genes encode essential enzymes responsible for charging tRNA with cognate amino acids. Thirteen mutations in the glycyl-tRNA synthetase (GARS) gene have been identified in patients with autosomal dominant CMT type 2D (CMT2D) characterized by impaired motor and sensory function in the upper extremities. While studies have revealed loss-of-function characteristics of GARS mutations, only a small subset of the mutations has been rigorously tested. We evaluated nine CMT-associated GARS mutations to assess if impaired function is a common feature. Our results demonstrate a strong correlation between impaired GARS function and CMT disease. Additionally, one mutation previously associated with CMT disease (S581L) did not demonstrate impaired function in our assays, failed to segregate with disease in two families with CMT, and was identified in the general population, indicating that S581L is not a disease-associated mutation and supporting the notion that impaired GARS function correlates with disease pathogenesis. Wild-type endogenous GARS forms puncta in neurons in vitro and in vivo. Seven GARS mutations impair granule formation in cultured neurons, suggesting that impaired localization may result in CMT2D disease pathogenesis. It is critical to identify proteins interacting with GARS in these structures to understand granule function and how disruption of these structures may lead to CMT disease. Impaired granule formation observed for mutant GARS proteins may be due to reduced binding to protein partners required for granule formation. Preliminary mass spectrometry data reveal that GARS associates with ribosomal complex proteins, including initiation factors and RPLP0. Previous studies demonstrated that archaeal ARS enzymes associate with ribosomes via RPLP0. This interaction may be critical in neurons for tRNA recycling for translation efficiency in the soma and axons. Current studies to identify differential protein partners of wild-type and mutant GARS will provide insight into the function of GARS-associated granules and may help explain the dominant toxicity seen in ARS-mediated CMT disease.

1309S

Novel Cytoplasmic Roles for the RNA-binding Protein, TDP-43. *R. Smith, N. Alami, J.P. Taylor.* Cell & Molecular Biology, St. Jude Children's Research Hospital, Memphis, TN.

Tar DNA Binding Protein of 43 kDa (TDP-43) is a DNA/ RNA binding protein that is found in ubiquitinated cytoplasmic inclusions in a collection of neurodegenerative diseases. While these inclusions contain ubiquitin and TDP-43, they are tau and alpha-synuclein negative, characterizing a unique "TDP-43 pathology." Furthermore, mutations in TDP-43 have been found to play a causative role in Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Lobar Degeneration (FTLD). The normal function of TDP-43 is unclear and how mutations in TDP-43 cause disease is still undetermined. TDP-43 contains two RNA recognition motifs and the protein is shuttled between the nucleus and cytoplasm. Although TDP-43 has been described to be predominately localized to the nucleus, cytoplasmic localization is seen in normal conditions within wild type tissue. This localization could highlight a critical but yet undefined role for TDP-43 in the cytoplasm. To address this hypothesis we utilized the Drosophila model to investigate several potential roles for TDP-43 in RNA processing. By over-expressing either human TDP-43 or the Drosophila ortholog, TBPH in various larval tissues, we have elucidated potential cytoplasmic functions, such as long-range transport of RNA for local translation at the synaptic terminal. This axonal transport and formation of cytoplasmic granules is dependent on the RNA-binding domain of TDP-43. Additionally, we found that ALS mutant TDP-43 granules display disrupted anterograde movement and abnormal TDP-43 distribution in the neuromuscular junction and cell body. Additionally, in muscle cells we observed cytoplasmic TDP-43 aggregates and cell toxicity that is reversed by inhibition of RNA binding. Our data identifies a novel cytoplasmic role for TDP-43 in RNA processing and transport that is disrupted by the introduction of disease-causing mutations, leading to cell death. Our results have furthered our understanding of the normal functions of TDP-43 and will advance future studies of TDP-43 associated disease.

1310M

Allelic expression analysis in the brain suggests a role for heterogeneous insults affecting epigenetic processes in autism spectrum disorders. *E. Ben-David, S. Shohat, S. Shifman.* Dept. of Genetics, The Hebrew University of Jerusalem, Jerusalem, Israel.

Monoallelic expression, including genomic imprinting, X-chromosome inactivation and random monoallelic expression of autosomal genes are epigenetic phenomena. Genes that are expressed in a monoallelic way may be more vulnerable to genetic or epigenetic mutations. Thus, comprehensive exploration of monoallelic expression in human brains may shed light on complex brain disorders. Autism-related disorders are known to be associated with imprinted genes on chromosome 15. However, it is not clear whether other imprinted regions or other types of monoallelic expression are associated with autism spectrum disorder (ASD). Here, we performed a genome-wide survey of allele expression imbalance (AEI) in the human brain using single-nucleotide polymorphisms (SNPs) arrays, in 18 individuals with ASD and 15 controls. We identified an individual with a large (~100) number of monoallelic expressed SNPs. This widespread monoallelic expression was limited to the prefrontal cortex, and was not found in the cerebellum of the same individual. Another individual was found to have a highly skewed pattern of X chromosome inactivation. This skewed inactivation was stronger in the neural cell types (neurons, oligodendrocytes and astrocytes) than in microglia. Using our data, we were also able to define the allelic expression status of known imprinted genes in the human brain. We found that many of the known imprinting genes are biallelic expressed, and other genes showed isoform-specific imprinting patterns. We were also able to discover an abnormal imprinting event in an individual with ASD, in the imprinting locus on chromosome 15q11-13. Lastly, we developed an analysis of individual-level expression, focusing on the difference of each individual from the mean. We found that individuals with ASD had more genes that were up- or down-regulated in an individual-specific manner. We also identified pathways perturbed in specific individuals. These results underline the heterogeneity in gene regulation in ASD, at the level of both allelic and total expression.

1311T

GRIP2-mediated AMPA Signaling Defects Contribute to Autism Social Behavioral Deficits. *T. Niranjan[1,2], A. Adamczyk[1], M. Han[1], R. Mejias[1], R. Rose[1], H.C. Bravo[3,4], M. Taub[4], C. Schwartz[5], D. Valle[6], R. Huganir[7], T. Wang[1,6].* 1) The Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Predoctoral Training Program in Human Genetics, Johns Hopkins University, Baltimore, MD; 3) Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD; 4) Department of Biostatistics, Johns Hopkins University School of Public Health, Baltimore, MD; 5) Greenwood Genetics Center, Greenwood, SC; 6) Department of Pediatrics, Johns Hopkins University, Baltimore, MD; 7) Department of Neuroscience, Johns Hopkins University, Baltimore, MD.

Autism spectrum disorders are clinically and genetically heterogeneous, and likely involve several hundred risk genes. Identification of disrupted brain signaling pathways that underlie Autism endophenotype is crucial for discovery of novel drug targets. Glutamate mediates the majority of excitatory neurotransmission in the CNS via its family of receptors including AMPA-, NMDA-, and mGluR-type receptors. Glutamate Receptor Interacting Proteins 1 and 2 (GRIP1/2) are neuron-enriched scaffolding proteins with seven PDZ domains. PDZ4-6 bind directly to the c-termini of AMPA receptor 2/3 (GluA2/3), Liprin-α, and EphrinB1/2 to form a postsynaptic protein complex, which plays a crucial role in AMPA receptor trafficking and synaptic function. mGluR and NMDA signaling defects are implicated in Autism-associated conditions, such as Fragile X Syndrome, Rett Syndrome, and Shank2/3 deficiencies. However AMPA signaling defects in Autism are poorly understood. We recently identified functional GRIP1 mutations that alter GluA2 synaptic trafficking and contribute to Autism social deficits. To determine if GRIP2 may mediate AMPA signaling defects in Autism, we sequenced exons of GRIP2 in a cohort of Autism patients (n=480) and ethnically matched controls (n=480), and identified a significantly increased mutation load at PDZ4-6 (p=0.022) in Autism patients. In proband families, affected sibs who carry GRIP2-PDZ4-6 mutations show more severe deficits in reciprocal social interactions defined by ADIR's social scores as compared to affected sibs who do not carry the mutations. These mutations have functional effect; two result in significantly reduced binding with GluA2/3, while four others significantly alter interactions with EphrinB1/2 and Liprin-α. *Grip2*-knockout mice exhibit reduced preference for social novelty (p<0.001) in a modified three-chamber test, and exhibit reduced social interaction (p<0.05) in a male dyadic social interaction test. Furthermore, enhancing AMPA signaling using a receptor desensitization inhibitor improves sociability in BTBR mice (p<0.001), an established Autism model. These study results support GRIP2-mediated AMPA signaling defects as a novel mechanism for social behavioral deficits in Autism. Characterization of GRIP2 deficiency in Autism shall provide valuable insights into the pathogenesis and identify novel drug targets for correction of social behavioral deficits in Autism.

1312S

A polymorphic di-nucleotide repeat (DNR) variant in the 5'UTR of *DPYSL2* gene affects its regulation via mTOR signaling. *X. Pham, Y. Liu, R. Wang, A. Pulver, D. Valle, D. Avramopoulos.* Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD.

Schizophrenia (SZ) is a common, disabling neuropsychiatric disorder with a complex etiology. It is estimated that as much as 80% can be attributed to genetic factors. Previous linkage and association studies have implicated *DPYSL2* on chr8p21 as a candidate gene for SZ. *DPYSL2* encodes CRMP2 which is important in axonal growth, and its dysfunction may result in neurodevelopmental abnormalities. We previously identified a polymorphic dinucleotide repeat (DNR) located in the 5'-untranslated region of *DPYSL2*, with a characteristic 5'-terminal oligopyrimidine (5'-TOP) tract, a target for mTOR mediated regulation pathway. The 13 CT repeat allele (risk) of the DNR was found to be associated with SZ compared to the 11 CT repeat common allele (WT). We performed dual luciferase assays in HEK293 cells and mouse primary cortical neurons and the risk allele showed ~3-fold decrease in luciferase activity as compared to the WT allele in both cell types. Further, polysome profiling of the constructs showed the fraction of luciferase mRNA in polysomes associated with the risk allele was reduced by ~3-fold compared to the WT allele. Here we show that increasing concentrations of Rapamycin, an allosteric mTOR inhibitor, reduced luciferase expression in constructs from both alleles, the risk allele remaining lower at 0-30 nM. At concentrations higher than 30 nM, both alleles reached a plateau at the same levels. The same trend was recapitulated in both HEK293 cells and mouse primary cortical neurons. Our results suggest that the difference we observe between the two DNR alleles is mediated by mTOR signaling. Using arrays of > 4,000 human transcription factors and proteins, we screened for those that bind differentially to the two alleles within the 5'-TOP of *DPYSL2* and might produce the decreased gene expression observed in the risk DNR allele. We identified a number of such proteins, one of which is a ribosomal binding protein (RBP) *HuD/ELAVL4* that is also involved in mTOR signaling. Further *HuD/ELAVL4* has been shown to play a crucial role in neuronal differentiation. How this RBP interacts with *DPYSL2* in the WT and risk form to mediate gene expression in neurons is the subject of our current studies. In conclusion, we show that *DPYSL2* is regulated by the mTOR signaling pathway and a SZ associated DNR variant in the 5'-UTR of *DPYSL2* affects this regulation.

1313M

Functional analysis in *C. elegans* of candidate genes for schizophrenia. *S.B. Pierce[1], S. Gulsuner[1], T. Walsh[1], J.M. McClellan[2], M.C. King[1,3].* 1) Department of Medicine (Medical Genetics), University of Washington, Seattle, WA; 2) Department of Psychiatry, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA.

Schizophrenia (MIM 181500) is genetically highly heterogeneous, with both *de novo* and inherited mutations playing an important role in the illness. Large numbers of rare or private mutations have been identified in individuals with schizophrenia and only a few genes have been found to harbor mutations in more than one unrelated person. Not all identified mutations will be causal and methods must be developed for assessing the biological relevance of affected genes and individual mutations. The nematode *C. elegans* provides an efficient platform for studying the neurobiological impact of putative disease genes and mutations. Advantages include: (1) a simple nervous system in which the cells are invariant in number, position, and connectivity; (2) genetic manipulability; and (3) transparency allowing for *in vivo* visualization of neurons with fluorescent reporter genes. Many genes of interest for neuropsychiatric disorders are conserved in *C. elegans* and disrupting these genes can lead to changes in neuronal morphology or alterations in animal behavior. We previously showed that persons with schizophrenia are enriched for damaging *de novo* point mutations in genes expressed in fetal prefrontal cortex. Of 54 genes identified in our study as harboring damaging *de novo* mutations, 29 (54%) are predicted to have unique *C. elegans* orthologs. We are screening this set of candidate genes by RNAi for neurological phenotypes, including defects in axon guidance, dendritic branching, or neurotransmitter-specific neuron function. To screen for axon guidance phenotypes we are using an RNAi-sensitized strain of *C. elegans* expressing a fluorescent reporter in a large number of interneurons and motor neurons, including all GABAergic neurons. We showed that knockdown by RNAi of *lpd-3*, the ortholog of our candidate gene *KIAA1109* (MIM 611565), causes axon guidance defects, including abnormal branching of commissural axons and decreased numbers of commissures reaching the dorsal cord. RNAi data is available for *C. elegans* orthologs of many, but not all, human genes; in particular not for all genes that are candidates for human neurological phenotypes. This approach will enable functional characterization of a large number of genes and suggest the best neurological phenotypes to use for further study of promising candidates. The functional consequences of select human mutations will be examined by replacing the function of the orthologous *C. elegans* gene in transgenic animals.

1314T

New mutations of *CYP2U1* in patients with spastic paraplegia and exploration of mitochondria dysfunctions. *C. Tesson[1,2], KH. El Hachimi[1,2], F. Fellman[3], G. Banneau[4], L. Raymond[1,2,4], M. Mairey[1], S. Morais[1], E. Obre[5], C. Durand[5], M. Zaki[6], J. Lavie[5], D. Lacombe[5], C. Goizet[5], A. Brice[1,4], A. Durr[1,4], F. Darios[1], G. Stevanin[1,2].* 1) Institut du Cerveau et de la Moelle épiniÃre (ICM, INSERM U1127, UPMC UMR_S1127 Sorbonne Universités, CNRS 7225, NEB), Pitié-SalpÃªtriÃre Hospital, Paris, France; 2) Ecole Pratique des Hautes Etudes, Hésam Université, Laboratoire de Neurogénétique, Paris, France; 3) Dpt de Génétique Médicale, CHUV, Lausanne, Suisse; 4) APHP, Dpt de Génétique, Paris, France; 5) Université de Bordeaux, France; 6) National Research Institute, Cairo, Egypt.

Hereditary spastic paraplegia (HSP) is considered one of the most heterogeneous group of neurological disorders, both clinically and genetically. It comprises pure and complex forms that clinically include slowly progressive lower-limb spasticity resulting from degeneration of the corticospinal tracts. At least 73 loci accounting for these diseases have been mapped to date, and mutations have been identified in 53 genes, most of which playing a role in intracellular trafficking. We recently reported that loss-of-function mutations in *CYP2U1* are responsible for an autosomal recessive form of early onset HSP. This gene codes for an enzyme involved in fatty-acid metabolism. We performed the screening of *CYP2U1* in a new cohort of 259 index patients with HSP or sporadic spastic paraplegia. We have now identified 2 new causative mutations in this gene: i) c.1469G>A/ p.C490Y, at the homozygous state in a North African sporadic case with early onset and thinning of the corpus callosum, ii) c.1A>C/ p.M1L at the homozygous state in 2 siblings with early onset HSP in a consanguineous family originating from Switzerland. Interestingly, this last mutation is also present in an asymptomatic sister aged 50 years suggesting an incomplete penetrance or a metabolic compensation in this individual. Exome sequencing on the 3 siblings identified some candidate variants involved in lipid metabolism or in the mitochondria that potentially explain this phenomenon.

In addition, previous experiments we demonstrated in human cells (fibroblasts and lymphoblast) of 2 patients with the p.D316V mutation that the pathophysiology of this clinic-genetic entity includes alteration of mitochondrial architecture and bioenergetics with increased oxidative stress. We are extending these observations to 3 additional cases using 3D electronic microscopy and preliminary results also show an enlargement of the endoplasmic reticulum (ER) in fibroblasts. Moreover, overexpression experiments in COS7 cells suggest that part of the physiopathology is explained by a partial loss of colocalisation of mutated forms with the mitochondria compared to controls. Our combined results focus attention on lipid metabolism as a critical HSP pathway with a deleterious impact on ER and mitochondrial functions.

1315S

Behavioral phenotyping of mice deficient in CHRNA7. *J. Yin[1,2], C.P. Schaaf[1,2].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, TX.

15q13.3 microdeletion syndrome is a rare genetic disorder caused by a deletion of a segment of chromosome 15. The deletion is commonly 1.5Mb in length and encompasses 6 genes, but patients with small deletions, which only encompass the CHRNA7 gene, have also been reported. The clinical phenotypes associated with this syndrome are variable, but commonly include developmental delay/intellectual disability and impaired social interaction. Other clinically important features include epilepsy, impulsive behavior, aggression, and schizophrenia. A mouse model deficient of CHRNA7 was generated and reported to be grossly normal in growth, anxiety-like behaviors, learning and memory, as well as sensorimotor gating. However, deficits in social interaction and repetitive behavior have not been assessed in this mouse model. We tested heterozygous and homozygous CHRNA7 mutant mice and their wildtype littermates for repetitive behaviors in self-grooming, holeboard exploration, and marble burying test, and for social interaction behaviors in the three-chamber paradigm, partition test, and social interaction video scoring. A detailed assessment of the aforementioned behaviors will be presented and be discussed in the context of human 15q13.3 microdeletion phenotypes.

1316M

Regulatory function of CACNA1C schizophrenia-associated variants. *N. Eckart[1], R. Wang[1], M. Zeledòn[1], M. Szymanski-Pierce[1], D. Valle[1], D. Avramopoulos[1,2].* 1) Johns Hopkins University, Institute of Genetic Medicine, Baltimore, MD; 2) Johns Hopkins University, Department of Psychiatry, Baltimore, MD.

Schizophrenia (SZ) and bipolar disorder (BP) are complex psychiatric disorders, together affecting over 3.5% of the US population. They have overlapping clinical presentations, and onset in the second or third decade of life. Association and family studies indicate a shared genetic risk. One variant that has been independently and repeatedly associated with both disorders is rs1006737, a single nucleotide polymorphism (SNP) in the third intron of the CACNA1C gene. Variants identified by association studies are often in non-coding regions of the genome and enriched in expression quantitative trait loci (eQTLs). We previously reported that the risk allele of rs1006737 is correlated with decreased expression of CACNA1C (p=0.001) in a study of 195 post-mortem tissue samples from the superior temporal gyrus (STG). In 100 post-mortem tissue samples from the dorsolateral prefrontal cortex (DLPFC) we found that the risk allele trends toward increased expression of CACNA1C, .in agreement with published data from Bigos et al. in the DLPFC (p=0.002). The variant rs1006737 tags a haplotype with several other SNPs, all located in the third intron. We are testing these for allele-specific regulatory potential using dual luciferase reporter assays. Following up on previous results, now using 4 biological replicates for each construct, we report that for two constructs, one containing rs2159100 and a second containing both rs1077306 and rs10744560, the risk alleles show statistically significant increases in luciferase expression compared to the common alleles when transiently transfected in HEK293 cells (p=4.8*10-4 and p=2.1*10-3, respectively). In addition to differences in regulatory potential, we also explore allele-specific protein binding, using electrophoretic mobility shift assays (EMSAs). We show that 4 of 7 the SNPs tested have allelic differences in protein binding profiles in HEK293 and SK-N-SH cells. In a protein microarray, we find that the risk allele for rs11062170 binds 7 proteins that do not show strong binding with the common allele. These proteins include calcium dependent proteins (ANXA11), transcription factors (MAPK8, MAX, MLX), and splicing factors (ROD1). Our data indicate that the haplotype tagged by rs1006737 plays a role in regulating transcription of CACNA1C. It appears there may be multiple functional variants, as we see allelic differences in enhancer activity as well as protein binding profiles for multiple SNPs.

1317T

Dysregulated Sonic Hedgehog signaling in MED12-related XLID disorders. *S. Srivastava[1], T. Niranjan[1], M. May[2], C. Skinner[2], R. Stevenson[2], C. Schwartz[2], T. Wang[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Greenwood Genetic Center, Greenwood, SC.

Mediator of RNA polymerase II transcription subunit 12 homolog (MED12) is a core subunit of the Mediator Complex that plays a key role in transcriptional regulation. Recurrent mutations in MED12 were found in patients with X-linked intellectual disability (XLID) syndromes including FG syndrome (R961W) and Lujan syndromes (N1007S), and sporadic mutations in patients with Ohdo syndrome. MED12-related XLID disorders exhibit a wide phenotypic spectrum including variable degrees of cognitive impairment, behavioral defects, and multiple congenital anomalies. Molecular mechanisms underlying these distinct neurodevelopmental phenotypes remain poorly understood. MED12 functions as a direct suppressor of Gli3-dependent Sonic hedgehog (SHH) signaling pathway and involves in REST (RE1silencing transcription factor)-imposed epigenetic restriction of neuronal gene expression. By screening a large cohort of patients with XLID of unknown etiology, we identified five new mutations that segregate with the disease phenotype in the probands families and are absent in more than 500 normal males. All five are missense mutations involving highly conserved amino acid residues. Four (N854S, N898D, R1214C, and R1295H) are located in the Leu-Ser (LS) rich-domain where the recurrent mutations for FG and Lujan syndromes were found and one (R206Q) in the Leu (L)-rich domain. To understand the molecular mechanisms of these mutations in XLID phenotype, we performed real-time RT-PCR quantification studies on the transcript levels of three genes, *CREB5*, *BMP4*, and *NGN2*, in the SHH signaling pathway in lymphoblasts from probands patients. Transcript levels for all three genes show significant increase in lymphoblasts carrying the four MED12 mutations within the LS domain but minimal or no change for the mutation in L domain. These results are consistent with previous findings that MED12-LS domain plays an important role in the negative regulation of the Gli3-dependent SHH signaling. Genotype-phenotype correlation and REST-dependent neuronal gene expression are currently being investigated. Results of these studies expand the clinical spectrum and support that dysregulated Gli3-dependent SHH signaling contributes to the phenotypic features of MED12-related XLID disorders.

1318S

Transcriptome analysis of Lphn3 null mutant mouse brain and implications for ADHD and Addiction. *D. Wallis[1], S. Galaviz[1], B. Baker[1], M. Tucker[1], T. Ioerger[2].* 1) Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX; 2) Department of Computer Science and Engineering, Texas A&M University College Station, TX 77843-3474.

Latrophilin 3 (LPHN3), a GPCR, has been identified by linkage and association studies performed by separate research groups in separate cohorts as a prime candidate gene for ADHD and Addiction. Lack of readily identifiable functional polymorphisms suggests that non-coding variations determining the quantity and/or quality of LPHN3 isoforms are likely contributors to these behaviors, but the mechanisms by which LPHN3 might influence susceptibility and gene expression profiles in neurons are essentially unknown. We generated the Lphn3 null mutant mouse in order to evaluate gene function. We have already reported behavioral characterization of these mice indicating that they are both hyperactive and display increased reward seeking behavior. Here we describe the results after comparative transcriptome analysis of 3 brain regions over time between male wild type and null mutant Lphn3 littermates. Brains were harvested and dissected at 3 time points to reflect different stages of brain development (4 days, 4 weeks, and 6 months). Three brain regions known to play roles in ADHD: the hippocampus (important for learning and memory), the prefrontal cortex (the brain area essential for most executive functions), and the striatum (the location of dopaminergic reward-processing circuitry key in ADHD) were also independently sampled. We performed 100 bp single end reads on a HiSeq2500 and got over 13 million reads per sample. Preliminary analysis has focused on changes in transcription between time points and regions based on genotype. Some of the more surprising results indicate that Serpina3n is one of the most downregulated genes in all 3 brain regions in null mice at 6 and 1 month, but is relatively unchanged at 4 days. While we fully anticipate differential gene expression over time, we are startled by the lack of correlation of genes identified as showing changes between nulls and mutants at Day 4 in comparison to the other time points. This suggests that analysis of even earlier time points might be beneficial and supports the concept that ADHD is a developmental disorder. Further data analysis will emphasize neurotransmitters and their receptors, transporters, and metabolism genes as well as neurodevelopmental and survival genes. Enrichments for GO pathways will be determined.

1319M

Dual-marker lineage specific sorting in heterogeneous Parkinson's disease patient-specific iPSC-derived dopaminergic neuronal cultures. *K. Belle[1, 2, 3], B.A. DeRosa[1, 2, 3], J.M. Van Baaren[1,3], J.M. Vance[1,3], D.M. Dykxhoorn[1,3].* 1) John P. Hussman Institute for Human Genomics, Miami, FL 33136, United States; 2) Dr. John T. MacDonald Foundation Department of Human Genetics, Miami, FL 33136, United States; 3) University of Miami Miller School of Medicine, Miami, FL 33136, United States.

Parkinson's disease (PD) is a neurodegenerative disease characterized by the loss of dopaminergic neurons (DAns). Model systems and cell lines have helped improve our understanding of PD neuropathologies. In addition to failing to faithfully recapitulate the full spectrum of PD phenotypes, animal models lack the genetic architecture of human disease. PD studies in human patient fibroblast and immortalized cell lines are hampered by the fact that they are not the relevant tissue type affected in PD. To better understand PD genetics, we have used reprogramming and derivation techniques to derive stem cell lines genetically identical to patients diagnosed with PD, which, in turn, have been differentiated to relevant neuronal cell types, midbrain dopaminergic neurons. Advancements in the field of iPSC research make it a powerful tool for modeling genetic diseases by functional and transcriptomic studies. This is especially important for the study of progressive neurological diseases such as PD. While studies of autopsy brains provide endpoint data, longitudinal in vitro studies to measure disease-related dysfunction and gene expression changes can be performed with iPSC derived cultures. One challenge associated with gene expression studies from iPSC culture is the heterogeneity inherent to the differentiation to specific cell types. This is due to heterogeneity in both the cellular composition and developmental maturity of the differentiating cultures. To overcome this limitation, we have designed and tested lentiviral-based fluorescent reporter constructs that allow for the identification and isolation of specific cell types and developmental stages. Utilizing promoter reporter constructs for the tyrosine hydroxylase (TH) and dopamine beta hydroxylase genes, both markers for DAns, we were able to enrich for DAns mature enough to produce dopamine. We have created a dual-marker system by incorporating a previously designed promoter reporter for Synapsin1 (SYN1) to further enrich our sample for mature SYN+/TH+ DAns. These enriched cultures show increase expression of mature DAn markers and related genes. This dual-marker system will be further utilized in gene expression studies of PD patient derived neuronal cultures to measure subtle gene expression differences between the DAns of patients and controls.

1320T

Investigating the role of RBFOX1 in human stem cell-derived glutamatergic neurons. *H.N. Cukier[1], B.A. DeRosa[1,2], K.C. Belle[1,2], Y.S. Park[1,2], A.J. Griswold[1], J.M. Lee[1], M.L. Cuccaro[1,2], J.R. Gilbert[1,2], D.M. Dykxhoorn[1,2], M.A. Pericak-Vance[1,2].* 1) John P. Hussman Institute of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL.

Autism spectrum disorders (ASDs) are a group of devastating neurodevelopmental conditions that currently afflict about 1 in every 68 children. Since most genetic causes are quite rare, investigations of ASD loci using genome wide association studies (GWAS) have identified relatively few consistent signals of interest. Our group performed a joint analysis of three independent, family-based GWAS datasets for a total of 2,963 families. The top hit with a near genome-wide significance (rs3095508, joint $p=6.04 \times 10^{-8}$) was in the *RNA binding protein, fox-1 homolog 1* (*RBFOX1*) gene. *RBFOX1* had a total of 9 SNPs with $p<1\times10^{-5}$. *RBFOX1* is a neural splicing cofactor that regulates alternative splicing. Both our group and others have identified copy number variants and point mutations in *RBFOX1* in ASD individuals. In addition to ASD, *RBFOX1* has been connected to bipolar disorder, epilepsy, and schizophrenia. To better understand the impact of *RBFOX1* on neuronal functionality, we are evaluating how *RBFOX1* expression modulates neuronal function in iPSC-derived glutamatergic neuronal cells. The viability and functionality of the glutamatergic neurons in which *RBFOX1* has either been overexpressed or silenced by RNA interference are assessed at multiple time points during in vitro neurogenesis. Overexpression experiments are performed with lentiviral transduction and genomic integration the wild type form of *RBFOX1* variant 4 driven under a CMV promoter, as well as an ASD-specific nonsense mutation at R173X that we identified in a sporadic ASD family. Each condition is compared at a gross morphological level to determine if modulating *RBFOX1* produces recognizable qualitative or quantitative phenotypes. Since *RBFOX1* plays a key role in splicing and transcription, RNA-seq analysis of the iPSC-derived neurons under three conditions (*RBFOX1* knockdown, overexpression and no treatment) will be used to complement the functional analysis to identify key networks regulated by *RBFOX1*. Collectively, it appears that *RBFOX1* may play a broader role in ASD pathogenesis and other neurological disorders than has been previously suspected.

1321S

Mutant dystrophin Dp71$\Delta_{78\text{-}79}$ stimulates cellular proliferation in the inducible system PC12 Tet-On. *A. Herrera-Salazar[1], J. Aragón[1], J. Romo-Yáñez[1], A. Sánchez-Trujillo[1], V. Ceja[1], R. García-Villegas[2], C. Montañez[1].* 1) Genética y Biología Molecular, CINVESTAV-IPN, México, Distrito Federal, Mexico; 2) Fisiología, Biofísica y Neurociencias, CINVESTAV-IPN, México, Distrito Federal, México.

Duchenne muscular dystrophy (DMD), a X-linked disease, is a degenerative muscle disorder caused by mutations in the DMD gene resulting in the absence of cytoskeletal dystrophin or the presence of truncated protein. DMD gene has a complex transcriptional regulation that give rise to several tissue-specific proteins named according their molecular weight (Dp427, Dp260, Dp116, Dp71, Dp40). Dp71 is the major DMD gene product expressed during embryonic development and in adult brain. In addition, patients with severe mental retardation have mutations that affect Dp71 expression. The differential cellular distribution of Dp71 isoforms, spliced out at exons 71, 71-74 and 79, suggest different functions for these proteins. Reports using PC12 cells, a neuronal model, have shown alterations in the nerve grow factor (NGF)-induced neuronal differentiation and cell cycle when Dp71 expression is depleted. Nevertheless, the mechanism of Dp71 dystrophin function has not been describe yet. As an attempt to define the functional role of Dp71 domains, our team have been studied a novel mutant protein, Dp71$\Delta_{78\text{-}79}$, which lacks exons 78 and 79. This deletion does not affect the interaction sites with β-dystroglycan, syntrophin and dystrobrevin. In this work we generated an inducible system by stable transfection of PC12 Tet-On cells with pTRE2pur-Myc/Dp71$\Delta_{78\text{-}79}$ construct. Doxycycline was used to switch on the transcription of mutant dystrophin and the effects were analyzed by Western blot and optical and confocal microscopy. Results shown that PC12 cells transfected with Dp71$\Delta_{78\text{-}79}$ have an increase in the cell growth rate when NGF is added in the medium rather than stimulate differentiation, when compared to control cells. Western blotting analysis show an increase in the expression of β-dystroglycan (dystrophin-associated protein related with cell adhesion, signal transduction and regulation of actin cytoskeleton) and a decline in Dp71a (endogenous dystrophin) and Grb2 (mediator of Ras-related signaling pathway) when comparing cells expressing Dp71$\Delta_{78\text{-}79}$ to control cell. Confocal microscopy exhibits β-dystroglycan variation in subcellular localization in Dp71Δ78-79 clone. Proliferating cell nuclear antigen (PCNA) and focal adhesion kinase (FAK) expression is also increased in this construct. Our results suggest that the mutant dystrophin isoform Dp71$\Delta_{78\text{-}79}$ has a role in cell cycle in this system.

1322M

Prenatal malnutrition reprogrammed postnatal gene expression in mammals' brain. *JW. Xu[1,2,3], XY. Zhou[1,2], YQ. Xiang[1,2], T. Wang[1,2], L. He[1,2], XZH. Zhao[1,2].* 1) Reproductive Medicine Centre,The First Affiliated Hospital of Zhengzhou Universtiy, Zhengzhou, China; 2) Children's Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai, China; 3) Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai, China.

Background: Converging evidence suggests that schizophrenia is a neurodevelopmental disease with environmental influences during early brain development. Prenatal nutritional deficiencies have long been implicated in the etiology of schizophrenia. The birth cohorts affected by the Dutch Hunger Winter of 1944-1945 and 1959-1961 Chinese famine who suffered a nutrition deficiency showed a two fold increase in the incidence of schizophrenia in adulthood. **Material and Methods:** We established a prenatal "famine" rat model, names RLP50 that induced by maternal exposure to a diet restricted to a 50% low (6%) protein diet to explore increased schizophrenia risk in those born during the Chinese famine. NimbleGen gene expression microarray was used to investigate the differential expression genes, MBD-seq was used to detect whole genome DNA methylation, MASSARRAY EpiTYPER was used to validate the MBD-seq data. **Results:** We show prenatal malnutrition reprograms adult prefrontal cortex gene expression in the offspring of RLP50, 415 genes showing modest significant expression differences ($P<0.05$), of which 48 genes showed highly significance ($P<0.01$), related to neurotransmitters and olfactory function. In the hippocampus, we identified 2987 genes with significant ($p<0.05$) expression differences, and 841 genes with highly significant ($p<0.01$) expression differences, the differentially expressed genes were related to synaptic function and transcription regulation. DNA methylome profiling of the hippocampus revealed a systematic epigenetic changes of which majority showed hypermethylation (86.9%). Remarkably, plasma membrane was significantly enriched in both profiling screens ($P = 2.37\times10^{-9}$ and 5.36×10^{-9}, respectively). Furthermore, Slc2al associated with cognitive impairment, showed significant down-regulation and hypermethylation in the hippocampus of the RLP50 offspring. The gene expression and DNA methylation differential affected the same GO function term-plasma membrane ($P=3.93\times10^{-9}$) and cell projection ($P=9.78\times10^{-7}$).**Conclusion:** We show that the gene expression reprogramming was less significant in the PFC of RLP50 offspring compared with that of the hippocampus, as well as the genome-wide DNA methylation. Our results suggest that prenatal exposure to famine will reprogram postnatal brain gene expression, disturbed synaptic and neurodevelopment, impairing learning and memory, and therefore contribute to predisposition to schizophrenia.

1323T

Gene expression profiling of human astrocytes treated with bexarotene and related compounds shows an increase in the neuroprotective cytokine GMCSF. *R.F. Richholt[1,2,3], I.S. Piras[4], A.M. Persico[4], M.J. Huentelman[1,2,3].* 1) Translational Genomics Research Institute, Phoenix, AZ; 2) Arizona Alzheimer's Consortium, Phoenix, AZ; 3) Evelyn F McKnight Brain Institute at the University of Arizona; 4) Lab. of Mol. Psychiatry and Neurogenetics, Univ. Campus Bio-Medico, Rome, Italy.

Characteristic neuropathology of Alzheimer's disease (AD) includes the accumulation of extracellular amyloid plaques in the brain. These plaques are thought to be formed by an imbalance between beta-amyloid (Aβ) production and clearance. Recent studies in multiple AD mouse models show that treatment with the RXR agonist bexarotene (BEX) restores cognitive functions and in some models results in reduced soluble and oligomeric Aβ. These observations position BEX as a potential agent for AD prevention therapy. RXR and LXR activation has been shown to increase expression of the cholesterol transporters ABCA1 and ABCG1, as well as APOE. These increases were attributed to the benefits of the BEX treatment on Aβ, but they also caused concern regarding its potential use in patients carrying the epsilon 4 allele variant of APOE. How these molecules facilitate AB clearance is not fully understood; therefore we utilized gene expression profiling to investigate BEX and related RXR/LXR agonists in human cells. Human primary astrocytes (Lonza) were treated for 48 hours with 100nM concentrations of the following RXR/LXR agonists - BEX, honokiol, and 9-cis retinoic acid (RA). Gene expression analysis was conducted with Illumina HumanHT12 v4 BeadChips and differential expression analysis was performed with the R package Limma. Hierarchical clustering and gene ontology analysis was also conducted. BEX and RA significantly upregulated ABCA1 and ABCG1 ($p<0.01$, validated by qRT-PCR), but APOE was unaffected by any of the three drug treatments. Cluster analysis identified a group of immune response genes that were upregulated at three hours by all drugs. Among these genes, BEX increased granulocyte-macrophage colony stimulating factor (GMCSF) (Log2 fold change 1.64, $p<0.01$), a cytokine that is known to be neuroprotective. Additionally, treatment of cultured human microglia with BEX demonstrated a significant increase in GMCSF across a similar time course. This study is the first to examine the molecular effects of BEX in human cells. Our results suggest that BEX treatment does not upregulate APOE expression and therefore should remain a strong candidate for anti-amyloid therapy in humans. Additionally, our results demonstrate that BEX may act at least in part via upregulation of GMCSF. Several studies show that upregulation of GMCSF can reverse both cognitive impairment and amyloidosis. BEX likely represents a novel approach to upregulate GMCSF in the central nervous system.

1324S

Persistent neurocognitive decline is associated with vascular and epithelial damage to the choroid plexus and β-amyloid plaques in an outbred rat model. *A.J. Wyrobek[1], S. Bhatnagar[1], B. Rabin[2].* 1) Lawrence Berkeley National Laboratory, Berkeley, CA; 2) University of Maryland, Baltimore Campus, Baltimore, MD.

CNS damage in youth may accelerate neurocognitive decline later in life and hasten the onset of neurological diseases. Understanding the molecular mechanisms of CNS damage after radiation exposures is of special importance for patients receiving cranial radiotherapy, astronauts returning from extended space missions, and has relevance to those experiencing traumatic head injuries. We employed an outbred rat model (Sprague Dawley) to investigate the time-course of molecular and cellular CNS damage after HZE irradiation with 56Fe (1 Gev/n; 10 or 100 cGy), using age-matched shams and young animals as reference. Irradiated rats showed persistent neurocognitive deficits on novel object recognition and bar-press assays. CNS transcriptomic findings pointed to persistent molecular changes in the choroid plexus (CP), the structure that produces cerebral spinal fluid (CSF). Groups of rats were exposed at 2 or 6 months (m) of age and CNS tissue was sampled at 4, 9 or 21 m later. Beginning at 4 m after exposure, there was increased CP endothelial fibrosis especially in the small fenestrated capillary vessels, and the CP epithelium produced less transthyretin (Ttr) protein, a major component of CSF. Ttr is a β-amyloid-binding protein that facilitates β-amyloid clearance. There was also a progressive increase in β-amyloid plaques beginning at 9 m after exposure. Unirradiated animals did not show age-related changes in CP fibrosis or Ttr expression, but showed a small increase in plaque frequency with age. Our findings are consistent with the hypothesis that CNS neurotoxic exposures or trauma early in life may cause persistent CP dysfunctions that lead to persistent neurocognitive decline associated with the diminished production of molecular factors required for β-amyloid clearance and prevention of β-amyloid plaque build-up in advanced aging. [supported by NASA NNX14AC86G (AJW) and NNJ06HD93G (BR) at LBNL under DE-AC02-05CH11231.].

1325M

No association between telomere length and exposure to life course stress or adversity in two longitudinal New Zealand cohorts. *S. Jodczyk[1], J.F Pearson[2], D.M Fergusson[3], L.J Horwood[3], J.K Spittlehouse[3], P.R Joyce[2], M.B Hampton[1], M.A Kennedy[1].* 1) Department of Pathology, University of Otago, Christchurch, New Zealand; 2) Department of the Dean, University of Otago, Christchurch, New Zealand; 3) Department of Psychological Medicine, University of Otago, Christchurch, New Zealand.

Telomeres are specialised structures that maintain the integrity of the genome. They shorten with each cell division, eventually reducing the telomere to a critical length and triggering cell senescence and apoptosis. Previous research suggests that various life stressors increase the rate of telomere shortening with potential impact on disease states and mortality later in life. The objectives of this study were to establish and validate a qPCR method for telomere length measurement and apply this assay to two longitudinal cohorts. The first was the Christchurch Health and Development Study (CHDS) (n=677), a birth cohort that has followed participants for over 30 years and the second was the Christchurch Health, Aging and Lifestyle Cohort (CHALICE) (n=351), a population sample of 50 year olds. We hypothesised that telomere length would be proportional to an individual's cumulative exposure to life course stress and adversity. We experienced difficulty establishing a reliable multiplexed method (MMQPCR) for telomere length measurement on the Lightcycler 480 platform. This was due to complexities in extraction and analysis of data and differential heating across the Peltier block. In addition, we discovered that choice of hot-start DNA polymerase was critical for the success of the assay. Six of the 12 hot-start enzymes tested yielded primer dimers with the telomere-specific primers, presumably due to activity of the enzyme below the specified activation temperature. Reliable data could only be obtained by implementing a modified form of the original single amplicon qPCR method, which was validated with Southern blotting. No associations were found between telomere length and life course stress or adversity in either cohort. The stressors examined in the CHDS cohort spanned the antenatal through to young adulthood periods and included childhood physical or sexual abuse, substance use and mental health problems. Correlations were very small ranging from -0.06 to 0.06, and none were statistically significant. The stressors assessed in the CHALICE cohort included BMI, cholesterol levels and smoking status. We conclude that up to the age of 50, life course stressors have negligible effects on telomere length. Although telomere length does not appear to provide a sensitive measure of the effects of environmental stress, it may be that telomere erosion measures obtained by comparing telomere length before and after exposure to stress would be more effective.

1326T

Polymorphisms in the TCF4 gene interact with body mass index to influence lithium response among patients with bipolar disorder. *E. Ryu[1], A. Cuellar-Barboza[2], M. Prieto[2], J. Geske[1], C. Colby[1], J. Biernacka[1,2], M. Frye[2].* 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Psychiatry and Psychology, Mayo Clinic Rochester, Rochester, MN, USA.

Background: Although lithium is considered as the first line of treatment for bipolar disorder (BD), half of BD patients do not respond to it. Recent studies have indirectly suggested a potential connection between lithium's protective cellular effects and TCF4 on chromosome 18. TCF4 codes for transcription factor 4, shown to affect multiple signaling pathways involving cell survival and neuronal differentiation, together with neuropsychiatric phenotypes including schizophrenia. Therefore, we assessed the association between genetic variants in TCF4 with lithium response to treatment of BD, allowing the genetic effect to be modified by body mass index (BMI), which has also been associated with lithium response.

Methods: Data from 341 BD patients previously treated with lithium was used to assess the association between lithium response (no response, intermediate response, or response) and 20 TCF4 SNPs previously reported to be associated with schizophrenia. Adjusting for age and sex, cumulative logit models including BMI-SNP interactions were applied to jointly test the main effect of each SNP and its interaction with BMI.

Results: Among 341 patients, 52% responded to lithium treatment, 32% showed intermediate response, and 16% had no response. The joint test showed nominally significant evidence of association between two intronic SNPs in TCF4 (rs17512836 and rs17597926) with lithium response (P= 0.009). These two SNPs were in almost complete in linkage disequilibrium (r2=0.99) and showed an interaction effect with BMI (P=0.003), indicating that the effect of these SNPs on lithium response depends on BMI.

Conclusion: TCF4 is a member of the basic helix-loop-helix (bHLH) family of transcription factors that have an important role in a number of developmental processes. Our results suggest that the SNPs in this gene may interact with BMI for lithium response. Future work is needed to replicate this finding and to investigate the biological mechanism behind this association.

1327S

Multiple functional linear models and three dimensional functional principal component analysis for image-genetic data analysis in clouds. *J. Jiang[1], N. Lin[1], L. Luo[3], J. Yu[1], V. Calhoun[2,3], C. Gupta[2], M. Xiong[1].* 1) School of Public Health, UTHSC, Houston, TX; 2) The Mind Research Network, Albuquerque, NM; 3) University of New Mexico.

Large-scale integrated genetic and imaging data analysis is a new approach used to uncover the individual variability and mechanism of disease development. This approach, which has not been well developed, has the potential to open a new avenue for dissecting genetic structure of complex disease and personal medicine. Joint analysis of imaging and genetic data will identify genes significantly associated with disease and provide useful information on target therapy. The widely used statistical methods for image-genetic analysis often employ some simple summarizing statistics to represent image signals. As a consequence, these methods often formalize image-genetic studies as a traditional genetic study of a single quantitative trait in which we test association of a single SNP with a single summarizing statistic of the image signals. However, summarizing statistics completely ignore distribution of signals of the image in two or three dimensions. In addition, single SNP analysis methods are difficult to deal with next-generation (NGS) data. To overcome these limitations, we extend one dimensional principle component analysis to three dimensional principle component analyses (3DFPCA) and use 3DFPCA scores to represent image data. These 3DPCA score are then taken as multiple quantitative traits. To deal with NGS data, we use one gene or one genomic region as an unit of genetic analysis. In other words, we develop multiple functional linear models for image-genetic analysis. The methods for parameter estimation and statistics for testing association of gene or genomic region are also developed. Since both imaging- and genetic-domain observations include a huge number of variables, joint image and genetic analysis on such Big Data represents a computational challenge that cannot be addressed with conventional computational techniques. To address this challenge, we develop sophisticated algorithms for joint analysis of imaging and genetic data running on Amazon cloud computer. The proposed method was applied to the mind clinical imaging consortium's schizophrenia image-genetic study with 9,766 genes (746,575 SNPs) typed in 94 schizophrenia patients and 75 healthy controls. The excellent QQ plot of our results showed that we did not inflate the false positive rates. We identified 44 significant ly associated genes, 20 of which can be confirmed to be associated with Schizophrenia from the literature.

1328M

Whole genome analysis of high-dimensional phenotypic data: Multiple testing in the context of genome-wide analysis. *S.E Medland[1], B.M Neale[2,3,4], N. Jahanshad[5,6], D.P Hibar[5,6], A. Arias Vasquez[7,8], J.L Stein[9], N.G Martin[10], B. Franke[7,8], M.J Wright[11], P.M Thompson[5,6,12], the ENIGMA consortium.* 1) Quantitative Genetics, QIMR Berghofer, Brisbane, QLD, Australia; 2) Program in Medical and Population Genetics, The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 4) Psychiatric and Neurodevelopmental Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 5) Imaging Genetics Center, Institute for Neuroimaging & Informatics, Keck School of Medicine of University of Southern California, Los Angeles, California, USA; 6) Department of Neurology, University of Southern California, Los Angeles, California, USA; 7) Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands; 8) Departments of Psychiatry, Human Genetics & Cognitive Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands; 9) Neurogenetics Program, Department of Neurology, UCLA School of Medicine, Los Angeles, USA; 10) Genetic Epidemiology, QIMR Berghofer, Brisbane, QLD, Australia; 11) NeuroImaging Genetics, QIMR Berghofer, Brisbane, QLD, Australia; 12) Departments of Neurology, Psychiatry, Engineering, Radiology, Pediatrics and Ophthalmology, University of Southern California, Los Angeles, California, USA.

Genome-wide association (GWAS) has been a highly successful technique for the identification of replicable loci influencing a wide variety of traits. Arguably, much of the consistency and robustness of effects identified using GWAS methods derives from the development and application of genome-wide approaches to correct for multiple testing which have reduced the number of false positives within the literature. As GWAS are almost exclusively univariate in nature, extension of the multiple testing corrections to account for running GWAS on multiple traits has focused on Bonferroni corrections for the number of independent traits being analysed. However, with the increasing focus on BIG data problems, where genome-wide analyses are repeatedly applied to dense sets of correlated phenotypes, a more structured approach would be useful. One such big data problem commonly encountered in imaging genetics is the analysis of voxel-level brain imaging data, where over 2 million data points with complex correlational structures are typically extracted. Here we present a series of analyses to determine the genome-wide significance threshold required to maintain a family-wise error rate of < 5% for voxel-level image analysis. Following Dudbridge and Gusnanto, we estimated the dimensionality of the voxel-level volumetric analysis in three independent cohorts (ADNI, BIG and QTIM), for whom structural MRI and genome-wide data were available. To examine the robustness of these results we examined the impact of field strength (from 1.5 to 4 tesla magnets), voxel size and image processing. Across data sets with over 2 million data points each we estimated the dimensionality of the voxel-level analysis to be approximately 10,000. Adopting a Bonferroni correction for testing 10,000 independent traits, the genome-wide, voxel-wide tensor based morphology significance threshold would then be $p< (5 \times 10^{-8})/10{,}000 = 5 \times 10^{-12}$. Importantly, the threshold remained stable in down-sampled data sets (an approach that is often advocated as a method of reducing the multiple testing burden). Encouragingly, the largest imaging genetics studies of brain volume to date, performed as part of the ENIGMA consortium, show effects on the order of $P = 10^{-23}$, indicating that this threshold is within the reach of available sample sizes. These findings also have implications for other high-dimensional phenotype sets such as gene expression data.

1329T

MicroRNAs associated with declarative memory phenotypes. *J. Neary[1], H. Kulkarni[1], J. Peralta[1], D. Cruz[2], D. Reese McKay[3], E. Knowles[3], P. Fox[4], J. Curran[1], H. Göring[1], L. Almasy[1], R. Olvera[5], P. Kochunov[6], R. Duggirala[1], J. Blangero[1], D. Glahn[3], M. Carless[1].* 1) Dept Genetics, Texas Biomedical Research Inst, San Antonio, TX; 2) Dept. Psychiatry, University of Texas Health Sciences Center San Antonio, San Antonio, TX; 3) School of Medicine, Yale University, Hartford, CT; 4) Research Imaging Institute, University of Texas Health Sciences Center San Antonio, San Antonio, TX; 5) School of Medicine, University of Texas Health Sciences Center San Antonio, San Antonio, TX; 6) School of Medicine, University of Maryland, Baltimore, MD.

Declarative memory refers to long-term memory that can be consciously recalled, such as factual information and events. Declarative memory impairment, a well-established endophenotype for several psychiatric illnesses, can be indexed either behaviorally, or through in vivo measures of brain structure. To understand genetic contributions to declarative memory, we sequenced microRNAs (miRNAs) from peripheral blood cells of 1,152 Mexican Americans within ~50 large pedigrees who have undergone magnetic resonance imaging (MRI) and cognitive assessment; we then assessed gray matter thickness, surface area, and volume for nine neuroanatomical structures relevant to declarative memory and five memory-related neurocognitive traits. 1,583 mature miRNAs (1,539 known, 42 SNP-altered in whole genome sequence data, 2 novel) were tested for association with 14 declarative memory-related phenotypes. In total, 2,390 significant ($p<0.05$) associations were detected, as well as 3 significant associations ($p<3.16\times10^{-5}$) after correction for multiple testing. Some miRNAs previously implicated in memory-related disorders (miR-107, miR-148a-3p, miR-25-3p, miR-35c-5p) were associated with multiple phenotypes, as were several other miRNAs, including miR-1972 (6 phenotypes, $p=2.74\times10^{-2}$-5.34×10^{-6}), miR-5586-5p (11 phenotypes, $p=4.21\times10^{-2}$-2.56×10^{-5}), miR-7706 (5 phenotypes, $p=2.61\times10^{-2}$-$<1.00\times10^{-7}$), and miR-922 (3 phenotypes, $p=1.93\times10^{-2}$-3.51×10^{-5}). We also identified significant associations between miRNA sequences altered by SNPs and several phenotypes. Additionally, both novel miRNAs identified by sequencing were also associated ($p<0.05$) with several phenotypes, and one was highly expressed in ~15% of our cohort, which may suggest that this miRNA is cohort-specific. Our data suggests that miRNAs may act as peripheral biomarkers for declarative memory, and further investigation of the correlation between blood and brain miRNAs may help to implicate these in neurodevelopment, neurodegeneration and synaptic plasticity. Our early analysis in baboon tissue suggests that only ~21% of miRNAs show significantly different expression levels between blood and brain tissue. The identification of SNP-altered miRNAs and novel miRNAs, which appear to be at least nominally associated with declarative memory phenotypes, further suggests that (cohort-specific) genetic variation may be important in regulatory function for miRNA biogenesis.

1330S

Expanding the phenotypic spectrum of AFG3L2-associated ataxia. *A. Knight Johnson, S. Halbach, K. Silver, D. Waggoner.* University of Chicago, Chicago, IL.

Pathogenic variants in AFG3L2 have been associated with spinocerebellar ataxia, type 28 (SCA28), which is an autosomal dominant disorder associated with slowly progressive gait and limb ataxia, lower limb hyperreflexia, dysarthria, nystagmus, ptosis and ophthalmoparesis. The mean age of onset of symptoms is 27 years, with a reported range of 3-60 years. The majority of AFG3L2 pathogenic variants described to date have been missense changes in exons 15 and 16 within the M41-protease domain of the gene. We present a case with a de novo pathogenic variant in AFG3L2 in a 7 year old male with a history of upper body tremors starting at 6 months of ages, significant rapidly progressive ataxia, and cerebellar atrophy. Between 6-8 months the patient was noted to have episodic upper body tremors that occurred multiple times per day, truncal hypotonia and delayed motor and speech development. At age 12 months bilateral esotropia was noted. A brain MRI obtained at 12 months was normal, a repeat MRI at age 2 years showed cerebellar atrophy. By age 6 years the patient had significant balance issues requiring support to walk or stand, and dysarthria. The family history was not significant for ataxia, and the patient had no siblings. Whole exome sequencing was performed, and a de novo variant in the AFG3L2 gene, c.1994G>T, was identified. This variant results in a missense change, p.Gly665Val, in a highly conserved amino acid residue in exon 16. Three different pathogenic missense changes involving the adjacent amino acid residue, p.Met666, have previously been described in patients with SCA28. This patient exhibited an earlier age of onset than any previously described case of SCA28, and has more rapidly progressive disease than is typical for the disorder. The clinical variability that has previously been described includes rare cases of individuals who present at age 6-8 and only two cases of presentation at age 3 in identical twin siblings. The twins had slower progression than our current case. Our patient expands the known phenotypic spectrum of SCA28 suggesting a possible modifier gene affect. A previous case of homozygous mutations in AFG3L2 in siblings has been described where the mutations affected the binding of AFG3L2 to paraplegin and resulting in a severe phenotype that included features of spasticity and neuropathy. This raises the possibility of other minor changes in the SPG7 gene (codes for paraplegin) to modify the phenotype of SCA28.

1331M

Expanding the phenotypic spectrum of *TUBB4A*-associated hypomyelinating leukoencephalopathies. *S. Miyatake[1], H. Osaka[2,3], M. Shiina[4], M. Sasaki[5], J. Takanashi[6], K. Haginoya[7], T. Wada[8], M. Morimoto[9], N. Ando[10], Y. Ikuta[11], M. Nakashima[1], Y. Tsurusaki[1], N. Miyake[1], N. Matsumoto[1], H. Saitsu[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Division of Neurology, Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Department of Pediatrics, Jichi Medical School, Tochigi, Japan; 4) Department of Biochemistry, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 5) Department of Child Neurology, National Center of Neurology and Psychiatry, Tokyo, Japan; 6) Department of Pediatrics, Kameda Medical Center, Chiba, Japan; 7) Department of Pediatric Neurology, Takuto Rehabilitation Center for Children, Sendai, Japan; 8) Genetic Counselling and Clinical Research Unit, Kyoto University School of Public Health, Kyoto, Japan; 9) Department of Pediatrics, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan; 10) Department of Pediatrics, Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 11) Department of Neurology, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan.

Leukoencephalopathies are a heterogeneous group of disorders affecting the white matter of the brain. Hypomyelination with atrophy of the basal ganglia and cerebellum (H-ABC) is one of those characterized by early-onset motor regression and/or delay followed by extrapyramidal symptoms. A recurrent *de novo TUBB4A* mutation was recently reported in 11 patients with H-ABC. Interestingly *TUBB4A* mutations also cause DYT4 dystonia, suggesting that *TUBB4A* mutations may have a broader phenotypic spectrum. Here we performed whole-exome sequencing analysis of patients with genetically unsolved hypomyelinating leukoencephalopathies, identifying eight patients with *TUBB4A* mutations, and allowing the phenotypic spectrum of *TUBB4A* mutations to be investigated. Fourteen patients with hypomyelinating leukoencephalopathies, 7 clinically diagnosed with H-ABC, and 7 with unclassified hypomyelinating leukoencephalopathy, were included. Six heterozygous missense mutations in *TUBB4A*, 5 of which are novel, were identified in 8 patients [6/7 H-ABC patients (the remaining patient is an atypical case), and 2/7 unclassified hypomyelinating leukoencephalopathy patients]. In 4 cases with parental samples available, the mutations occurred *de novo*. The effect of the mutations on microtubule assembly was examined by mapping altered amino acids onto 3D models of the αβ tubulin heterodimer. The p.Glu410Lys mutation, identified in patients with unclassified hypomyelinating leukoencephalopathy, directly impairs motor protein and/or microtubule-associated protein interactions with microtubules, while the other mutations affect longitudinal interactions for maintaining αβ tubulin structure, suggesting different mechanisms in tubulin function impairment. In patients with the p.Glu410Lys mutation, the clinical course appeared generally milder compared to the typical H-ABC patients. Their basal ganglia atrophy was unobserved or minimal although extrapyramidal features were detected, suggesting its functional impairment. We conclude that *TUBB4A* mutations cause typical H-ABC. Furthermore *TUBB4A* mutations associate cases of unclassified hypomyelinating leukoencephalopathies with morphologically retained but functionally impaired basal ganglia, suggesting *TUBB4A*-related hypomyelinating leukoencephalopathies encompass a broader clinical spectrum than previously expected. Extrapyramidal findings may be a key for consideration of *TUBB4A* mutations in hypomyelinating leukoencephalopathies.

1332T

Genetic enrichment of multiple sclerosis risk loci in multiple sclerosis patients with co-morbid diseases. *M.F. Davis[1], J.L. Haines[2], J.C. Denny[3].* 1) Microbiology and Molecular Biology, Brigham Young University, Provo, UT; 2) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH; 3) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.

Patients with MS are at a higher risk of developing a number of other diseases, specifically other autoimmune diseases. There are many shared genetic risk factors between autoimmune diseases, including multiple sclerosis (MS), and patients with MS are at higher risk of the development of other diseases. Studies of specific co-morbidities of MS have been published, but there are few studies that have the clinical data necessary to conduct a broad study of the clinical landscape of MS. Electronic medical records (EMR) have extensive clinical data for patients and provide an opportunity to collect data on all diseases a patient experiences, rather than a narrow interview conducted for a disease-specific research study. We previously created and validated algorithms to identify patients with MS and detailed clinical traits of MS disease course from EMRs. Using a de-identified EMR dataset of 5,789 MS patients, we are now creating algorithms based on ICD-9 billing codes, text key words, and medications to identify and characterize diseases that occur co-morbidly with MS. We are comparing the distributions of co-morbid diseases identified to those in published literature. 1,000 of these patients are genotyped on the ImmunoChip. We will perform analysis to determine if patients with particular co-morbidities are enriched for specific MS risk loci.

1333S

Microbiome profiling in whole blood using RNA-seq reveals disease-specific patterns. *S. Mangul[1], A.P. Ori[2], K.A. Staats[2], L.M. Olde Loohuis[2], G. Jospin[3], J.A. Eisen[3], E. Eskin[1], R.A. Ophoff[2].* 1) Department of Computer Science, University of California, Los Angeles, Los Angeles, California, 90095; 2) Center for Neurobehavioral Genetics, University of California, Los Angeles, Los Angeles, California, 90095; 3) Department of Evolution and Ecology, College of Biological Sciences, University of California, Davis, 95616.

Microbial communities in and on the human body represent a complex mixture of bacteria, viruses, archaea and microbial eukaryotes containing in total more than ten times the number of genes found in the human genome. Advances in high-throughput sequencing offer a powerful culture-independent approach to study the underlying diversity of microbial communities of human tissues in health and disease. Availability of comprehensive compendiums of reference microbial genomes and rRNA genes provides ample possibilities to use sequencing data to profile microbial communities present across different human tissues. Reads mapped to microbial reference databases can be used to assign taxonomy to microbial species and identify genes expressed across different samples and subjects. Here, we analyzed RNA-seq datasets from 192 humans from four disease groups (i.e. schizophrenia, bipolar disorder, amyotrophic lateral sclerosis and unaffected controls). Non-mapped RNA-seq reads (failed to map to the human genome) are mapped to a compendium of reference marker genes and microbial genomes and subsequently used to study underlying diversity of microbes in blood. We perform analysis of microbial communities using Phylosift, a phylogeny-driven method that uses Bayesian methods to asign taxonomic IDs and to compare community diversity between samples. More specifically, Phylosift uses phylogenetic comparison of reads to reference databases of rRNA and protein coding "marker" genes to profile individual microbial communities and then carries out edge principal components analysis to compare communities. Additionally we validate individual microbial communities using genomic sequence data of the same individual. Using identical procedures we are able to use non-mapped reads to profile diversity of microbial communities. We are able to confirm the microbial genomes at different abundance levels. Using sequencing data from 192 individuals we were able to study microbiome diversity in blood in subjects with and without brain disorders. In particular, we observe distinctive patterns of microbiota composition in schizophrenia patients. Further analysis is needed to decipher the origin and meaning of these differences. The widespread detection of microbial species in whole blood and the observed differences in disease groups underscores the involvement of the microbiome in human health and disease.

1334M

COFS syndrome due to ERCC1 mutation without Nucleotide Excision Repair defect. *Y. CAPRI[1], N. CALMELS[2], M. GERARD[3], V. BELAVOINE[4], I. DORBOZ[5], D. RODRIGUEZ[5,6,7], V. LAUGEL[2,8], O. BOESPFLUG-TANGUY[4,5], L. BURGLEN[5,7,9].* 1) Medical Genetics, Robert Debré University Hospital, Paris 19eme, Paris, France; 2) Department of Molecular Genetics, University Hospital Strasbourg, 67091 Strasbourg; 3) Department of Genetics, University Hospital Caen, 14000 Caen; 4) Department of Pediatric Neurology and metabolic diseases, APHP- Robert Debré University Hospital, 75019 Paris; 5) INSERM U1141, Paris; 6) Department of Pediatric neurology, APHP -Trousseau University Hospital, 75019 Paris, France; 7) Centre de référence des malformations et maladies congénitales du cervelet, APHP-Trousseau University Hospital, 75012 Paris; 8) Department of Pediatric, University Hospital Strasbourg, 67098 Strasbourg; 9) Department of Genetics, APHP- Trousseau University Hospital, 75012 Paris.

Cerebro-oculo-facio-skeletal syndrome (COFS syndrome) is a heterogeneous disease characterized by the association of microcephaly, congenital cataract or microphtalmia, arthrogryposis, severe psychomotor delay, growth retardation and facial dysmorphism. COFS syndrome is due to Nucleotide Excision Repair (NER) defect. ERCC6, ERCC5, ERCC2 and ERCC1 mutations have been involved in COFS syndrome. ERCC1 mutations have been reported in 1 patient with COFS syndrome and in 2 other patients; one patient developed a type II Cockayne syndrome and the other had a severe form of Xeroderma pigmentosum with neurodegeneration symptoms (both syndromes belong to the same spectrum of DNA repair disease as COFS). The patient is the 1st girl of consanguineous Algerian parents. Pregnancy was uneventful but at birth, she displayed distal arthrogryposis and rapidly developed microcephaly, seizures and severe encephalopathy. Brain MRI showed brainstem and cerebellar hypoplasia, and myelinisation delay. She could not walk, sit nor speak. She died at 5 years from chest infection. The 2nd pregnancy of the couple was interrupted because of the recurrence of the disease with arthrogryposis observed at 28 weeks of gestation (WG) and cerebellar hypoplasia at 32WG, confirmed by foetal brain MRI. The pregnancy was terminated at 34WG. We report the 4th familial case of ERCC1 mutations. In this family, homozygous p.Phe231Leu mutation was associated with COFS syndrome but no NER defect was observed. This family confirmed ERCC1 involvement in COFS syndrome but NER defect seemed inconsistent compared to what can be observed when ERCC6 and ERCC5 is mutated. The p.Phe231Leu mutation was already reported in 2 out of the 3 published patients with ERCC1 mutations (one at the homozygous state, and the other in combination with a nonsense mutation). The pathogenicity had already been confirmed by complementation assay and this mutation was not found in 6,000 control exomes. The absence or low consequence of ERCC1 mutations on NER function in homozygous individuals could suggest a predominance of transcriptional anomalies in COFS/Cockayne syndrome. For clinical relevant patients, functional study is probably not sufficient to decide if ERCC1 is involved. But among the 3 reported patients with the p.Phe213Leu ERCC1 mutation, clinical findings are highly variable and more patients are needed to precise the phenotype and eventually the genotype-phenotype correlation of ERCC1 mutations.

1335T

MindCrowd: web-based testing of 19,202 individuals suggests family history of Alzheimer's disease is associated with decreased episodic memory performance in young adults. *M.J. Huentelman[1,2], J.J. Corneveaux[1,2], I. Schrauwen[1,2], A.L. Siniard[1,2], J. Peden[1], E. Reiman[4,2], R. Caselli[5,2], E. Glisky[2,3], L. Ryan[2,3].* 1) Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ; 2) Arizona Alzheimer's Consortium, Phoenix, AZ; 3) University of Arizona, Tucson, AZ; 4) Banner Alzheimer's Institute, Phoenix, AZ; 5) Mayo Clinic, Scottsdale, AZ.

Episodic memory is influenced by both heritable and non-heritable components; however, most studies are largely underpowered to detect all but those that exert large effects on the cognitive task of interest. To address these sampling and cohort diversity concerns, we created a web-based (at mindcrowd.org) paired associate episodic memory task (PAL) in an attempt to interrogate a large cohort of individuals that span a morezersus 12 years or less of education for 14/16/20 years of education respectively), marital status (p=0.0002, r=1.3% higher performance in single individuals), handedness (p=0.0006, r=1.7% higher scores in right handed subjects), and the presence of a first-degree relative diagnosed with Alzheimer's disease (AD, p=0.002, r=1.3% higher performance in individuals without a first degree relative with AD). This AD first-degree relative effect is enhanced when individuals under the age of 42 are analyzed separately from those over 42 (p=4.3e-05). Prior work, on smaller cohorts, has demonstrated the influence of the AD risk-associated APOE locus has a minimal effect on cognitive performance in unaffected relatives suggesting that these findings are most likely explained by an as of yet undefined heritable factor(s). After removing the three easiest word pairs in the PAL task, we also noted a significant association above age 70 in individuals with first degree relatives with AD. These findings suggest that heritable factors associated with first degree AD risk exert an effect on cognitive performance even in healthy individuals at young adult ages. Our results demonstrate the effectiveness of web-based recruitment for the study of cognition across a diverse cohort and further highlight the potential utility in identifying genes associated with cognition and their role in AD.

1336S

Phenotypic spectrum associated with PTCHD1 deletions and truncating mutations. *J.B Vincent[1], A. Chaudhry[2], A. Noor[3], B. Degagne[1], K. Baker[4], K.F. Boerkoel[5], L.A. Bok[6], A.F. Brady[7], D. Chitayat[8], B.C. Hon-Yin[9], C. Cytrynbaum[2], D. Dyment[10], I. Filges[11], B. Helm[12], D.J. Stavropoulos[13], L.J.B. Jeng[14], F. Laumonnier[15], C.R. Marshall[16], S. Parkash[17], F.L. Raymond[18], A.L. Rideout[17], W. Roberts[19], R. Rupps[5], I. Schanze[20], C.T.R.M. Schrander-Stumpel[21], S.J.C. Stevens[6], E.R.A. Thomas[22], A. Toutain[16], S.W. Scherer[23], M.J. Carter[2].* 1) MiND LAB, CAMH, Toronto, ON, Canada; 2) Department of Pediatrics, Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, University of Toronto, Toronto ON Canada; 3) Department of Pathology and Laboratory Medicine, The Hospital for Sick Children, Toronto ON Canada; 4) Department of Medical Genetics, University of Cambridge, Cambridge, United Kingdom; 5) Department of Medical Genetics, Children's and Women's Health Centre, University of British Columbia, Vancouver BC Canada; 6) Department of Clinical Genetics, Unit of Cytogenetics, Maastricht University Medical Center, Maastricht, the Netherlands; 7) North West Thames Regional Genetics Service, Northwick Park Hospital, Harrow, United Kingdom; 8) The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; 9) Department of Pediatrics and Adolescent Medicine, Department of Obstetrics and Gynaecology, Centre for Reproduction, Development and Growth, Centre for Genomic Sciences, The University of Hong Kong, Hong Kong; 10) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa ON Canada; 11) Division of Medical Genetics, Department of Biomedicine, University Hospital Basel, Basel, Switzerland; 12) Division of Medical Genetics and Metabolism, Children's Hospital of The King's Daughters/Eastern Virginia Medical School, Norfolk, VA, USA; 13) DPLM, The Hospital for Sick Children, University of Toronto, Toronto ON Canada; 14) Department of Laboratory Medicine, University of California, San Francisco, USA; 15) UMR_INSERM U930 Faculté de Médecine, Université François Rabelais, Tours, France; 16) The Center for Applied Genomics, The Hospital for Sick Children, Toronto ON Canada; 17) Maritime Medical Genetics Service, IWK Health Centre, Halifax, NS Canada; 18) Institute for Medical Research Wellcome Trust, University of Cambridge, Cambridge, United Kingdom; 19) Autism Research Unit, The Hospital for Sick Children, Toronto ON Canada; 20) Institute of Human Genetics, University Hospital Magedeburg, Germany; 21) Department of Clinical Genetics and School for Oncology & Developmental Biology (GROW), Maastricht UMC+, Maastricht, the Netherlands; 22) Clinical Genetics Department, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom; 23) McLaughlin Centre for Molecular Medicine, University of Toronto, Toronto ON Canada.

Studies of submicroscopic genomic copy number variants (CNVs) have identified candidate genes associated with autism spectrum disorders (ASD) such as PTCHD1, NRXN1, SHANK2 and SHANK3. Four families have been reported in literature with exonic deletions of PTCHD1 in individuals with ASD and/or intellectual disability (ID), with minimal description of phenotypic features. Here, we describe the first and largest cohort of individuals with PTCHD1 deletions or truncating mutations. We report phenotype information for 21 males and one female from 15 unrelated families. Eighteen of these individuals have deletions involving PTCHD1 gene while four patients have truncating mutations within the PTCHD1 coding region. Our data suggests that individuals with disruption of PTCHD1may have subtle dysmorphic features including a long hypotonic face, prominent forehead, puffy eyelids and thin upper lip. They do not have a recognizable pattern of associated congenital anomalies or growth abnormalities. They have mild to moderate global developmental delays, and many have prominent behavioural issues. 45.5% of subjects (10/22) have a diagnosis of ASD or have ASD-like behaviours. The only consistent neurological finding in our cohort is orofacial hypotonia and mild motor incoordination. While detailed neuropsychological studies of individuals with PTCHD1 disruptions are needed to better define the cognitive and behavioural phenotype, our findings suggest that hemizygous PTCHD1 loss of function causes an X-linked neurodevelopmental disorder with a strong propensity to autistic behaviours. Because there is limited information so far on the functionality of PTCHD1, we have not included missense mutations in the current analysis, however it would be of interest to see whether patients with missense changes predicted to be damaging and not present in controls individuals also share some of the subtle features reported here. Likewise, patients with deletions of the region upstream of PTCHD1 were not included, as it is currently not known whether these CNVs exert an effect through PTCHD1 or through another mechanism. Again, it will be interesting to evaluate the clinical features in these patients and to compare to the PTCHD1 deletion and truncation patients reported here.

1337M

Report of a Colombian family with new clinical features for autosomal dominant sleepwalking and night terrorsautosomal dominant. *M. Lattig, L. Hernandez.* Biological Sci, Univ de los Andes, Bogota, Colombia.

Sleepwalking and night terrors are sleep disorders classified as excitement Parasomnias. These disorders generate a considerable decree in the individual's quality of life. There is only one study that previously reported an autosomal dominant inheritance pattern in an American family demonstrating linkage with a region between the SNP type markers rs728331 and rs286819 (Licis, et al. 2011). We report second case in a three-generation Colombian family with 18 affected individuals, with a dominant inheritance pattern and reduced penetrance. However, there are some mayor differences between the reported family and ours. Onset prevalence and sex-affected ratio are the mayor phenotypic distinctions. We propose that we may have a new condition that is caused by different genes.

1338T

The transcriptional regulator ADNP links the nBAF (mSWI/SNF) complexes with autism. *F. Kooy[1], G. Vandeweyer[1], C. Helsmoortel[1], A. Van Dijck[1], A.T. Vulto-van Silfhout[2], B.P. Coe[3,4], R. Bernier[5], J. Gerdts[5], L. Rooms[1,6], J. van den Ende[1,6], M. Bakshi[7], M. Wilson[8], A. Nordgren[9], L.G. Hendon[10], O.A. Abdulrahman[10], C. Romano[11], B.B.A. de Vries[2,12], T. Kleefstra[2,12], E.E. Eichler[3,4], N. Van der Aa[1,6].* 1) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; 2) Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Institute for Genetic and Metabolic Disease, Radboud University Medical Center, Nijmegen, The Netherlands; 3) Department of Genome Sciences, University of Washington School of Medicine, Seattle, USA; 4) 4Howard Hughes Medical Institute, University of Washington, Seattle, USA; 5) 5Department of Psychiatry, University of Washington, Seattle, USA; 6) Department of Medical Genetics, University Hospital Antwerp, Belgium; 7) Department of Genetic Medicine, Westmead Hospital, Sydney, Australia; 8) Department of Clinical Genetics, Children's Hospital at Westmead, Westmead, Australia; 9) Clinical Genetics Unit, Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 10) University of Mississippi Medical Center, Jackson, MS, USA; 11) Unit of Pediatrics and Medical Genetics, I.R.C.C.S. Associazione Oasi Maria Santissima, Troina, Italy; 12) Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands.

Mutations in ADNP were recently identified as a frequent cause of syndromic autism, characterized by deficits in social communication and interaction and restricted, repetitive behavioral patterns. Based on its functional domains, ADNP is presumed transcription factor. The gene interacts closely with the SWI/SNF complex by direct and experimentally verified binding of its C terminus to three of its core components. A detailed and systematic clinical comparison of the symptoms observed in our patients allows a detailed comparison with the symptoms observed in other SWI/SNF disorders. While the mutational mechanism of the first 10 patients identified suggested a gain of function mechanism, an 11th patient reported here is predicted haploinsufficient. The latter observation may raise hope for therapy, as addition of NAP, a neuroprotective octapeptide, has been reported by others to ameliorate some of the cognitive abnormalities observed in a knockout mouse model. It is concluded that detailed clinical and molecular studies on larger cohorts of patients are necessary to establish a better insight in the genotype phenotype correlation and in the mutational mechanism.

1339S

Prenatal and perinatal risk factors for autism spectrum disorders. *A. Anhalt[1], L.R. Simard[1], X-Q. Liu[1, 2].* 1) Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Manitoba, Canada; 2) Department of Obstetrics, Gynaecology, and Reproductive Sciences, University of Manitoba, Winnipeg, Manitoba, Canada.

Background: Autism spectrum disorders (ASD) are complex neurodevelopmental disorders. Twin studies have shown that ASD is affected by genetic factors; though, discordancy (e.g. only one twin is ASD-affected) was observed in at least 20% of identical twins. Environmental and epigenetic factors may contribute to the discordance in identical twins. Of these, various pre-/perinatal factors (e.g. abnormal/breech presentation) have been associated with ASD. However, previous studies usually do not account for age at first diagnosis or covariate effects, such as birth year or maternal age at birth. **Objectives:** The objectives of this study are to investigate the relationships between: 1) ASD and twinning, 2) pre-/perinatal conditions and twinning, and 3) pre-/perinatal conditions and ASD, while incorporating age at first diagnosis and covariates. **Methods:** We identified families with twins and/or ASD-affected individuals using administrative data from the Population Health Research Data Repository in the province of Manitoba (HIPC#2013/2014-41). We estimated the prevalence of ASD in 5 to 9 year old twins during 5-year periods. We compared the presence of pre-/perinatal complications between twins and singletons. Using survival analysis models, we examined the role of prenatal factors in ASD. **Results:** After applying quality control measures, we identified ~4000 families with at least one twin pair and ~4000 families with at least one ASD-affected individual; >100 of these families included twins and ASD-affected individuals. We estimated the prevalence of ASD in twins to be 0.56% in 1996-2001 and 1.15% in 2001-2006; our prevalence estimates in twins are higher than published estimates for all 5 to 9 year old children during these periods in the province. In our preliminary analyses, twins were more likely than singletons to experience prenatal (OR=1.62; 95% CI=1.32-1.99) or perinatal complications (OR=2.32; 95% CI=2.12-2.54). The overall presence of prenatal conditions was not associated with ASD. Further analyses will examine specific pre-/perinatal conditions and incorporate covariate effects. **Significance:** Identifying potential risk factors for ASD may aid in developing effective methods for the prevention, diagnosis, and treatment of these disorders. Future directions include comparing DNA methylation and gene expression profiles between discordant identical twins and incorporating any significant pre-/perinatal factors in the epigenetic analyses.

1340M

Evidence of a Genetic Basis for Developmental Topographical Disorientation. *S.F. Barclay[1], F. Burles[2], K. Rancourt[1], N.T. Bech-Hansen[1], G. Iaria[2].* 1) Department of Medical Genetics and Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada; 2) Department of Psychology, Hotchkiss Brain Institute & Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada.

BACKGROUND: The ability to find our way around familiar and foreign environments, known as topographical orientation, is a complex cognitive task involving many brain regions. A variety of brain lesions can result in an inability to navigate, known as acquired topographical disorientation. Distinct from these cases is Developmental Topographical Disorientation (DTD), in which topographical disorientation is present in a subject with no apparent brain structural abnormalities and apparently intact general cognitive skills (Neuropsychologia 47: 30-40, 2009). The defining characteristics of DTD are that individuals (a) have not suffered any brain injury, (b) are not affected by neurological conditions, and (c) have apparently well-preserved cognitive functions; and yet they (d) get lost daily in very familiar surroundings and have done so since childhood. In a study of 120 individuals with DTD (Exp Brain Res. 206: 189-96, 2010), one third of the subjects reported one or more family members with similar orientation difficulties, suggesting that DTD may be a genetic disorder. OBJECTIVE: To collect phenotypic data from families to determine whether there may be a genetic basis for DTD. METHODS: We have developed an online test that includes a diagnostic questionnaire (the Santa Barbara Sense of Direction Scale) and a virtual reality cognitive map test, and have used this online protocol to test and diagnose relatives of participants who report a family history of DTD. RESULTS: We have recruited five families to our study in which at least two members have a DTD phenotype confirmed by our online testing protocol. In total, among these five families, we have tested 30 individuals and confirmed DTD in 13. Each of these families supports an autosomal dominant mode of inheritance, in that every affected individual has an affected parent, and men and women appear to be affected in approximately equal proportions. CONCLUSION: Our evidence suggests that some cases of DTD are genetic, suggesting that there is a heritable basis to the complex cognitive function of navigation.

1341T

Genetic Basis of Dynamic Auditory Processing with Application to Reading Ability. *J.F. Flax[1], M.J. Bruni[2], S.L. Wolock[2], C.W. Bartlett[2], L.M. Brzustowicz[1].* 1) Department of Genetics, Rutgers University, Piscataway, NJ; 2) Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Childrens, OH.

Acoustic characteristics of human language follow distinct patterns of amplitude and frequency changes. Insensitivity to amplitude modulation (AM) and frequency modulation (FM) of sounds in terms of discrimination of sound and speed of processing has been associated with dyslexia. Given that developmental dyslexia occurs in higher prevalence among children with specific language impairment (SLI) and many children on the autism spectrum also have language impairments with concomitant dyslexia, we proposed that this phenotype may be useful for finding susceptibility loci to these disorders. In this study, the genetic and environmental correlations between AM and FM tasks and reading tasks was examined in SLI and ASD families prior to quantitative trait linkage analysis. Three extended pedigrees ascertained for specific language impairment (SLI) with 117 subjects and 52 families ascertained for both SLI and autism spectrum disorder with 248 subjects were analyzed. The AM and FM perceptual thresholds were determined for 2 Hz, 20 Hz and 240 Hz. Reading was assessed by single word and single non-word reading (Woodcock Battery). Heritabilities and genetic correlations were modeled with SOLAR. Linkage analysis was conducted with the PPL framework. Univariate heritabilities for both AM and FM were found to be significant ($p<.05$). AM and two reading measures showed an environmental correlation ($p=1\times10^{-6}$), but did not have a significant genetic correlation. The opposite was true for FM which showed a genetic, but not environment correlation with the reading measures ($p=0.02$ and $p=0.013$). Linkage analysis established linkage to chromosomes 20 (FM 2Hz PPL=65%). Chromosome 10 showed linkage to both AM 2Hz and FM 2Hz (PPLs of 66% and 33%, respectively) suggesting it is not specific to either AM or FM but may relate to processing slow changes in auditory input. A strong environmental correlation of AM and reading ability indicates that the similarity between these two measures is driven mostly by a common set of environmental experiences, not shared genetics. FM and reading ability appear to have genetic drivers that could be elucidated by future molecular genetics studies examining association of genes in the linked regions. FM perceptual thresholds and the acoustic and perceptual phenotypes they may represent could be useful for identifying additional reading genetic loci.

1342S

Linkage analysis of IQ discrepancy in autism: an attempt to replicate. *A.Q. Nato[1], N.H. Chapman[1], H.K. Sohi[1], R.A. Bernier[2], J.M. Viskochil[3], H. Coon[3], E.M. Wijsman[1,4,5].* 1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA; 2) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA; 3) Department of Psychiatry, University of Utah, Salt Lake City, UT, USA; 4) Department of Biostatistics, University of Washington, Seattle, WA, USA; 5) Department of Genome Sciences, University of Washington, Seattle, WA, USA.

In individuals with autism spectrum disorders, performance IQ (PIQ) is often greater than verbal IQ (VIQ). Linkage analysis of IQ discrepancy (PIQ-VIQ) in University of Washington (UW) nuclear families identified several genomic regions with evidence of linkage, including the strongest evidence on chromosomes (chr) 10p12 and 17 [Hum Genet 2011:129,59-70]. These signals were not apparent when analyses were rerun without IQ data on parents, and attempts to replicate these signals in datasets without parental IQ data (AGRE and AGP) were unsuccessful. Here, we attempt to replicate these two linkage signals in large multi-generational families from Utah (UT), where IQ data are available on multiple generations.

A total of 1,532 individuals were available in 67 UT families, ranging from 4 to 177 people. 631 subjects are typed for 6,044 SNPs, and 509 of these subjects have IQ data. In the UW dataset, Wechsler PIQ was measured by block design and object assembly subscales, while VIQ was measured by vocabulary and comprehension subscales. A diversity of IQ measures were present in the UT families, so we chose single subtests to represent PIQ and VIQ, with an effort to stay as close as possible to original UW measures. For PIQ, we used block design from Wechsler tests and pattern construction from DAS tests (BD). For VIQ, we used vocabulary from Wechsler tests and naming vocabulary or word definition from DAS tests (VOC). Analysis of this revised definition of IQ discrepancy (BD-VOC) in the UW families showed the regions of interest on chr 10 and 17.

We estimated multipoint IBD sharing in the UT families with Markov chain Monte Carlo (MCMC) methods due to the large size of the pedigrees. We used PBAP to select a subpanel of markers to maximize information while using informative markers in linkage equilibrium that were spaced sufficiently for the MCMC methods. We performed variance components linkage analyses in SOLAR, using models allowing for additive effects with and without dominance in both the major gene and polygenic components. Using SNPs selected from the 6k panel, we found a moderate lod score (~0.75) on chr 10p12, in addition to a larger lod score 20 cM away. The chr 17 signal did not replicate. Analyses with more genotyped individuals, using SNPs from a 750k chip, are underway.

1343M

Identifying endophenotypes associated with Age-related Macular Degeneration in the Amish. *M. Pericak-Vance[1], R.J. Sardell[1], J.N. Cooke Bailey[2], W. Cade[1], L.D. Adams[1], R. Laux[2], D. Fuzzell[2], L. Reinhart-Mercer[1], L. Caywood[1], D. Dana[3], A.S. Bowman[3], M.G. Nittala[4], S. Sadda[4], J.L. Haines[2], D. Stambolian[3].* 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 3) Departments of Ophthalmology and Genetics, University of Pennsylvania, Philadelphia, PA; 4) Doheny Eye Institute, Los Angeles, CA.

Age-related macular degeneration (AMD), a leading cause of blindness in older adults, is a complex disease with several known genetic and environmental risk factors. However, a substantial portion of heritability in disease risk remains unexplained. In addition to reflecting unidentified rare genetic variants, this missing heritability may also be an artifact of challenges in defining the complex phenotype. Ocular Coherence Tomography (OCT) imaging quantifies fine-scale features of AMD, potentially allowing the identification of endophenotypes associated with the disease. We ascertained both genetic samples and detailed phenotype data via OCT from both affected and unaffected related individuals from Amish populations across Pennsylvania, Ohio and Indiana. The Amish provide an excellent opportunity to analyze the heritability of complex traits given their large nuclear families, while their relatively uniform genetic and environmental background increases power to detect rare genetic variants. We sampled 51 individuals (22 males, 29 females) from 19 Amish families with mean age of 73 (range 49-99). Approximately 8% of individuals had no AMD, 51% had early AMD, 25% intermediate AMD, 9% geographic atrophy and 6% advanced neovascular AMD. In preliminary analyses, we assessed the extent to which the AMD phenotype may be independent of genotype by measuring the correlation of features between left and right eyes. We defined the phenotype using both traditional presence/absence of drusen and quantitative OCT parameters (drusen area and volume). Across all 51 individuals, presence of drusen was strongly correlated between left and right eyes; 94% of individuals were concordant between eyes for presence of small drusen, 86% for medium drusen, and 90% for large drusen. Drusen area in both a 3mm circle (Spearman's r_s=0.48) and a 5mm circle (r_s=0.72), and drusen volume in a 3mm (r_s=0.34) and a 5mm circle (r_s=0.72) were also correlated between eyes. These results suggested individual repeatability of quantitative phenotypic traits. OCT parameters were also correlated with the traditional Age-Related Eye Disease Study (AREDS) scale AMD grade (r_s=0.42-0.57); correlation coefficients did however suggested additional variation in the relationship, confirming that OCT parameters may help to define the disease process. Further analyses on the heritability of quantitative AMD traits may therefore enable the identification of endophenotypes associated with this disease.

1344T

***P54NRB/NONO* mutations link intellectual disability to impaired gene expression and altered circadian rhythm.** *M. LANGOUET[1], M. RIO[1,2], S. MOUTTON[1], K. SIQUIER-PERNET[1], C. BOLE-FEYSOT[3], N. CAGNARD[4], P. NITSCHKE[4], A. MUNNICH[1,2], D. MIRCSOF[5], P. SEEBECK[6], L. GASPAR[5], S. BROWN[5], J. AMIEL[1,2], L. COLLEAUX[1].* 1) Imagine Institute, Necker-enfants malades hospital, Paris, France; 2) Genetic Service, Necker-enfants malades hospital, AP-HP, Paris, France; 3) Genomic platform, Imagine Institute, Necker-enfants malades hospital, Paris, France; 4) Bioinformatic platform, Imagine Instiute, Necker-enfants malades hospital, Paris, France; 5) Chronobiology and sleep research group, Institute of pharmacology and toxicology, University of Zurich; 6) Center of integrative rodent physiology, University of Zurich.

We report on two unrelated patients presenting a novel syndromic XLID featuring slender build-macrocephaly, scoliosis, severe elocution disability with mandibulofacial dysostosis and, a thick corpus callosum at brain MRI. High-throughput sequencing identified two distinct null mutations in the *P54NRB/NONO* X-linked gene: c.1131G>A; p.Ala377Ala and c.1394dup; p.Asn466Lysfs*13. P54NRB/NONO belongs to the DBHS (Drosophila Behaviour Human Splicing) protein family with three members in mammals i.e. P54NRB/NONO, PSPC1, and PSF/SFPQ. DBHS proteins are nuclear proteins implicated in multiple aspects of RNA production and processing, as well as RNA surveillance via binding and retaining hyper-edited RNA in subnuclear bodies named paraspeckles. In addition, they serve as chromatin-regulating transcriptional cofactors within the mammalian circadian clock. Finally, PSF and P54NRB have been implicated in dendritic RNA transport. We demonstrated that both mutations lead to complete absence of the P54NRB/NONO protein and over-expression of the two other DBHS proteins in patients' cells. Whole transcriptome comparison of patients' fibroblasts vs controls reported global deregulation of gene expression. Moreover, clock gene expression analysis using reporter assays revealed reduced circadian clock amplitude in these cells. Finally, CT scans of P54nrb/Nono-deficient mice described a dramatic flattened nose phenotype that may mimic the severe malar hypoplasia observed in patients. Collectively, our data identify *P54NRB/NONO* as a new neurodevelopmental disease gene and support the role of DBHS proteins in brain development and function through a clinically recognizable *NONO*-deficiency syndrome. Finally, they further argue that altered circadian clock function may contribute to the pathogenesis of intellectual disability.

1345S

Gene expression analysis of methamphetamine addicted and schizophrenic patients in correlation with their psychiatric symptoms. *A. Haghigatfard[1], M. Amini faskhodi[2].* 1) Department of biology, Science and Research Branch, Islamic Azad University, Tehran, Iran; 2) Department of biology, Tehran medical Branch, Islamic Azad University, Tehran, Iran.

Methamphetamine is a neurotoxin and potent psycho-stimulant of the phenethylamine and amphetamine classes that is used to treat attention deficit hyperactivity disorder (ADHD) and obesity. Schizophrenia is a chronic multifactorial and multigenic , disabling psychiatric disorder that affects about 1% of the population worldwide. Symptoms of schizophrenia including positive, negative and cognitive deficits. majority of methamphetamine addicted have the same psychiatric symptoms like hallucination and bizarre behavior. Researches suggesting that some parts of brain in schizophrenic patients and methamphetamine addicts are neurodegenrated but the mechanisms of this degeneration is not clear. gene expression analysis could help to understand the molecular mechanisms of this neurodegeneration and explain the cause of same psychiatric symptoms in these two groups .this researches also can help to understand the pathogenesis of schizophrenia and methamphetamine effect on brain. We have studied Methamphetamine affects on gene expression in individuals who are addicted (several time in one year). also 50 schizophrenic patients with same age. blood samples collected from 50 addicted ,50 schizophrenic patients and 50 normal person in the same range of age(23-30years old). RNA of whole blood extracted by column method kit . cDNA synthesized and expression of 3000 gene investigated with DNA microarray technique. expression of 12 gene had been affected in patients and addicts . over expression of NDUFS1 and NDUFS2 , and decreasing of expression in 5HT2a,DRD2,DRD3,BDNF,DISC,NRG1,MAOA,bcl2,NDUFV1 and NDUFV2 . Over expressions were significant in schizophrenics & addicts in compare with normal but Over expressions were significant in addicts in compare with schizophrenic patients. Microarray results for two gene NDUFS1 and NDUFS2 confirmed by Real Time PCR syber green method and results analyzed by SPSS and Bonferroni statistical software. In our samples a significant affects of Metamphetamin and schizophrenia on expression of several genes especially in mitochondria complex 1 subunits genes have been detected.It seems that the mechanism of neurodegeneration in both group is the same but rate of degeneration is higher in methamphetamine addicts . May be that's why in PANSS psychiatric test the negative symptoms were in same scores for both group but positive symptoms including hallucination and exaggerated behaviors got higher scores in addicts.

1346M

Early-onset Behr syndrome due to compound heterozygous mutations in OPA1. *P. Amati-Bonneau[1], E. Colin[1], F. Oca[1], M. Ferré[1], A. Chevrollier[1], N. Gueguen[1], V. Desquiret-Dumas[1], S. N'Guyen[2], M. Barth[1], X. Zanlonghi[3], M. Rio[4], I. Desguerre[5], C. Barnerias[5], M. Momtchilova[6], D. Rodriguez[7], A. Slama[8], G. Lenaers[9], V. Procaccio[1], P. Reynier[1], D. Bonneau[1].* 1) Unité Mixte de Recherche CNRS 6214 - INSERM 1083, Département de Biochimie et Génétique, Centre Hospitalier Universitaire, Angers, France; 2) Département de Neuropédiatrie, Centre Hospitalier Universitaire, Angers, France; 3) Clinique Sourdille, Nantes, France; 4) AP-HP, Service de Génétique Médicale, Hôpital Necker, Paris, France; 5) AP-HP, Service de Neuropédiatrie, Hôpital Necker, Descartes University Paris O5, Paris, France; 6) AP-HP, Service d'Ophtalmologie, Hôpital Armand Trousseau, HUEP, Paris France; 7) AP-HP, Service de Neuropédiatrie & Centre de Référence de Neurogénétique, Hôpital Armand Trousseau, HUEP; UMPC Univ Paris 06; Inserm U1141; Paris, France; 8) Laboratoire de Biochimie, AP-HP Hôpital Kremlin-Bicêtre, Paris, France; 9) Institut des Neurosciences de Montpellier, INSERM U1051. Hôpital Saint Eloi, Montpellier, France.

The Behr syndrome (MIM#210000) is characterized by the association of early-onset optic atrophy with spinocerebellar degeneration resulting in ataxia, pyramidal signs, peripheral neuropathy and developmental delay. Although the disorder is believed to be inherited in an autosomal recessive manner, it may be clinically heterogeneous, encompassing several genetic aetiologies and patterns of inheritance. Recently, an adult-onset Behr-like syndrome, including optic atrophy and ataxia, was reported in two brothers carrying a heterozygous mutation in the optic atrophy type 1 (OPA1) gene. Heterozygous mutations in OPA1, a gene encoding for a dynamin-related GTPase involved in mitochondrial dynamics and mtDNA maintenance, are the main causes of autosomal dominant optic atrophy (DOA). In DOA, the optic neuropathy occurs insidiously in the first decade of life leading to various levels of visual impairment. As many as 20% of patients with DOA exhibit extra-ocular neuromuscular signs including deafness, chronic progressive external ophthalmoplegia, ataxia, peripheral neuropathy and mitochondrial myopathy with multiple mtDNA deletions, also called the "DOA plus" phenotype. Apart from these autosomal dominant forms, only a few syndromic cases have so far been reported with compound heterozygous OPA1 mutations suggestive of either recessive or semi-dominant patterns of inheritance. However, the clinical spectrum of these emerging double-mutant OPA1-related disorders remains to be characterized. We here report four cases of children affected by the Behr syndrome associated with compound heterozygous OPA1 mutations. The four unrelated children are affected with a strikingly similar early-onset neurological syndrome associating severe visual impairment due to optic atrophy (4/4), cerebellar ataxia with cerebellar atrophy evidenced by brain MRI (4/4), peripheral neuropathy (4/4), digestive involvement (2/4) and deafness (1/4).These results confirm the importance of searching a compound heterozygosity for OPA1 in severe paediatric cases of complicated optic neuropathy.

1347T

Association of common variants in *CCM* genes with disease severity in familial Cerebral Cavernous Malformations Type 1. *H. Choquet[1], L. Pawlikowska[1,2], J. Nelson[1], C.E. McCulloch[3], A. Akers[4], B. Baca[5], Y. Khan[5], B. Hart[6], L. Morrison[5,7], H. Kim[1,2,3], Brain Vascular Malformation Consortium (BVMC).* 1) Department of Anesthesia and Perioperative Care, University of California San Francisco (UCSF), San Francisco, CA., USA; 2) Institute for Human Genetics, UCSF, San Francisco, CA, USA; 3) Department of Epidemiology and Biostatistics, UCSF, CA, USA; 4) Angioma Alliance, Durham, NC, USA; 5) Department of Neurology, University of New Mexico (UNM) , Albuquerque, NM, USA; 6) Department of Radiology, UNM, Albuquerque, NM, USA; 7) Department of Pediatrics, UNM, Albuquerque, NM, USA.

Objective: Familial cerebral cavernous malformation (CCM) is an autosomal dominant disease caused by mutations in *CCM1*, *CCM2* or *CCM3* genes, and characterized by multiple brain lesions that often result in intracerebral hemorrhage (ICH), seizures, and neurological deficits. Carriers of the same genetic mutation can present with variable symptoms and severity of disease, suggesting the influence of modifier factors. As the three CCM proteins interact and can form a CCM1-CCM2-CCM3 protein complex in vitro, the purpose of this study was to investigate whether common variants in *CCM* genes modify disease severity, as manifested by ICH and greater total or large lesion counts, in a genetically homogenous group of CCM1 patients. Methods: Hispanic CCM1 subjects (n=188) harboring the common Hispanic mutation (CHM, Q455X) in *CCM1* were analyzed. Clinical assessment and cerebral MRI were performed at enrollment to determine ICH and lesion counts. Samples were genotyped on the Affymetrix Axiom Genome-Wide LAT1 Human Array. We analyzed 25 common variants (MAF≥1%) in the 3 *CCM* gene loci for association with ICH as well as with total and large (≥5 mm in diameter) lesion counts. Gene loci were defined as ± 10 kb upstream and downstream of the sequence. Results: At baseline, 30.3% of CCM1-CHM subjects had ICH, with a mean ± SD of 60.1 ± 115.0 (range 0 to 713) for total lesions and 4.9 ± 8.7 (range 0 to 104) for large lesions. The minor allele of rs7777835 at the *CCM2* locus was modestly associated with fewer number of total lesions (22% fewer, P= 0.037) at baseline, independent of age and gender and adjusted for family structure. Common variants in *CCM1* and *CCM3* were not associated with ICH nor with total or large lesion counts. Conclusions: Common variants in the 3 *CCM* genes do not seem to contribute strongly to the variability of CCM1 disease severity; however, larger studies are needed to confirm these findings.

1348S

Change of neuronal gene expression by administration of various anti-depressant in primary neocortical neurons. *N.A. Nguyen, K. Miyake, T. Kubota.* Epigenetic Medicine, University of Yamanashi, Chuo, Yamanashi, Japan.

Anti-depressants are widely used for treatment of depression. Among many anti-depressants, imipramine, a representative anti-depressant, is recently known to have an effect to restore the expression of Brain-derived neurotrophic factor (*BDNF*) gene, which is decreased in the patients with depression, by increasing histone acetylation. However, it is largely unknown whether other anti-depressants have similar epigenetic effect to change gene expression. To address this question, we examined the effect of five anti-depressants (imipramine, citalopram, fluoxetine, duloxetine, amitriptyline) to genes associated with depressive status. As a result, quantitative RT-PCR assay showed that all five anti-depressants increased *BDNF* expression by 48-hour treatment in primary neocortical neurons in mice, but did not increase glucocorticoid receptor (*GR*) gene expression, which is known to be decreased by mental stress during infancy in hippocampus region in mice. These results indicate that one of major mechanism among anti-depressants may be restoration of *BDNF* expression in the brain. We are currently performing comprehensive gene expression and histone modification analyses to know the differences of effect between the anti-depressant. Our study will shed a new insight in understanding of the molecular and epigenetic mechanism of anti-depressants.

1349M

Premorbid psychiatric diagnosis in young persons with 22q11.2 deletion syndrome who later developed schizophrenia. *E. Chow[1,2,3], A.S. Bassett[2,3,4].* 1) Clinical Gen Service, Ctr Addiction & Mental Hlth, Toronto, ON, Canada; 2) Clinical Gen Res Prog, Ctr Addiction & Mental Hlth, Toronto, ON, Canada; 3) Dept of Psychiatry, Univ of Toronto, ON, Canada; 4) Dalglish Hearts & Minds Clinic for Adults with 22q11DS, UHN, Toronto, ON, Canada.

Background: 22q11.2 deletion syndrome (22q11.2DS) is a common human microdeletion syndrome known to be associated with high rates of psychiatric disorders. Attention deficit hyperactivity disorder (ADHD) has been found in up to 60% of children with 22q11.2DS and approximately 25% of adults with the syndrome suffers from schizophrenia. Mood and anxiety disorders are also common. There is however limited literature on the predictors of schizophrenia in this high risk population. Objective: To investigate the relationship between premorbid non-psychotic disorders and the subsequent development of schizophrenia in a group of young individuals with 22q11.2DS. Methods: Twenty-five children and adolescents with 22q11.2DS (10 M, 15 F; mean age at first assessment=14.4 years, SD=3.2 years) at a psychiatric genetics clinic who were found not to have a DSM-IV psychotic disorder on assessment were followed prospectively for the development of psychosis on an average of 10.8 years (SD 3.7 years) as part of a research study. All subjects were at least 20 years of age at last assessment. Six subjects (23.1%) have developed schizophrenia or schizoaffective disorder (SZD) so far, with an average age at onset of 19.3 years (SD=3.1 years). Rates of lifetime (for the non-psychotic group) or pre-morbid (at least one year prior to onset of SZD) non-psychotic psychiatric diagnosis were compared between the two psychosis outcome groups. Results: Subjects who eventually developed SZD had similar rates of lifetime/premorbid ADHD, major depressive disorder, anxiety disorder, substance use disorder or an intellectual disability. They also had similar rates of prodromal positive and negative symptoms. However, subjects who eventually developed schizophrenia had a significantly higher rate of oppositional defiant disorder than those who remained psychosis-free (50.0% vs 5.3%, p<0.01). Conclusions: Although the sample size is small, results from this study would suggest that behavioural problems in childhood and adolescence may be predictive of the subsequent development of schizophrenia in 22q11.2DS, as for schizophrenia in the general population. Further research into this area is needed.

1350T

Association between advanced paternal age and early onset of schizophrenia among sporadic cases. *S. Wang[1], L. Yeh[2], C. Liu[3], C. Liu[3], T. Hwang[3], M. Hsieh[3], Y. Chien[3], Y. Lin[3], S. Chandler[4], S. Glatt[5], N. Laird[6], S. Faraone[5], M. Tsuang[4], H. Hwu[3], W. Chen[1].* 1) Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan; 2) Department of Healthcare Administration, College of Health Science, Asia University, Taichung, Taiwan; 3) Department of Psychiatry, College of Medicine and National Taiwan University Hospital, National Taiwan University, Taipei, Taiwan; 4) Center for Behavioral Genomics, Department of Psychiatry, University of California San Diego, La Jolla, California, USA; 5) Departments of Psychiatry and Behavioral Sciences and Neuroscience and Physiology, Medical Genetics Research Center, SUNY Upstate Medical University, Syracuse, New York, USA; 6) Department of Biostatistics, Harvard University, Boston, Massachusetts, USA.

Background: Advanced paternal age is associated with increased risk of schizophrenia, and the association has been attributed to the increased risk of de novo mutations in paternal germ cells. Age at onset is an important component phenotype for schizophrenia. Patients with early-onset schizophrenia have a higher incidence of structural genomic variation than those with later onset. If early-onset schizophrenia could be associated with genetic factors, the increased number of de novo mutations along with paternal age could be associated with not only the elevated risk of schizophrenia but also the earlier onset of the disorder. Aim: This study aims to explore whether increased paternal age is associated with early onset in schizophrenia among sporadic cases. Methods: A total of 2450 patients with schizophrenia with no first-degree relatives affected with psychiatric disorders were recruited from the Schizophrenia Trio Genomic Research in Taiwan (S-TOGET) project. Early onset was defined as those below 18 years of age of the first psychotic episode. The association between paternal age and early onset in schizophrenia was estimated by a logistic regression model adjusting for confounders. Results: A U-shape relationship was observed between paternal age and proportion of early onset in schizophrenia. The proportion of early onset for patients with paternal age < 20 years, about 30%, is larger than that for patients with paternal age of 20-25 years, about 20%. As paternal age increased over 25, increased paternal age exhibited a linear increase in the proportion of early onset in schizophrenia: increased to about 35% for a paternal age between 25-30 and to about 45% for a paternal age greater than 45, with a statistical significance for the trend test adjusting for maternal age and other potential confounders. Conclusions: These findings indicate that paternal age greater than 25 and younger than 20 were both associated with early onset in schizophrenia among sporadic cases. The associations of advanced paternal age with both increased susceptibility to schizophrenia and early onset in schizophrenia are consistent with the rate of increases in spontaneous mutations in sperm as men age.

1351S

Cyclooxygenase-2 non-selective inhibitor prior to pentylenetetrazole-induced seizures increases the latency to seizure onset and decreased the number of seizures in zebrafish. *P. Barbalho, H.M Gomide, D.M Nakata, I.L Cendes, C.V Maurer-Morelli.* Department of Medical Genetics, State University of Campinas (UNICAMP)), Campinas, Sao Paulo, Brazil.

Introduction: Neuroinflammatory response may play a role in the pathophysiology of epilepsy. It has been demonstrated that cyclooxygenase-2 (COX-2) increases in response to an insult, as seizure. The aim of this study was investigated the effects of indomethacin (a COX-2 non-selective inhibitor) prior to pentylenetetrazole (PTZ)-induced seizure on mRNA expression of both cox-2 genes in zebrafish larvae brain. Methods: This study was approved by Animal Ethical Committee/UNICAMP (# 3098-1). At six dpf, zebrafish larvae were separated into Seizure+Indomethacin (SG+Indo; n=5) and Control+indomethacin (CG+indo; n=5) groups and incubated in indomethacin solution (110µg/ml) for 24 hours. After the incubation, animals from SG+Indo were exposed to 15mM PTZ during 20 minutes and the number of seizure-like behavior and latency to seizure onset (Stage 3 of seizure) were analyzed. Animals from CG+indo were handled in PTZ-free water. Other groups: SG - only exposed to PTZ (n=5) and CG- control (n=5) were also investigated. Each sample was composed by pooling 20 heads. Immediately after seizure, animals were anesthetized and their heads collected for reverse transcriptase quantitative-PCR amplifications that were carried out in triplicates with ef1α as endogenous controls using TaqManTM System. The relative quantification (RQ) was calculated by the equation RQ=2 -$\Delta\Delta$CT. Statistical analyses were performed by Mann-Whitney test with p<0.05. Results: The mean±SEM obtained were: (ii) cox2a: SG+indo 0.5±0.06 vs SG 1.3±0.12 (p=0.004); (ii) cox2b: SG+indo 0.73±0.06 vs SG 1.73±0.18 (p=0.004). Interestingly, when exposed to PTZ, animals pre-treated with indomethacin showed longer latency to reach seizure: SG+indo: 4.6±0.33 vs SG: 2.92±0.17 (p=0.0004) and presented less number of seizure-like behavior response when compared to SG (SG+indo: 11.2±1.5 vs SG: 38.16±4.5; p=0.003). Conclusion: Indomethacin treatment prior to PTZ-induced seizure reduced the cox2a and cox2b mRNA expression levels compared with non-treated animals (SG). Besides, the pharmacological cox-2 inhibition increased the latency to seizure onset and significantly decreased the number of seizures during PTZ exposure compared to SG. Our findings support evidence that zebrafish is a valuable model for further investigations of the main role of inflammation in seizure, as well as a valuable model for anti-inflammatory screening of compounds that are potentially therapeutic for seizures. Support: FAPESP and CNPq.

1352M

miR-1202: A Primate Specific and Brain Enriched miRNA Involved in Major Depression and Antidepressant Treatment. *J.P. Lopez[1,2,3], R. Lim[4], C. Cruceanu[1,2,3], L. Crapper[2], C. Fasano[3], B. Labonte[2], G. Maussion[2], J.P. Yang[2], V. Yerko[2], E. Vigneault[3], S. El Mestikawy[3], N. Mechawar[2], P. Pavlidis[4], G. Turecki[1,2,3].* 1) Human Genetics, McGill University, Montreal, Quebec, Canada; 2) McGill Group for Suicide Studies, Douglas Mental Health University Institute, McGill University, Montreal, Quebec; 3) Douglas Mental Health University Institute, McGill University, Montreal, Quebec, Canada; 4) Department of Psychiatry, University of British Columbia, Vancouver, B.C, Canada.

Major depressive disorder (MDD), is a prevalent mood disorder that associates with differential prefrontal brain expression patterns. Treatment of MDD includes a variety of biopsychosocial approaches, but in medical practice, antidepressant drugs are the most common treatment for depressive episodes, and not surprisingly, they are among the most prescribed medications in North America. While they are clearly effective, particularly for moderate to severe depressive episodes, there is important variability in how individuals respond to antidepressant treatment. Failure to respond has important individual, economic and social consequences for patients and their families. Several lines of evidence demonstrate that genes are regulated through the activity of microRNAs (miRNAs), which act as fine-tuners and on-off switches in gene expression patterns. Here we report on complementary studies using postmortem human brain samples, cellular assays and samples from clinical trials of depressed patients, and show that miR-1202, a miRNA specific to primates and enriched in the human brain, is differentially expressed in depressed individuals. Additionally, miR-1202 regulates the expression of the Metabotropic Glutamate Receptor 4 (GRM4) gene and predicts antidepressant response at baseline. These results suggest that miR-1202 is associated with the pathophysiology of depression and is a potential target for novel antidepressant treatments.

1353T

The NINDS Repository Biomarker Discovery Collection is a Public Resource for Neurodegenerative Disorders. *G. Balaburski[1], S. Heil[1], A. Green[1], C. Kopil[3], M. Fraser[3], M. Sutherland[2], K. Gwinn[2], R. Corriveau[2], C. Tarn[1].* 1) Coriell Institute for Medical Research, Camden, NJ; 2) National Institute for Neurological Disorders and Stroke-NIH, Bethesda, MD; 3) Michael J Fox Foundation for Parkinson's Research, NY, NY.

Neurological diseases present numerous challenges: they are devastating to patients and their families, present unique research, diagnostic and clinical difficulties and are an increasing burden on health care systems. Biomarkers for diagnosis, disease onset, disease progression and therapeutic response are urgently needed to improve research, as well as clinical and diagnostic outcomes of patients with neurodegenerative diseases such as Parkinson's and Huntington's disease. The mission of the NINDS Repository is to provide genetic support for scientists investigating pathogenesis of the central and peripheral nervous systems through submission and distribution of biomaterials and de-identified clinical data. Moreover, the NINDS Repository facilitates identification and validation of biomarkers for neurological diseases by virtue of being the centralized facility for storage, processing and distribution of biofluids (cerebrospinal fluid, plasma, serum, whole blood, urine) and nucleic acids (DNA and RNA) extracted from whole blood. The Repository minimizes pre-analytical variables by establishing unified standards for sample collection and shipping, securely stores samples with real time monitoring and recording systems, performs standardized laboratory processing with validated operating procedures and maintains the highest levels of quality control and quality assurance. To further incite the investigation and discovery of novel biomarkers the NINDS Repository is establishing large, long term longitudinal collections of biological samples obtained from affected and neurologically healthy individuals, as well as from individuals with known genetic mutations. Currently, the NINDS Repository collects samples under either NINDS sponsored biomarker initiatives as well as in collaboration with partners such as the Michael J. Fox Foundation. Current biomarker discovery studies include: Parkinson's Disease Biomarkers Program (PDBP), Neurobiological Predictors of Huntington's Disease study (PREDICT-HD), Frontotemporal Dementia MAPT Carrier Study and the Fox Investigation for New Discovery of Biomarkers (BioFIND). Thus, the NINDS Repository is a public resource that freely distributes biomaterials for research, discovery and validation of biomarkers for neurological disorders. Samples are available upon request from the NINDS Repository web catalog (http://ccr.coriell.org/NINDS), or via NIH-sponsored resources with links to the online catalog.

1354S

SP1 inhibitors as modulators of APP and BACE1 levels in human cells: A novel drug target in Alzheimer's disease. *B.L. Bayon[1,2], J.A. Bailey[2], B. Ray[2], K. Sambamurti[3], N.H. Greig[4], D.K. Lahiri[1,2].* 1) Medical & Molecular Genetics, Indiana University School of Medicine , Indianapolis, IN; 2) Laboratory of Molecular Neurogenetics, Department of Psychiatry, Neuroscience Research Building, Indiana University School of Medicine, 320 W 15th St. Indianapolis, IN, USA; 3) Department of Neurosciences, Medical University of South Carolina, Charleston, South Carolina, USA; 4) National Institute on Aging, National Institutes of Health Translational Gerontology Branch, Intramural Research Program, District of Columbia, USA.

According to the amyloid hypothesis, abnormalities in AD are believed to result from the over-production of amyloid-β peptide (Aβ), a product of Aβ precursor protein (APP). Dysregulation of proteins involved in the production of Aβ such as APP and β-site APP-cleaving enzyme (BACE1), may contribute to excessive Aβ deposition. The rate-limiting step in the production of Aβ is the processing of APP by BACE1. Understanding how expression of these proteins is regulated will eventually expose new drug targets. The transcription factor specificity protein 1 (SP1) coactivates the expression of the APP and BACE1 gene. We tested SP1-mediated regulation of APP with Mithramycin A, a selective inhibitor of SP1, and Tolfenamic acid, an inducer of SP1 degradation in human glioblastoma cells U373 and in human neurosphere (NSP) cultures. NSPs were cultured in Neurocult basal media plus differentiation supplement (Stem Cell Technologies). U373 (ATCC) cells were cultured and transfected, and Western blot analysis was performed as previously described (Long et al., JBC-2014). Mithramycin A (Santa Cruz) and Tolfenamic acid (Sigma Aldrich) were prepared in 1 μM and 5 μM doses. After 72-hour treatment or transfection, cell viability was assessed using CTG assay (Promega), and protein lysates made. Western blot analysis reveals a significant decrease in the expression of APP in U373 and NSP treated with Mithramycin A. NSP treated with Mithramycin A also exhibit a decrease in BACE1 expression. Treatment with Tolfenamic acid, however, does not significantly decrease APP or BACE1 expression in either cell model. APP siRNA effectively knocks down APP expression in U373 and NSP cultures. BACE1 siRNA and SP1 siRNA did not significantly affect APP levels. CTG showed no significant changes in cell viability among treatment groups in U373 and NSP. We show that expression of APP is decreased after treatment with the SP1 inhibitor Mithramycin A in both U373 and human neurospheres cells. However, APP expression is not affected by treatment with Tolfenamic acid, perhaps due to the differences in the mechanisms between these SP1-inhibiting drugs. We also show that transfection with siRNAs can effectively change the expression of APP and BACE1 in both the human cells. It is essential to ascertain whether drugs or small RNAs targeting this transcription factor could be used to effectively decrease amyloid load and possibly the symptoms of AD in patients.

1355M

Association of Serotonin 2c Receptor Polymorphisms with Antipsychotic Drug Response In Schizophrenia. *J. Li[1], H. Hashimoto[2], H.Y. Meltzer[1].* 1) Psychiatry, Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Graduate School of Pharmaceutical Sciences, Osaka University, Japan.

The serotonin (5-HT)2C receptor (HTR2C) has been implicated in schizophrenia and response to antipsychotic drugs (APDs) through its regulatory effect on dopamine release, interaction with scaffolding proteins at the synapse, and other unknown mechanisms. Evidence from genetic association studies also implicates HTR2C in a variety of neuropsychiatric diseases. We tested the association between HTR2C polymorphisms, Cys23Ser, -759C/T, and -697G/C, and treatment response in 171 schizophrenic patients after treatment with APDs, mostly clozapine, for 6 months. We confirmed that -759C/T, but not Cys23Ser, was a *cis*-eQTL for HTR2C according to Braincloud data, an integrated database of genome-wide gene expression and genetic control in human postmortem dorsolateral prefrontal cortex (DLPFC) of normal subjects. Ser23 was significantly associated with treatment response at 6 months (positive symptoms, $X^2 = 7.540$, $p = 0.01$; negative symptoms, $X^2 = 4.796$, $p = 0.03$) in male, but not in female patients. Haplotype analysis showed that -759C-Ser maintained the same level of significant association with positive symptom improvement ($X^2 = 6.648$, $p = 0.01$) but additive association with negative symptom improvement ($X^2 = 6.702$, $p = 0.01$). Logistic regression after controlling for covariates showed these haplotypic associations remained significant with the same direction. Finally, a meta-analysis was performed on six studies with accessible genotyping data for rs6318 and treatment outcome. The overall odds ratio under fixed effect model is 2.00 (95%CI, 1.38-2.91, $p = 0.0003$) and under random effect model is 1.94 (95%CI, 1.27-2.99, $p = 0.0024$). In conclusion, HTR2C polymorphisms were associated with treatment response to clozapine in male schizophrenic patients. HTR2C could be relevant to a broad range of the psychopathology which responds to clozapine in schizophrenia.

1356T

First case of Spinocerebellar Ataxia type 1 in a Mexican female. *I. Cervantes[1], C.M. Morán[2], J. Sanchez[2], G. Castañeda[4], D.M. Sanchez[5], M.H. Orozco[1,2], N.Y Nuñez[1], J.A. Alcaraz[1,2], I.P. Dávalos[3], N.O. Dávalos[1], S. Ramírez[6], J. Corral[7], H. San Nicolás[7], L. de Jorge[7], V. Volpini[7], D. García[1].* 1) Instituto de Genética Humana, CUCS, Doctorado en Genética Humana Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) División de Medicina Molecular, CIBO, CMNO, IMSS, Guadalajara, Jalisco; 3) División de Genética, CIBO, CMNO, IMSS, Guadalajara, Jalisco; 4) Servicio de Neurocirugía, UMAE Hospital de Especialidades, CMNO, Jalisco; 5) Psicología, ITESO, Tlaquepaque, Jalisco; 6) Universidad de la Sierra Sur, Mihuatlán de Porfirio Dïaz, Oaxaca; 7) Centro de Diagnóstico Genético Molecular, IDIBELL, Barcelona, España.

Introduction. Spinocerebellar ataxia type 1 (MIM ID #164400) is an autosomal dominant disease caused by expansion of the CAG trinucleotide repeats in the ataxin-1 gene. Clinically characterized by cerebellar signs, upper motor neuron signs and extensor plantar responses, scanning speech, incoordination, slow motor-nerve conduction and choreiform movements in some instances. Material and Methods. A female aged 24 year-old was studied clinically due to SCA with 10 years of evolution which presented gait ataxia, dysmetria, dysarthria, disdiadochokinesis, scanning speech, loss of deep tendon reflexes, mild hypotonia, hypermetric bilateral saccades and normal ocular fundus; the electromyography reported sensorimotor polyneuropathy with segmental demyelination. Brain MRI scans revealed olivopontocerebellar atrophy. Laboratorial studies were performed with 5mL of peripheral blood were used to isolate patient's DNA by GeneCatcher Kit (invitrogen). The molecular analysis was made by PCR. Results. Molecular analysis revealed a number of expansion repeats of 32/*exp. Discussion. In previous studies done in Mexican population there was not detected any patient with SCA1, so its frequency is unknown. This is the first time that is described a patient affected by SCA1 in Mexico.

1357S

Allelic distribution of the normal ATXN10 gene in a sample of a Peruvian Amerindian population: an exploratory study. *D. Veliz-Otani[1], O. Ortega[1], M. Cornejo-Olivas[1,2], K. Milla-Neyra[1,3], K. Espinoza-Huertas[1], S. Lindo-Samanamud[1], M. Inca-Martinez[1], V. Marca[1], I. Tirado[1,3], M.L. Saraiva-Pereira[4,5], L. Jardim[4,5], P. Mazzetti[1,6], Rede Neurogenetica.* 1) Neurogenetics Research Center, Instituto Nacional de Ciencias Neurologicas, Lima, Peru; 2) Northern Pacific Global Health Research Training Consortium, Bethesda, MD; 3) Genetics and Biotechnology School, Universidad Nacional Mayor de San Marcos. Lima, Peru;; 4) Laboratório de Identificação Genética - Centro de Pesquisa Experimental e Serviço de Genética Médica, HCPA;; 5) Programa de Pós-Graduação em Genética e Biologia Molecular, UFRGS; 6) School of Medicine, Universidad Nacional Mayor de San Marcos. Lima, Peru.

Background: SCA10 is caused by an ATTCT abnormal expansion (locus 22q13.31.) of the *ATXN10* gene , and has been diagnosed only in patients from North and South Americas. Up to now, for instance, we have diagnosed SCA10 in 16 Peruvian SCA families, becoming the most frequent dominant ataxia reported in this population, which has an average of 70% of Amerindian ancestry. Normal alleles range from 10 to 29 ATTCT repeats, and the mutant ones are from 800-4500 ATTCT repeats. Large normal alleles of other SCA-related genes seem to be prone to expanding to a pathogenic range, acting as natural sources of *de novo* mutations. **Objective**: To describe the normal allelic distribution as well as the frequency of large normal alleles (≥17 repeats) in the *ATXN10* gene in healthy controls of an Amerindian "Quechua" community. **Methods**: Observational exploratory study collecting DNA samples from 49 healthy individuals from an Amerindian "Quechua" population from Puno, Peru. ATTCT repeats of the *ATXN10* gene were genotyped by conventional PCR followed by polyacrylamide gel electrophoresis; ATTCT repeat counting was carried out by means of a lineal model based on normal controls of known repeat size previously measured by capillary eletrophoresis. The probability of sampling low-frequency alleles was estimated using Minsage software. **Results**: We included a total of 98 alleles from 49 individuals. The sample size yielded a 95% probability of sampling an allele with 5.8% population frequency under H-W equilibrium or 10.1% frequency assumming no H-W equilibrium. The allele distribution ranged from 11 to 17, the mean was 14.1 ATTCT repeats . The most frequent allele was 14 (45.92%), and the frequency of large normal alleles (≥17 repeats) was 3 (3.06%) under no H-W equilibrium (P-value=0.018). **Conclusion**: This preliminary data shows a low frequency of large normal alleles in this population, suggesting that frequency of large normal alleles is not related to the high prevalence of SCA10 in Peruvian population. Further analysis with larger sample size and complimentary approaches are required to confirm these results.

1358M

Molecular characterization of genes modifying the age at onset in Huntington's Disease in a group of patients from Uruguay. *P. Esperon Percovich[1], M. Vital Coppa[1], E. Bidegain Culasso[1], V. Raggio Risso[2].* 1) Molecular Biology Unit Clinical Biochemistry Department School of Chemistry, General Flores 2124Ave, Montevideo, Uruguay; 2) Genetic Department School of Medicine, General Flores 2124Ave, Montevideo, Uruguay.

Background: Huntington's disease (HD, OMIM #143100) is a neurodegenerative disorder with an autosomal dominant inheritance mode, complete penetrance and variable clinical expressivity. The genetic cause of HD is an expansion of a sequence of CAG trinuc leotide repeats located in the first exon of the *IT*15 gene (region 4p16.3). Though the number of CAG repeats can largely explain the age at onset (AAO), a percentage of variation on AAO persists and can be attributed to modifier genes. Aim: To evaluate the influence on the AAO of HD of: the CAG repeats number and del2642 in the *IT*15 gene; the E2, E3 and E4 *APOE* gene alleles, and *ADORA2A* rs5751876, *HAP*-1 rs4523977, *PGC*-1 rs7665116 and *UCHL*-1 rs5030732 gene polymorphisms. Patients and methods: 21 patients were recruited from different national institutions. The CAG repeat number determination and gene polymorphisms were done by polymerase chain reaction and High Resolution Melting assays, respectively. Informed consent was obtained from all patients. The study follows the Code of Ethics of the World Medical Association (Declaration of Helsinki). Results: A 63.3% variation in the AAO resulted from the CAG repeat number. The genotype for the aforementioned gene polymorphisms was analysed together with the CAG repeat number and the AAO. No influence of the del2642 of *IT*15 gene, *UCHL*-1, and *APOE* gene polymorphisms was found. Out of seven patients with solely the *ADORA2A* polymorphism, four presented anticipation on the AAO. Inversely, a delay on the AAO in two patients was observed, one carrying the T allele of *HAP*-1 and one the C allele of *PGC*-1α gene. Conclusions: For the first time, in our population, the molecular diagnostic of HD was done. For a more extensive study of molecular aspects, a determination of polymorphisms strongly involved in the variation of the AAO was also performed. Considering a HD disease prevalence of 5/100.000, in a country with only a 3.3 million inhabitant population, the recruitment of a large group of patients is a big challenge. Our results, albeit qualitative, showed a novel approach to the molecular analysis of HD patients.

1359T

Deregulation of specific microRNAs in whole blood and skeletal muscle of Myotonic Dystrophy type one patients. *K.K. Ambrose, I. Taufik, I.H. Lian, K.J. Goh, K.T. Wong, A. Ahmad-Annuar, M.K. Thong.* University of Malaya, Kuala Lumpur, Malaysia.

MicroRNAs (miRNAs) are short RNA molecules of approximately 22 nucleotides that function as post-transcriptional regulators of gene expression. They are expressed in a tissue specific manner and show different expression patterns in development and disease, and hence can potentially act as disease specific biomarkers. Several miRNAs have been shown to be deregulated in plasma and skeletal muscles of myotonic dystrophy type one (DM1) patients. Here we further expand the miRNAs associated with DM1, report the differential expression of particular miRNAs in whole blood and skeletal muscle samples of DM1 patients, and discuss the potential of these miRNAs as DM1-specific biomarkers. Eleven candidate miRNAs were analysed using quantitative real-time PCR in whole blood (n=10) and muscle biopsy samples (n=9) of DM1 patients, as well as normal control samples (whole blood, n=10; muscle, n=9). In DM1 whole blood samples, miRNA 133a, 29b, and 33a were significantly up regulated, whereas miRNA 1, 133a and29c were significantly down regulated in DM1 skeletal muscle samples compared to controls. Our findings are aligned to those shown in other studies and point towards pathways that potentially contribute towards pathogenesis in DM1. However currently available data is not sufficient for these miRNAs to be made biomarkers for DM1, as they seem to be common to many muscle pathologies, hence lacking specificity. This reinforces the need for DM1 biomarker study to be further explored as we move towards less invasive, economical, disease-specific, and timely methods of biomarker detection and analysis.

1360S

FMR1-based "Double Hit" model and genomic studies in premutation carriers. *R. Lozano[1], R. Hagerman[1], M. Duyzend[2,3], E. Eichler[2,3], F. Tassone[4].* 1) MIND Institute, Dept. of Pediatrics, UC Davis Medical Center; 2) Department of Genome Sciences, University of Washington School of Medicine; 3) Howard Hughes Medical Institute Seattle; 4) Dept. of Biochemistry, UC Davis.

The FMR1 premutation is defined as having 55 to -200 CGG repeats in the 5' untranslated region of the Fragile fragile X Mental mental Retardation retardation 1 gene (FMR1). The clinical involvement has been well characterized for Fragile fragile X-associated Tremortremor/Ataxia ataxia Syndrome syndrome (FXTAS) and Fragile fragile X-associated Primary primary Ovarian ovarian Insufficiencyinsufficiency (FXPOI). The behavior/psychiatric and other neurological manifestations remained to be specified as well as the molecular mechanisms that will explain the phenotypic variability observed in individuals with the FMR1 premutation. We describe a pilot study of copy number variants (CNVs) in 56 participants with a premutation ranging from 55 to 192 repeats. The participants were divided into 4 four different clinical groups for the analysis: those with behavioral problems but no autism spectrum disorders (ASD), those with ASD but without neurological problems, those with ASD and neurological problems including seizures, and those with neurological problems without ASD. Results: We found 12 rare CNVs (8 eight duplications and 4 four deletions) in 11 cases (19.6%) that were not found in about approximately 8,000 controls. Three of them were at 10q26 and two at Xp22.3, with small areas of overlap. The CNVs were more commonly identified in individuals with neurological involvement and ASD. The frequencies were not statistically significant across the groups. There were no significant differences in the psychometric and behavior scores among all groups. Further studies are necessary to determine the frequency of second genetic hits in individuals with the FMR1 premutation; however, these preliminary results suggest that genomic studies can be useful in understanding the molecular etiology of clinical involvement in premutation carriers with ASD and neurological involvement in an FMR1-base "Double Hit" model.

1361M

Determination of the origin of Huntington disease based on haplotypes in a Peruvian population. *I. Tirado[1,2], C. Kay[3], M. Cornejo-Olivas[1,4], J.A. Collins[3], M.E. Ketelaar[3], S. Lindo-Samanamud[1], M. Inca-Martinez[1], O. Ortega[1], V. Marca[1], D. Veliz-Otani[1], K. Espinoza-Huertas[1], G. Sotil[2,5], P. Mazzetti[1,6], M.R. Hayden[3].* 1) Neurogenetics Research Center, Instituto Nacional de Ciencias Neurologicas, Lima, Peru; 2) School of Genetics and Biotechnology, Universidad Nacional Mayor de San Marcos, Lima, Peru; 3) Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, Canada; 4) Northern Pacific Global Health Research Training Consortium, Bethesda, MD; 5) Department of Genetics and Cell Biology, Universidad Nacional Mayor de San Marcos, Lima, Peru; 6) School of Medicine, Universidad Nacional Mayor de San Marcos, Lima, Peru.

Introduction. Recent studies have associated specific *HTT* haplotypes with the CAG expansion causing Huntington disease (HD) in Caucasian, Asian and African populations. In Europe, HD chromosomes are predominantly associated with haplogroup A while in East Asian populations and the black subpopulation of South Africa, they are associated with haplogroups C and B respectively. There is a reported hotspot of HD cases in Peru, a Latin American country of admixed population with predominance of Spanish European and Amerindian ancestry. **Aims.** To determine the origin of HD in the affected Peruvian population based on *HTT* haplogroups and the relative admixture of Amerindian, European, Asian and African genetic background. **Methods.** Haplotypes of 44 HD chromosomes and 90 control chromosomes from Peruvian HD pedigrees were reconstructed from genotypes at 91 SNPs across the *HTT* gene region, IRB approval was obtained at Instituto Nacional de Ciencias Neurologicas. **Results.** HD chromosomes were found mainly on haplogroup A (85%), most specifically on the variant A1 (75%). HD occurs on A1 at the highest frequency (>90%) in the regions of Lima and Central Peru. By contrast, a small proportion (14%) are associated with haplogroup C, which occurs at the highest frequency in the South of the country, where Amerindian ancestry predominates. **Conclusions.** HD chromosomes from the Peruvian population are predominantly associated with haplogroup A, suggesting a European Caucasoid origin. The locally high frequency of HD on A1 in Lima and Canete, exceeding the frequency of A1 in previously studied Caucasian cohorts, may indicate a founder effect of HD in these regions. The few Peruvian HD haplotypes on C, common among East Asian HD, are likely of Amerindian origin.

1362S

Event-level quantification of alternate splicing using junction reads identifies new splicing QTLs in RNA-seq data from 1000 Genomes Project samples. *T. Bhangale, L. Goldstein, B. Friedman, R. Gentleman, T. Wu, (TB and LG: equal contribution).* Bioinformatics & Computational Biology, Genentech Inc, South San Francisco, CA.

Common variants can lead to complex diseases by affecting RNA splicing; such splicing QTLs (sQTLs) have been shown to explain a considerable proportion of variation in isoform usage. Most current approaches to identifying sQTLs attempt to quantify differences in transcripts at the isoform level. However, since sequencing reads/fragments are much shorter than transcripts, isoform-level quantification involves either probability models of transcripts or mapping of reads to a reference transcriptome, instead of the reference genome. Both approaches introduce various challenges and pitfalls. Moreover most methods only detect sQTLs related to known isoforms. We therefore present a new approach that focuses on event-level quantification of splicing within a gene and apply it to the RNA-Seq data for the 1000 Genomes Project lymphoblastoid cell lines from Geuvadis Consortium (GC). We first map reads to the reference genome using our aligner GSNAP, which performs splice-junction-aware split-read alignments, using known splice sites as well as novel splice sites it predicts. Using known variants in the sample it also performs SNP-tolerant alignment and enables allele-specific expression analysis by eliminating the common problem of reference bias. These alignments are then analyzed by our new method exonfindR, which detects known and novel (i.e. ones that cannot be inferred from known transcript annotations) alternate splicing events and quantifies their expression levels in the sample. Quantification is in terms of local events e.g., skipped exon, retained intron etc, and only uses reads that provide unambiguous support for a given event i.e. splice-junction reads in most cases. Association tests are then performed to identify sQTLs. In the data from GC, we detected 37,121 known events in 14,527 genes and 3,703 novel events in 2,515 genes. At FDR cutoff of 5% in the 366 samples of European descent, we found 845 genes (known events) and 81 genes (novel events) with common (MAF $\geq$ 5%) cis sQTLs, where GC had previously reported 620 genes. Among our sQTLs, 38 (25 previously unreported) disrupt splice sites and likely provide a direct hypothesis for a mechanism and 35 overlap with SNPs in the NHGRI GWAS catalog. Advantages of our approach include greater power to detect association when the variant leads to an event-level change in splicing and the ability to detect novel events and sQTLs associated with them. Both tools, GSNAP and exonfindR, are freely available.

1363M

Detection and prediction of deleterious mutations affecting pre-mRNA processing. *D.S. Hanna, J.G. Underwood, J.D. Smith, M.O. Dorschner, D.A. Nickerson.* Genome Sciences, University of Washington, Seattle, WA.

Synonymous and intronic variation is often ignored during human genetic analyses although such variation is known to alter signals recognized by the spliceosome or its associated trans-acting factors. This variation can alter splice site choice or generate cryptic splice sites, both of which can result in aberrant splicing events that lead to nonfunctional coding sequence or nonsense-mediated decay. For the analysis of this underappreciated class of disease-causing variation, we present Spliceosaurus, a lightweight C++ codebase architected using the Standard Template Library (STL) for rapid analysis of variation on splicing signals. Spliceosaurus uses a novel scoring algorithm based on the maximum entropy principle further scaled by weighting cis-acting splicing enhancers and silencers by genomic distance and affinity to establish U1 and U2 snRNA recognition sites on pre-mRNAs. The algorithm iteratively searches not only for variation affecting existing splice sites, but also for variation contributing to the generation of cryptic splice sites. The model is trained on the non-canonical splicing variation found in DBASS that lead to aberrant mRNA isoforms allowing extreme specificity in the identification of variation highly impacting the splicing code. Potential irregular splicing events are interrogated by predicting new exon boundaries and the effects on the aberrant mRNA. Unlike existing splicing analysis suites, the input and output for Spliceosaurus is a standard VCF file that allows seamless integration into existing data analysis pipelines. Additionally, the code is capable of processing whole genome datasets in hours, permitting splicing analysis on the scale requjred for next generation sequencing approaches. We applied Spliceosaurus and correctly identified previously reported causal intronic and synonymous variants in several Mendelian disorders as highly disruptive to canonical splicing [i.e., Auriculo-condylar syndrome (ACS), hereditary nonpolyposis colorectal cancer (HNPCC) and X-linked parkinsonism with spasticity (XPDS)] The code is freely available for academic use on GitHub: https://github.com/davidhanna/Spliceosaurus.

1364T

An Integrative Approach to Identify Splice Factors and their Putative Upstream Regulators. *M. Subramaniam, SJ. Gosline, SK. Hughes, FB. Gertler, E. Fraenkel.* Massachusetts Institute of Technology, Cambridge, MA.

The rise of whole-genome expression analysis has demonstrated the power of understanding genomic-level changes in gene expression across diseases such as cancer. In addition to the large changes in whole gene expression observed in these genome-wide expression experiments via massively parallel pair-ended RNA sequencing, many novel splice isoforms of genes have been identified across different forms of cancer. However, identification of proteins that cause these splice isoforms is limited to *in vitro* competitive binding assays or high throughput binding experiments such as cross-linking immunoprecipitation. These assays have been helpful in identifying the space of possible protein-RNA interactions, but often fail to capture condition specific binding events. Here, we propose an integrative approach that uses sequence-level knowledge of alternative splicing events together with RNA-binding profiles and protein-protein interaction networks. We use paired-end sequencing data to identify differential splicing in a model of breast cancer, then apply a hypothesis driven scoring method that identifies putative RNA binding proteins that lead to differential splicing. We then factor in evidence from upstream protein interactions to identify the RNA binding proteins and their associated pathways implicated in the observed sequencing data. Using sequencing data collected from *Mena*-expressing epithelial breast cancer cells, we applied this method to identify and explain alternative splicing events as well as the metastatic phenotype related to the introduction of *Mena*. The results identified a network of protein interactions that connect *Mena* over-expression with key alternative splicing events unobserved in the control condition. The network results pinpointed specific splice factors likely responsible for causing splicing changes in response to activation by upstream regulators.

1365S

The Exome Coverage and Identification (ExCID) Report: a gene survey tool for clinical sequencing applications. *C. Buhay[1], R. Sanghvi[1], Q. Wang[1], K. Walker[1], H. Doddapaneni[1], J. Hu[1], M. Wang[1], Y. Han[1], H. Dinh[1], E. Boerwinkle[1,2], D. Muzny[1], R. Gibbs[1].* 1) Baylor College of Medicine - Human Genome Sequencing Center, Houston, TX; 2) 2Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.

The Exome Coverage and Identification (ExCID) Report was developed at the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) to represent gene transcript and exon sequence coverage for samples analyzed with the VCRome Whole Exome sequencing (WES) reagent. Since March 2013, the report has been used to analyze more than 13,000 WES research samples for the BCM-HGSC and more than 1500 WES clinical samples at our Whole Genome Laboratory (WGL). ExCID assesses target sequencing coverage, annotates targets with gene, transcript and exon information and reports intervals below 20X. In addition, ExCID has batch analysis features that can compare data from hundreds of samples to reveal trends in the performance of large scale sequencing projects. Results can be visualized as 'coverage tracks' in popular browsers such as the Integrative Genome Browser and the UCSC Genome Browser. For a pilot survey, coverage data was examined from 34 WES samples from the WGL. This showed that the VCRome capture reagent covered (at $\geq$20X) more than 75 percent of the medically interesting genes listed in COSMIC, HGMD, GeneTests, and ACMG Incidental Findings lists. The analysis also identified gene regions that were currently below the 20X threshold. We also analyzed data from Whole Genome Shotgun (WGS) to compare sequence coverage in WGS versus WES. ExCID output was generated on 9 WGS (30X) and 9 WES (100X) to discover and aggregate coding regions below 20X in each method. The data show regions that are poorly covered irrespective of the sequencing strategy, as well as regions unique to each strategy. Ongoing analyses include the aggregation and characterization of poorly covered regions in 600 WGS and 1000 WES samples. Results will provide insight regarding regions of the human genome that require special consideration for the development of future clinical sequencing strategies. ExCID is available under public license in the GitHub repository: https://github.com/cbuhay/ExCID.

† Trainee Award Finalist

1366M

Scaled Sparse High Dimensional Tests for Localizing Disease Susceptible Sequence Variants. *S. Cao[1,3], H. Qin[2,3], J. Li[2,3], H. Deng[2,3], Y. Wang[1,2,3].* 1) Dept of Biomedical Engineering, Tulane University, New Orleans, LA; 2) Dept of Biostatistics and Bioinformatics, Tulane University, New Orleans, LA; 3) Center for Bioinformatics and Genomics, Tulane University, New Orleans, LA.

Existing association tests are powerful to identify susceptibility variants in large samples. These tests, however, have low power in high dimensional set (HDS), which is a sample containing fewer study subjects than the number of variants tested. HDS are very common among recently deep sequencing data for human complex trait studies. Given the assumption that most of genetic variants are neutral, i.e., only a small portion of them are functional variants, sparse regression methods are powerful to select a promising sparse set of susceptible variants from numerous genetic variants. However, they are ineffective to control the Type I error rates of HDS without functional variants. In this study, we propose a novel scaled sparse regression approach for identifying disease susceptible variants set in high dimensional data. Our method focuses on constructing statistical tests for high dimensional data, based on scaled *Lp* ($0 < p < 1$) norm regularization. Basically, we propose two significance tests: marker wise test (single variant test) and HDS-based test (whole regional test). For marker wise test, we first apply sparse regression with scaled *Lp* norm regularization to generate a de-biased solution. Next, we utilize the asymptotical distribution of the de-biased solution to build the significance test. For HDS-based test, we integrate the marker wise statistics to identify susceptible genetic regions. In this step, the dependence among markers is incorporated to appropriately control set-based Type I error rates while maintaining certain statistical power gain. Under a wide range of simulations, the proposed approaches appropriately controlled Type I error rates and had improved detection power relative to several prominent methods such as SKAT, Global test and HDI test. When applied to a real DNA sequence data of Mexican Americans for studying hypertension, our HDS-based tests identified 8 additional significant susceptibility genes which were not reported by SKAT. Our promising results indicate that the marker wise test can pinpoint single suspected variants with higher resolutions, and the HDS-based test yields considerably high statistical power gain for the whole regional test, especially for high dimensional data. In addition, our methods also maintain substantial power for detecting susceptibility variants in low dimensional data or large samples. Last but not least, our method can detect both rare and common variants with almost equal efficiency.

1367T

Serapis: an archival system for large-scale genetic data. *I.G. Colgiu, N.A. Clarke, M.O. Pollard, J.C. Randall.* Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, United Kingdom.

As the volume of genome sequencing data has rapidly increased over the last few years, the need for high-performance systems that can cope with these storage demands and handle large datasets in an efficient way has become acute. Within the genetics research community, it is desirable to be able to easily access and retrieve large datasets from previous studies of interest. The Serapis archival system has been developed and is currently in use at the Sanger Institute to satisfy these requirements for human genetics research groups. It has been designed for the purpose of archiving large genetic datasets in a long-term storage system and facilitating search of the entire archive based on various criteria. The latter feature has been implemented by attaching relevant metadata to each archived file, stored in the form of key-value pairs. This metadata describes predefined attributes of the data, such as data source (e.g. whole exome/genome sequencing), sample id, associated publication, etc. Though it is possible to use Serapis as a general purpose system, the tool has been tailored to genetics, with built-in support for genetic file formats such as BAM and VCF. It has the ability to automatically collect metadata from within these types of files, by querying external resources, or via manual entry by users. In addition to this, the system also enforces access control restrictions, which is essential in the context of non-public human genetics datasets. In the configuration in use at the Sanger Institute, Serapis uses the iRODS system[1] for backing storage and performs all the work associated with data submission in a distributed way on the Sanger cluster. It is implemented using a distributed task management system (Celery[2]), metadata is assembled temporarily in a NoSQL database before it is permanently archived, and the submission system is accessible via a RESTful web interface. This architecture allows substantial flexibility so that it can be configured to suit the needs of a specific research team or an entire department. The current throughput of our system is 12TB of data archived per day (140MB/s), the limiting factor being the I/O performance of the backing store. All components of the system are available as open source code [3]. [1] http://irods.org/ [2] http://www.celeryproject.org/ [3] https://github.com/wtsi-hgi/serapis-web.

1368S

Leveraging Genome Mapping in Nanochannel Arrays and NGS for a Better Human De Novo Sequence Assembly. *H. Dai[1], A. Pang[1], W. Stedman[1], T. Anantharaman[1], A. Hastie[1], P.Y. Kwok[3], A. Ummat[2], E. Schadt[2], R. Sebra[2], B.A. Bashir[2], H. Cao[1].* 1) BioNano Genomics, San Diego, CA; 2) Mount Sinai School of Medicine, New York, NY; 3) University of California, San Francisco, CA.

Irys genome mapping technology provides direct analysis of extremely long genomic DNA (up to multi-megabases) without amplification. De novo assembly of these single molecules yields high-fidelity contiguous map information across long ranges. Its advantage over all other genome assembly methods is particular dramatic in highly repetitive regions. Genome maps thus greatly complement assemblies using relatively short second- and third-generation sequencing reads. We have constructed genome maps of human NA12878 (cell line derived from the daughter in the CEU trio) which resulted in a consensus assembly measuring 2.9 Gb and with an N50 of 4.6 Mb. With Pacific Biosciences sequence from the same cell line, we also created a sequence-based assembly with N50 length of 930 kb in parallel. By combining data from these two technologies with a custom-designed hybrid scaffolding pipeline, we were able to generate an assembly having scaffold N50 length of greater than 30 Mb covering more than 2.7 Gb of the human genome. At the same time, we were able to identify potential misassembles in the sequence assembly as well as in the genome maps by reviewing the inconsistencies between these two complementary technologies. The hybrid scaffolds also discovered additional long range structural variations not identified in sequence assembly.

1369M

Inference for high-dimensional feature selection in genetic studies. *C. Ekstrøm.* Biostatistics, University of Copenhagen, Copenhagen, Denmark.

Feature selection is a necessary step in many genetic applications because the biotechnological platforms provide a cheap and fast means for producing high-dimensional data. This need for dimension reduction is heightened further for example when data from different omics are combined into simultaneous integrated data analysis or when higher-level interactions among the available predictors are considered (which is the case for gene-gene or gene-environment interactions or in epigenetics).

Penalized regression models such as the Lasso or the elastic net have proved useful for variable selection in many genetic applications - especially for situations with high-dimensional data where the numbers of predictors far exceeds the number of observations. These methods identify and rank variables of importance but do not generally provide any inference of the selected variables. Thus, the variables selected might be the “most important” but need not be significant. We propose a significance test for evaluating the number of significant selection(s) found by the Lasso.

This method rephrazes the null hypothesis and uses a randomization approach which ensures that the error rate is controlled even for small samples. The ability of the algorithm to compute p-values of the expected magnitude is demonstrated with simulated data and the algorithm is applied to two dataset: one on prostate cancer and a full GWAS. The proposed method is found to provide a powerful way to evaluate the set of selections found by penalized regression when the number of predictors are several orders of magnitude larger than the number of observations.

1370T

An integrated framework for sequence variant prioritization. *B. Feng, D. Goldgar.* Department of Dermatology & Huntsman Cancer Institute, University of Utah, Salt Lake City, UT.

Next generation sequencing has played a key role in genetic research and is now becoming integrated into clinical practice. A major problem in utilizing such information is in the interpretation of the large number of variants discovered from such sequencing, which have diverse characteristics and unknown clinical significance. Variant prioritization is the primary step leading to interpretation, based on which variants or genes or pathways can be filtered, tested or classified. Variant prioritization can be done by many criteria, including, but not limited to, sequence conservation, predicted effect on protein, frequency, co-segregation with disease, and gene/pathway relevance. However, it is not clear how to integrate all this information into a robust and quantitative system without arbitrary filtering or scoring. In this study, we propose a Bayesian framework to combine different lines of evidence for variant prioritization. The first part of this framework is a combined deleteriousness score using the weighted naïve Bayes model. By testing on a set of well classified BRCA1 and BRCA2 variants, this score has a higher area under the ROC curve than any component score such as PolyPhen2, SIFT, Mutation Taster, Mutation Assessor, LRT, GERP++, PhyloP, SiPhy, etc., or other combinatory scores such as CADD, MetaLR, and MetaSVM. Adding minor allele frequencies obtained from the NHLBI GO Exome Sequencing Project (ESP) to the model further improved its performance. The second part of the framework is co-segregation assessment using a modified version of that from Thompson et al., where such data is available. The third part, gene relevance, is inferred from guilt-by-association (GBA) in a protein-protein interaction network. The GBA score is computed by naïve Bayes label propagation on an InterConnectedness-transformed weight matrix, and then converted to an empirical Bayes factor. This method showed a high level of prediction accuracy among known disease genes. All these lines of evidences are combined into an overall Bayes factor, which can be used for variant selection and/or as a weight in a gene- or pathway-based burden test for association.

1371S

Omics Pipe: A Computational Framework for Multi-Omics Data Analysis. *K. Fisch, T. Meissner, L. Gioia, T. Carland, S. Loguercio, J.C. Ducom, A. Su.* Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA.

Next generation sequencing (NGS) has presented researchers with the opportunity to collect large amounts of sequencing data, which has accelerated the pace of genomic research with applications to personalized medicine and diagnostics. This has resulted in the development of a large number of computational tools and analysis pipelines, necessitating the creation of best practices and reproducible integrative analysis frameworks. Several automated pipelines have been developed to tie together individual software tools, although many of these tools focus on a single NGS platform (i.e. variant calling, RNA-seq), require computational expertise, and/or are poorly documented. The Nature Protocols journal has attempted to create one solution to establish best practices, by publishing step-by-step directions for well-established NGS analysis pipelines. In addition, other best practice pipelines exist, such as The Genome Analysis Toolkit (GATK). Omics pipe (https://bitbucket.org/sulab/omics_pipe) is an open-source, modular computational platform that automates best practice multi-omics data analysis pipelines. It currently automates and provides summary reports for two RNA-seq analysis pipelines, variant calling from whole exome sequencing, variant calling and copy number variation analysis from whole genome sequencing, two ChIP-seq pipelines and a custom RNA-seq pipeline for personalized cancer genomic medicine reporting. It also provides automated support for interacting with the The Cancer Genome Atlas (TCGA) datasets, including automatic download and processing of all of the samples in this database. We analyzed 100 TCGA breast invasive carcinoma data sets using Omics pipe with UCSC hg19 RefSeq annotation. Omics pipe automatically downloaded and processed the desired TCGA samples on a high throughput compute cluster to produce a results report for each sample. We aggregated the individual sample results to compare the results of our analysis with the original publications, which revealed high overlap between the analyses, as well as novel findings due to the use of updated annotations and methods. In conclusion, Omics pipe enables researchers to analyze next generation sequencing data with little development overhead to provide reproducible, open source and extensible use of established multi-omics analysis methods.

1372M

Tandem repeat sequencing error profiles and error correction models for short read sequencing data. *A. Fungtammasan[1,2], G. Ananda[1,3], S. Hile[4], C. Sun[5], R. Harris[2], P. Medvedev[3,5], K. Eckert[4], K. Makova[2].* 1) Integrative Biosciences, Bioinformatics and Genomics Option, Pennsylvania State University, Pennsylvania, 16802 USA; 2) Department of Biology, Pennsylvania State University, Pennsylvania, 16802 USA; 3) Department of Biochemistry and Molecular Biology, Pennsylvania State University, Pennsylvania, 16802 USA; 4) Department of Pathology, Gittlen Cancer Research Foundation, The Pennsylvania State University College of Medicine; 5) Department of Computer Science and Engineering, Pennsylvania State University, Pennsylvania, 16802 USA.

Profiling Tandem Repeats (TRs) from short read sequencing data is a big challenge because of sequencing and mapping errors. We developed a pipeline that can detect the full spectrum of TR lengths from short read data. We used our pipeline to estimate the TR error rates and patterns for PCR-containing and PCR-free Illumina sequencing libraries using both high-depth genome-wide data and ultra high-depth of plasmid data, which yielded very similar results. We found that the error rates increase with TR repeat number, A/T -richness of TR motives, and PCR step in library preparation. Contraction errors are more common than expansion errors, and the most common type of error is contraction by 1 motif. We used these error patterns to construct an error correction model for genotyping that can assign genotypes correctly for 97-100% of TR loci depending on repeat type. We also used our error models to recommend minimum sequencing depth for genotyping. We found that the minimum sequencing depth required for genotyping increases with TR length. Also, to obtain reliable genotypes, PCR-free data require lower depth than PCR-containing data. Our error correction model has the capacity to distinguish alleles even for TR with consecutive repeat numbers.

1373T

MITOCHONDRIAL DISEASE SEQUENCE DATA RESOURCE (MSeqDR) CONSORTIUM: A Centralized Genomic Resource for Analyzing Genetic Variants of Individuals with Suspected Mitochondrial Disease. *X. Gai[1], L. Shen[1], M.A. Gonzalez[2], M. Attimonelli[3], A. Stassen[4], D. Navarro-Gomez[1], J. Leipzig[5], M. Lott[5,9], D.C. Wallace[5,9], D. Krotoski[6], P.E. Yeske[7], S. Zuchner[2], M.J. Falk[8,9].* 1) Department of Ophthamology, Massachusetts Eye and Ear Infirmary, Boston, MA; 2) Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 3) Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica, UniversitÃ di Bari, Bari, Italy; 4) Department of Clinical Genetics, Maastricht University, The Netherlands; 5) Center for Mitochondrial Medicine and Epigenomic Medicine, The Children's Hospital of Philadelphia , Philadelphia, PA; 6) NICHD, NIH, Bethesda, MD; 7) The United Mitochondrial Disease Foundation, Pittsburgh, PA; 8) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 9) University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.

Mitochondrial diseases are highly heterogeneous with pathogenic variants in many genes from both the nuclear and mitochondrial genomes. While whole exome sequencing (WES) greatly empowers the discovery of novel disease genes and the diagnosis of mitochondrial disease patients, it also magnifies the associated bioinformatics challenges. The later include the highly fragmented and limited nature of the current mitochondrial resources, and also the lack of a convenient and standardized way to share data across the community. Since June 2012, we have established a grass-roots effort supported by the United Mitochondrial Disease Foundation (UMDF) called the "Mitochondrial Disease Sequence Data Resource (MSeqDR) Consortium". A centralized mitochondrial genomic resource has been developed (https://mseqdr.org/) by engaging and working with more than 100 disease experts across the globe. Major components of MSeqDR include: a) A Mitochondrial Disease Locus Specific Database (MSeqDR-LSDB) hosting 3,560 potentially disease-causing genetic variants in 1,332 genes, extracted and fully integrated from multiple resources, such as ClinVar, Ensembl, POLG database, and contributions from various diagnostic laboratories; b) A genome browser (MSeqDR-GBrowse) that allows convenient searching, browsing, and visualization of genomic variant annotations of known and potential mitochondrial disease genes, especially the mitochondrial genome, again extracted from, integrated with, and contributed by all major mtDNA genomics resources, namely MitoMap, HmtDB, and PhyloTree; c) Web-based variant annotation and analysis tools, namely HBCR, MT.AT, and MitoMaster@MSeqDR; d) Phenotype-centric search and browse of genetic variants, fully integrated with MSeqDR-LSDB and MSeqDR-GBrowse; and last but most importantly e) powerful web-based WES analysis of individual patients, families, and cohorts via GEM.app (https://genomics.med.miami.edu/gem-app). The resource is fully functional with more functionalities to be added. Together, we believe this resource will support reliable genetic diagnosis in individual cases, facilitate identification of additional cases of rare genetic disorders, provide evidence to exclude from consideration sequence variants of unknown significance that are unlikely to be disease-causing, permit identification of modifier genetic factors underlying disease variability, and even identify genetic variants that are likely to alter response to emerging therapies.

1374S

Using RNA-Seq to improve sensitivity/specificity of CNV calls made from whole-exome sequencing data. *R. Golhar, J. Glessner, E. Liu, H. Hakonarson.* The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Whole-Exome sequencing is being used to study human disease primarily by cataloguing single-nucleotide variants (SNVs). Copy-number variants (CNVs) are another form of genetic alterations that have been associated with several disease phenotypes and are harder to detect in whole-exome sequencing data due to the discontiguous nature of the target regions (exons) sequenced. Several algorithms have been published to detect rare CNVs in whole-exome sequencing data, including XHMM, Conifer and Canoes. All three methods work by normalizing read depth data, followed by using a hidden Markov Model to detect CNVs. In one dataset consisting of approximately 2000 samples, we were able to positively verify CNVs of at least 5 kb in size called by XHMM in 189 probands while CNVs in 48 probands were negative resulting in an 80% verification rate. In the second dataset consisting of 650 samples, we validated CNVs in 30 samples using qPCR where 15 CNVs (50%) were positively verified, 9 (30%) were negative, and remaining 6 (20%) could not be validated due to the quality of DNA or short length of the exon involved. In order to increase sensitivity and specificity of called CNVs, we are investigating the use of RNA-Seq to provide additional supporting evidence of CNV calls made from whole-exome sequencing data.

1375M

Branch: An interactive, web-based tool for building decision tree classifiers. *B.M. Good, K. Gangavarapu, V. Babji, M. Nanis, A.I. Su.* Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA.

A crucial task in modern biology is the prediction of complex phenotypes, such as breast cancer prognosis, from genome-wide measurements. Machine learning algorithms can sometimes infer predictive patterns, but there is rarely enough data to train and test them effectively and the patterns that they identify are often expressed in forms (e.g. support vector machines, neural networks, random forests composed of 10s of thousands of trees) that are highly difficult to understand. In addition, it is generally unclear how to include prior knowledge in the course of their construction.

Decision trees provide an intuitive visual form that can capture complex interactions between multiple variables. Effective methods exist for inferring decision trees automatically but it has been shown that these techniques can be improved upon via the manual interventions of experts. Here, we introduce Branch, a new Web-based tool for the interactive construction of decision trees from genomic datasets. Branch offers the ability to: (1) upload and share datasets intended for classification tasks, (2) construct decision trees by manually selecting features such as genes for a gene expression dataset, (3) collaboratively edit decision trees, (4) create feature functions that aggregate content from multiple independent features into single decision nodes (e.g. pathways) and (5) evaluate decision tree classifiers in terms of precision and recall. The tool is optimized for genomic use cases through the inclusion of gene and pathway-based search functions.

Branch enables expert biologists to easily engage directly with high-throughput datasets without the need for a team of bioinformaticians. The tree building process allows researchers to rapidly test hypotheses about interactions between biological variables and phenotypes in ways that would otherwise require extensive computational sophistication. In so doing, this tool can both inform biological research and help to produce more accurate, more meaningful classifiers.

A prototype of Branch is available at http://biobranch.org/.

1376T

Computational evaluation of the pathogenicity of noncoding sequence variants in autism spectrum disorder. *A.J. Griswold[1], D. Van Booven[1], N. Dueker[1], E.R. Martin[1,2], M.L. Cuccaro[1,2], J.R. Gilbert[1,2], J.L. Haines[3], J.P. Hussman[4], M.A. Pericak-Vance[1,2].* 1) John P. Hussman Institute for Human Genomics, University of MIami, Miami, FL, USA; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL, USA; 3) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA; 4) Hussman Institute for Autism, Baltimore, MD, USA.

Massively parallel whole exome sequencing, whole genome association, and copy number analyses in thousands of individuals have implicated hundreds of genetic loci, including dozens of genes with rare, *de novo* loss of function and inherited damaging missense variants, in the genetic risk for autism spectrum disorder (ASD). A largely unstudied hypothesis is that rare variants in noncoding regions of the genome functionally contribute to ASD. As part of an ongoing project, we have sequenced evolutionarily conserved regions upstream, downstream, and within introns of 689 ASD associated genes and 942 ASD associated intergenic loci in 2099 ASD cases and 812 controls of European descent. We identified 545,916 single nucleotide variants (SNVs), 519,950 of which are noncoding. To assess which SNVs might contribute to the ASD and to prioritize them for molecular studies, we utilized three bioinformatic tools to annotate and score their functional potential. 1) FunSeq assesses the SNV against the 1000 Genomes database and prioritizes on evolutionary selection and effect of the variant on protein binding motifs. 2) GWAVA (Genome Wide Annotation of Variants) is based on annotation of SNVs with ENCODE and on genome-wide properties such as evolutionary conservation and GC-content. 3) CADD (Combined Annotation Dependent Depletion) integrates multiple functional categories and contrasts variants that survived natural selection with simulated mutations and scores them relative to all possible mutations in the genome. 7,223 variants were in the top 10% of both GWAVA and CADD. Of those, 137 also had a score greater than 3 in FunSeq, the top 2% of FunSeq scores, making these our most likely functional noncoding SNVs. Among these are case specific rare variants (MAF<0.01) in highly conserved transcription factor binding sites upstream of ASD candidate genes including five upstream of *NRXN1* and two upstream of *CTNND1*. In addition, promoters of other neurodevelopmental genes *FGFR2*, *SGK1*, and *SETD7* contain case unique high scoring SNVs. Molecular experiments to determine the underlying effect of these variants on protein binding and gene regulation are underway. This evolving approach of applying computational methods to noncoding variation will enhance our ability to classify SNVs as functional and broaden our understanding of the underlying genetic factors in ASD. This approach can be applied to other disorders generating large scale beyond exome sequencing data.

1377S

Similarity metrics for comparing exome sequence variants. *V. Heinrich[1], P.N. Robinson[1,2,3,4], S. Mundlos[1,3], P.M. Krawitz[1,2,3].* 1) Institute for Medical Genetics and Human Genetics, Charité Universitätsmedizin Berlin, Germany; 2) Berlin Center for Regenerative Therapies (BCRT), Charité-Universitätsmedizin Berlin,; 3) Max Planck Institute for Molecular Genetics Berlin, Germany; 4) Institute for Bioinformatics, Department of Mathematics and Computer Science, Freie Universität Berlin, Germany.

Removing frequently detected variants is one of the most effective approaches to reduce the number of candidate mutations in the data analysis of next-generation sequencing studies. The incidence of a rare disorder in a population serves as an upper bound for the allele fre- quency or genotype frequency that can be used as a filter for dominant or recessive disorders. However, the frequentist inference requires that genotypes of a single individual are repre- sented in the database only once. With many and decentralized data submitters the probability increases that samples of the same individual are sequenced multiple times and are contributed independently under differ- ent pseudonyms. We compare two different metrics that compute the similarity between any two exome samples in a defined target region. Each similarity value can be used to assess whether a list of sequence variants has already been submitted and provides information about the relationship between samples. This allows the identification of replicates from different enrichment procedures, sequencing platforms and bioinformatic pipelines.

1378M

Novel Integrative Genomics Approach for the Discovery of MicroRNA and mRNA Signatures and target Pathways in Prostate Cancer. *C. Hicks[1,2,3], R. Jangiti[1], R. Ramani[1], T. Koganti[1], S. Giri[3], S. Vijayakumar[1,3].* 1) Cancer Institute, University of Mississippi Medical Center, Jackson, MS; 2) Department of Medicine, University of Mississippi Medical Center, Jackson Mississippi; 3) Department of Radiation Oncology, University of Mississippi Medical Center, Jackson Mississippi.

Background: Prostate cancer is the most common solid tumor and the second most common cause of cancer death in men in the United States. Advances in high-throughput transcription profiling have made possible molecular classification of prostate cancer. However, although these primary analyses have provided valuable insights about the molecular basis of prostate cancer, they have been unsuccessful in determining which genes have causative roles as opposed to being consequences of the prostate cancer state. Advances in genotyping and reduction in genotyping costs made possible identification of single nucleotide polymorphisms (SNPs-herein called genetic variants) associated with an increased risk of developing prostate cancer using genome-wide association studies (GWAS). However, many of the identified genetic variants (>80%) map to noncoding regions with unknown functions. Recently, microRNAs (miRNAs) a class of noncoding RNAs have gained prominence in oncology research as potential clinically actionable biomarkers. However, the functional relationships between genetic variants, miRNAs and their mRNA and pathway targets are not well defined. The objective of this investigation was to leverage GWAS information with miRNA and mRNA expression data for the discovery of MicroRNA and mRNA oncogenic signatures and pathways in prostate cancer. Methods: We used a combination of bioinformatics tools to integrated disparate genomics data in our analysis. First, we developed a comprehensive catalogue of genetic variants and associated genes from GWAS and a comprehensive catalogue of miRNAs and their mRNA targets. Next, we analyzed miRNA and mRNAs expression data to identify miRNA and mRNA signatures enriched for genetic variants. Subsequently we performed network and pathways analysis to identify target molecular networks and biological pathways enriched for genetic variants. Results: The analysis revealed miRNAs and mRNA oncogenic signatures. We identified molecular networks and pathways enriched for genetic variants and targeted by miRNAs. Among the identified target pathways included the androgen, androgen biosynthesis, IGF-1, JAK, STAT and prostate cancer signaling pathways. Conclusions: Our investigation demonstrates that integrative genomic analysis provides a powerful and unified approach for establishing putative functional bridges between GWAS findings, miRNAs and their mRNA and pathway targets.

1379T

Medical re-sequencing analysis pipeline provides one-stop solution for identifying disease-causing mutations of Mendelian disorders. *H. Hu[1], T.F. Wienker[1], L. Musante[1], V. Kalscheuer[1], P.N. Robinson[2], H.H. Ropers[1].* 1) Max Planck Institute for Molecular Genetics, Berlin, Germany; 2) Universitätsklinikum Charité, Berlin, Germany.

Next-generation sequencing has greatly accelerated the search for disease-causing defects, but even for experts the data analysis can be a major challenge. To facilitate the data processing in a clinical setting for Mendelian disorders, we have developed a novel Medical Re-sequencing Analysis Pipeline (MERAP). MERAP assesses the quality of sequencing, and has optimized capacity for calling variants, including Single Nucleotide Variants, insertions and deletions, Copy Number Variation, and other structural variants. MERAP identifies polymorphic and known causal variants by filtering against public-domain databases, and flags non-synonymous and splice-site changes. MERAP uses a logistic model to estimate the causal likelihood of a given missense variant. MERAP considers the relevant information such as phenotype and interaction with known disease-causing genes. MERAP compares favorably with GATK, one of the widely used tools, because of its higher sensitivity for detecting indels, its easy installation, and its economical use of computational resources. Upon testing more than 1,800 individuals with mutations in known and novel disease genes, MERAP proved highly reliable, as illustrated here for 5 families with disease-causing variants. We believe that the clinical implementation of MERAP will expedite the diagnostic process of many disease-causing defects for Mendelian disorders.

1380S

A novel integrated analysis framework for detecting genome-wide changes in gene expression or regulation with next-generation sequencing data. *W. Huang, D.M. Umbach, L. Li.* Biostatistics, NIEHS, NIH, Research Triangle Park, NC.

The cost-effective next-generation sequencing has quickly become a popular assay method for studying gene expression and genetic/epi-genetic regulation. The development of corresponding data analysis tools, as demanded by these sequencing applications, has also been fast paced, particularly in the last five years. The sequencing technologies continue to evolve so quickly, however, that the development of analysis tools still lags far behind the need. In particular, sequencing is now being used for more complicated or more challenging studies, and existing tools are no longer sufficient for these new and more advanced applications. In fact, data analysis and interpretation is becoming the bottleneck that limits more widespread sequencing applications. The main inadequacies of existing tools include: 1) ineffective at detecting true biological changes, especially when number of biological replicates is small, 2) high false positive rates, typically resulting from unrealistic underlying statistical assumptions, 3) little capability for analyzing data from more advanced experimental designs, 4) limited data visualization capability, and 5) typically restricted to one type of sequencing application. We present a new computational framework, based on our previous work (EpiCenter), to address the above challenges. Instead of using a single procedure or statistical test for every application, our new framework employs different statistical tests and filters to analyze data from different kinds of study. Specifically, our framework makes the choice according to the experimental design, number of biological replicates, type of data (e.g., histone ChIP-seq or mRNA-seq gene expression), and the organism/species (genome complexity and annotation can differ dramatically among organisms) used in the study. In addition, our framework, unlike existing methods that treat individual genes/regions separately, considers individual genes/regions as an integrated part of a genome-wide study to improve overall detection accuracy. Furthermore, our framework extends EpiCenter's data visualization capability including the generation of the UCSC hub track data. Evaluation on both simulated and actual sequencing data showed the superior performance of our framework over that of some popular existing tools.

1381M

Database of disease-associated genomic polymorphisms based on assessment of reproducibility between or within human populations. *T. Imanishi[1,2], Y. Nagai[1], Y. Takahashi[1].* 1) Tokai University School of Medicine, Isehara, Kanagawa, Japan; 2) National Institute of Advanced Industrial Science and Technology, Tokyo, Japan.

Genome-wide association studies (GWASs) have identified numerous single nucleotide polymorphisms (SNPs) that are associated with development of common diseases. However, because GWAS uses statistical evaluation, we cannot completely eliminate false positives that may contaminate to a certain extent. On the other hand, it is becoming clearer that genetic risk factors of common diseases are not totally universal but heterogeneous among human populations. We thus developed a new database of genomic polymorphisms that are *reproducibly* associated with disease susceptibilities, drug responses and other traits for each human population, and released it as "VarySysDB Disease Edition (VaDE)". Using PubMed and NHGRI GWAS catalog, we collected 1806 GWAS papers and curated them manually. We extracted information of associated SNPs, odds ratios, p-values, study design, nationality of subjects, and many others. Also, extensive manual curation has been carried out separately for hypertension and rheumatoid arthritis. Then, we assessed the reproducibility of each association in multiple, independent studies for each human population. Finally, we could obtain 4675 and 790 reproducible associations for 427 and 151 traits in the European and East Asian populations, respectively. Furthermore, to support finding functional SNPs in the VaDE database, we integrated data of ChIP-seq, DNaseI hypersensitivity experiments, regulatory motifs, RefSeq genes, H-Inv transcripts, and linkage disequilibrium data in three major human populations that have been obtained from Haploreg v2, VarySysDB, and Univ Michigan. We also installed a genome browser to visualize these data in VaDE. From the VaDE database, we searched for pleiotropic SNPs that are associated with multiple traits. As a result, we could identify 802 pleiotropic SNPs, such as those affecting metabolic traits and autoimmune diseases, which provides a clue to solve hidden relationships among various phenotypes. The VaDE database is publicly available from http://bmi-tokai.jp/VaDE/. We believe that our database will contribute to the future establishment of personalized medicine and understanding of genetic factors underlying diseases.

1382T

Semantic Similarity Analysis of Patient Phenotypes for Genome Wide Genetic Diagnostics. *R. James, M. Bainbridge, C. Eng, C. Shaw.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

The clinical implementation of genome wide diagnostics is increasingly important in the molecular diagnosis of human disease. However, these genome wide approaches are challenged by the large number of variants observed in individual patients. We hypothesize that efficient integration of available clinical data characterizing patient indications for genetic testing can improve the implementation of diagnostic workflows and, ultimately, prioritization of variants. We have developed scalable quantitative analysis tools for mining indication content. Fundamentally, our approach exploits semantic similarity computed on patient phenotypes to comprehensibly represent patient features in simple composite scores. Our tool can help to elucidate cohort substructure by segmenting large, heterogeneous groups of subjects into subgroups based on shared multidimensional phenotypes. Identification of this structure can help to identify and establish cohorts for associative or experimental study, accelerating the investigation of genetic variation. Identification of cohorts can also lead to discovery of novel functional variants that drive complex phenotypes. We have also developed a prototype graphical interactive data viewer exploiting our computational toolkit. This viewer can aid decision support through visual query and analysis of structured input data. This tool can help simplify identification of clinically relevant variants by facilitating indirect matching of patient phenotypes to cataloged variants already known to be associated with semantically similar patients or reported OMIM diseases. We use our tools, in the context of the Human Phenotype Ontology representation of patient indications, to analyze a pilot cohort assembled from retrospective data from the Whole Genome Laboratory at Baylor College of Medicine. We demonstrate the performance of our method for matching to known diseases by comparing our computational OMIM disease matches for subjects against the findings reported in exome reports. These tools can help improve the quality of clinical care by more effectively utilizing available indication data, increasing the value of the electronic health record (EHR), and better illuminating the correspondence between the requisition content and the ultimate genetic diagnosis.

1383S

Efficient and accurate multiple phenotypes regression method for high dimensional genomic data considering the population structure. *J.W. JOO[1], E.Y. Kang[2], E. Eskin[1,2,3].* 1) Bioinformatics IDP, UCLA, Los Angeles, CA; 2) Computer Science Dept, UCLA, Los Angeles, CA; 3) Human Genetics Dept, UCLA , Los Angeles, CA.

Typical GWAS test correlation between a single phenotype and each genotype one at a time. However, it is often very useful to analyze multiple phenotypes simultaneously. Especially, high-dimensional multiple phenotypes analysis is preferable with the advent of high-throughput technologies that produce a huge amount of data everyday. For example, researchers may want to detect variants associated with many phenotypes such as a profile of gut microbiota which contains tens of thousands species. Another useful application of high-dimensional multiple phenotype analysis is for detecting regulatory hotspots in eQTL studies. Many genes are known to be regulated by a small number of genomic regions called regulatory hotspots and those are very important evidence of the presence of master regulators of transcription. In addition, due to clinical overlap and correlation between many phenotypes, high-dimensional multiple phenotype analysis may boost the power to detect variants. There are several multivariate approaches that try to detect variants related to multiple phenotypes. However, none of them consider population structure and may result false positive identifications. Here, we introduce a new methodology referred to as GAMMA that could both simultaneously analyzes many phenotypes as well as corrects for population structure. In simulated study, GAMMA accurately identifies true genetic effects without false positive identifications. While other methods either fail to detect the true effects or result many false positive identifications. We further apply our method to a real Hybrid Mouse Diversity Panel data and show that GAMMA identifies several variants that are likely to have a true biological mechanism.

1384M

Using a Pan-Genome Reference for Sequence Alignment and Accurate Haplotype Discovery. *D. Kural.* Research & Development, Seven Bridges Genomics, Cambridge, MA., Select a Country.

Currently, the human reference genome is the anchor point against which new variants are discovered, and annotations are made. This approach relies on accurate matching placement of short reads generated by high throughput sequencing to the reference genome, leading to the development of aligners and variant detection (including SV/CNV) algorithms suited for these new high throughput technologies with an emphasis on speed. We develop an approach capable of aligning against a population genome, removing many of the biases traditionally associated with re-sequencing that rely on a singular reference genome. We accomplish this by removing the restrictions imposed by a linear reference sequence, and by incorporating known variants. A graph genome (pan-genome) representing known variants enables the accurate alignment of short sequences containing these regions - thus boosting discovery rates of known variants in new samples, compared to existing methods, most significantly in regions containing Structural Variants. Additionally, we've developed an SV genotyper that can operate directly on graph genomes. We present results demonstrating accurate genotyping of Structural Variants, that were previously undiscovered, deriving from direct alignments to a reference graph containing both hg19 and HuRef, based on Craig Venter. Our approach enables correct alignment of reads to complex regions, leading to a more accurate determination of the personal genome queried against an enhanced graph genome representing a reference population, compared to traditional methods. We believe that the reference genome will keep serving a useful purpose in providing a common coordinate system. This approach also enables more accurate analysis of tumor/normal pairs, by incorporating the normal genome into the reference graph. This enables determining and genotyping the clonal complexity around translocations and fusion genes.

1385T

Predicting functional regulatory variants from DNA sequence. *D. Lee[1], M.A. Beer[1,2], A. Chakravarti[1,3].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Center for Complex Disease Genomics, Johns Hopkins University School of Medicine, Baltimore, MD.

Genome-wide association studies (GWAS), having identified > 12,000 single nucleotide polymorphisms (SNPs) significantly associated with > 900 phenotypic traits and diseases, have highlighted the challenge of identifying specific causal variants underlying these associations. ~90% of such SNPs are intergenic or intronic, and, thus, likely regulatory, so that integration of epigenetic data can reduce the number of candidate variants. However, the presence of linkage disequilibrium (LD) and the fact that not all variants in regulatory elements are functional does not lead to unequivocal discovery of causal sites.

We propose a novel framework, using sequence-based predictive models and epigenetic data (Lee et al. 2011, Genome Res), to address this challenge. The method involves: (1) building a support vector machine (SVM) classifier based on publically available ChIP-seq and DNaseI-seq data; (2) predicting functional regulatory SNPs based on changes in the SVM score (delta-SVM) between reference and alternative alleles; (3) validation. A major advantage of our method is that it provides specific testable hypotheses of the underlying molecular mechanism (binding site, transcription factor, transcriptional target). As proof of principle, we analyzed GM12878 lymphoblastoid dsQTLs (DNaseI Hypersensitivity Quantitative Trait Loci, Degner et al. 2012, Nature) using delta-SVM trained on the DNaseI-seq data set from the same cell line. We show that: (1) our SVM classifier can accurately predict GM12878 DNaseI hypersensitive sites (area under curve=0.94); (2) ~60% of dsQTLs (~3,700 out of 6,070) are predicted to have at least one significant SNP determined by delta-SVM ($P < 0.05$); (3) ~34% of these SNPs (~1,250 out of ~3,700) are in strong LD ($r^2 > 0.5$) with their sentinel SNPs (i.e. SNP with the most significant association with DNaseI hypersensitivity); (4) these SNPs are mostly predicted to disrupt or to create transcription factor binding sites (TFBSs) known to play roles in the relevant cell type (i.e. NFKB1, RUNX1 and IRFs), as well as general and promoter specific TFBSs (i.e. AP-1, CREB, YY1, NFYA, SP1 and ETSs). Combined with association fine-mapping, our method can greatly help in the identification of causal and functional regulatory variants.

1386S

DISTMIX: Direct imputation of summary statistics for unmeasured SNPs from mixed ancestry population. *D. Lee, T. Bigdeli, B. Riley, A. Fanous, S. Bacanu.* Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA.

Background: To improve signal detection power, many meta-analyses of genome-wide association studies use mixed-ancestry cohorts. To analyze these diverse cohorts while also increasing imputation accuracy, large multiethic reference panels are used. However, the ever increasing sizes of reference panels and study cohorts makes genotype imputation computationally unfeasible for moderately sized computer clusters. Moreover, due to modelling assumptions, the commonly used genotype imputation methods lack the ability to impute genotypes for family data. Besides, genotype imputation requires subject level genetic data, which unlike summary statistics provided by virtually all studies, are not publicly available.

Method: To overcome these limitations, we propose a novel method/software for Directly Imputing summary STatistics for unmeasured SNPs from MIXed ancestry population (DISTMIX) without imputing first subject level genotypes. Based only on the correlation matrix of allele frequencies from the study and the ethnicities in the reference panel, DISTMIX first infers the proportions of reference panel ethnicities in the study cohort. These proportions are subsequently used as weights in calculating the ancestry-weighted genotype correlation matrix as the mixture of ethnically-specific correlation matrices, as estimated from the mixed ethnicity reference panel. This mixture correlation matrix is subsequently used by DISTMIX to impute summary statistics of unmeasured SNPs using the conditional expectation formula for multivariate normal variates.

Results: Experiments based on our simulated null data sets with different types of ancestry shows that the proposed method controls the false positive rates. 1000 Genomes based imputation of summary statistics from the ethnically diverse Psychiatric Genetic Consortium Schzophrenia Phase 2 (PGC SCZ2) suggests that, when compared to commonly used genotype imputation methods, our method offers comparable imputation accuracy while requiring only a fraction of computational resources. We also present some new and interesting genetic findings for PGC SCZ2 cohort, which were obtained by increasing the size of the reference panel (and, likely, imputation accuracy) through the combination of the 1000 Genomes and UK10K reference panels.

1387M

Detecting complex fusion transcripts in pediatric cancer using a novel assembly-based algorithm CICERO. *Y. LI[1], T. Bo[2], M. Rusch[1], J. Easton[3], K. Boggs[3], B. Vadodaria[3], P. Gupta[1], G. Song[2], J. Ma[2], C.G. Mullighan[2], S.J. Baker[4], R.J. Gilberton[4], J.R. Downing[2], D.W. Ellison[2], J. Zhang[1].* 1) Department of Computational Biology, St Jude Children's Research Hospital, Memphis, TN; 2) 2Department of Pathology, St Jude Children's Research Hospital, Memphis, TN; 3) 3The Pediatric Cancer Genome Project Validation Laboratory, St Jude Children's Research Hospital, Memphis, TN; 4) 4Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, TN.

Fusion genes are important for cancer diagnosis, subtype definition and targeted therapy. Although RNAseq is useful for detecting fusion transcripts, computational methods to identify fusion transcripts arising from internal tandem duplication (ITD), that have multiple partners, low expression or non-template insertion are limited. We developed an assembly-based algorithm CICERO (CICERO Is Clipped-reads Extended for RNA Optimization) that is able to extend the read-length spanning fusion junctions for detecting complex fusions. Using test data that include RNASeq from 3 ependymoma (EPD), 39 low-grade glioma (LGG), and 128 acute lymphoblastic leukemia (ALL), we have shown that CICERO is able to detect multi-segment fusion transcripts resulting from chromothripsis, internal tandem duplication or rearrangement at a highly repetitive immunoglobulin (IG) locus; all of which would be missed by existing fusion analysis methods. The overall sensitivity and accuracy of CICERO are much higher compared with existing tools such as deFuse and Tophat-Fusion. Using CICERO, we analyzed >600 brain tumor and leukemia transcriptomes from the St. Jude/Washington University Pediatric Cancer Genome Project (PCGP) and detected recurrent C11orf95-RELA fusions in EPD, FGFR1 ITD in LGG, NTRK fusion in high-grade glioma and activating kinase fusions with multiple partners in ALL. CICERO also shows high sensitivity when detecting fusions with low expression, like BRAF fusions in LGG, making it useful for identifying subclonal lesions and for analyzing tumor specimens with low purity. Furthermore, the power of CICERO increases with the extended read-length enabled by improvement in next-generation sequencing (NGS) technology. Using paired-end 300bp RNAseq reads, CICERO shows the ability to assemble near full-length fusion transcripts and identify complex fusions with multiple segments.

1388T

Mixture modeling of next generation sequencing data and its application to estimating genotype frequency. *J. Lihm[1], S. Yoon[1], S.J. Finch[2].* 1) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; 2) Stony Brook University, Stony Brook, NY.

Studying a probability of an individual having a variant, inherited or not, is of interest in genetic studies. We present a method for modeling the frequency of single nucleotide polymorphism variants in the exome capturing sequence data of an individual. Mixture modeling is used to model the proportion of alternative allele at a position under the biallelic single nucleotide polymorphism model. We suggested the application of this mixture modeling in this paper. We measured the proportion of alternative allele for positions in chromosome 1 exome sequencing data of two trios from pilot 3 study in the 1000 genomes project. The measurement was based on the counts of reference and alternative allele in the pileup file from SAMtools. We fit the proportions to mixture model of two point distributions and five continuous distributions. The fitted mixture model well described the properties of the distribution of the alternative allele proportions. The estimates of mixing proportions were used to estimate the fraction of each genotype in the data. Each individual had different estimates of parameters, but the estimates of genotype fractions were similar. We found the he estimated fractions of each trio clustered. We developed an expectation-maximization algorithm to obtain the maximum likelihood estimates of the mixture model parameters. Its application to estimate the fraction of each genotype was presented in this paper.

1389S

A high-performance database framework for fast and easy prioritization of disease related variants from Exome Sequencing data. *B. Linghu, F. Yang, R. Bruccoleri, J. Szustakowski.* Novartis Institutes for BioMedical Research, Cambridge, MA.

Exome sequencing (Exome Seq) has become a promising approach to identify disease related genetic variations. Pinpointing the small subset of pathogenic mutations amongst the thousands or millions of variants generated in an Exome Seq experiment remains a conceptual and computational challenge. One common approach is to use relational database systems to conveniently organize and query variant data for prioritization. However, traditional database systems often perform poorly when applied to such "Big Data". Recently, a number of high-performance database systems have been developed specifically to enable analysis of extremely large data sets. Here we describe applying one such system, namely Vertica, to prioritize disease variants. Our approach leverages Vertica's high performance capabilities to efficiently model, store, and query a comprehensive landscape of information including variant calls, variant quality metrics, predicted functional consequences, allele frequencies, disease prior knowledge, inheritance patterns, and clinical phenotypes. This framework enabled the convenient and efficient identification of candidate disease variants, with significant improvements over traditional databases. To our knowledge, this is the first demonstration that high-performance databases such as Vertica provide an efficient solution to prioritize variants from exome sequencing.

1390M

Analysis of Human neurodevelopmental disorders from the systems biology perspective using the Lynx Platform. *N. Maltsev[1], D. Sulakhe[2], D. Boernigen[1], B. Xie[1], A. Taylor[1], A. Paciorkowski[3], W. Dobyns[4], T.C. Gilliam[1].* 1) Human Genetics Department, University of Chicago, Chicago, IL., USA; 2) Computation Institute, University of Chicago, Chicago, IL, USA; 3) Departments of Neurology, Pediatrics, and Biomedical Genetics, University of Rochester Medical Center, Rochester, NY, USA; 4) Center for Integrative Brain Research, Seattle Children's Research Institute and Department of Pediatrics, University of Washington, Seattle WA, USA.

Understanding of genetic mechanisms underlying common heritable disorders (e.g. autism, schizophrenia, brain malformations) is one of the most important challenges of biology. Recent advances in biology provided new perspectives onto studies of complex heritable disorders, including: networks-based view of Human disorders; emergence of "phenomics" and a notion of interrelatedness of disease traits. However, the progress of biomedicine facilitated by these methodologies depends on availability of the new bioinformatics platforms supporting the development of integrative models of genetic disorders and identification of their characteristic multidimensional patterns of inheritance. We present an approach and a supporting computational platform Lynx (http://lynx.ci.uchicago.edu/) for analysis of common heritable disorders from the systems biology perspective. Our approach is three-fold and includes tools for the enrichment analysis, gene prioritization and the development of the networks-based models of biological processes in health and disease. These tools utilize genomic, functional, clinical and experimental information from the Lynx knowledge base integrating over 35 public and proprietary databases and information derived from the literature using advanced text mining. We will illustrate our approach using analysis of brain connectivity disorders (e.g. agenesis of corpus callosum, autism, schizophrenia) as an example. While interest in the biological mechanisms that lead to heritable brain disorders such as autism, infantile epilepsy and brain malformations has been high, discovery of the underlying genetic mechanisms has proven elusive despite some recent advances. Previous studies of developmental brain disorders has led to understanding that "interrelatedness" between molecular components postulated by the systems approach will be essential for understanding both individual susceptibilities and possible biological interventions. The presented work is based on comparative analysis of a group of developmental brain disorders that often co-occur and share at least a subset of causative genes (e.g. autism, agenesis of corpus callosum, mid-hindbrain malformations). Our analysis allowed uncovering some of the common molecular mechanisms that underlie these disorders. This knowledge will eventually lead to the development of efficient diagnostic and therapeutic strategies.

1391T

Identification of differentially expressed genes and somatic mutations in esophageal adenocarinoma cancer patients. *M. Matvienko[1], B. Øster[2], A. Joecker[2], A.-M. Hein[2], P. Dekker[2], R. O'Neill[3], A. Krejci[4], A. Arens[2], N. Thomson[1], C. Boysen[1], S. Mønsted[2], R. Forsberg[2], B. Knudsen[2], T. Knudsen[2], T. Hupp[3].* 1) CLC bio, a Qiagen Company, Davis, CA; 2) CLC bio, a QIAGEN Company, Silkeborgvej 2, Aarhus, 8000, Denmark; 3) 2p53 Signal Transduction Group, University of Edinburgh Cancer Research Centre in the Institute of Genetics and Molecular Medicine, Crewe Road South, Edinburgh EH4 2XR, United Kingdom; 4) Regional Centre for Applied Molecular Oncology, Masaryk Memorial Cancer Institute, þlutý kopec 7, 656 53 Brno, Czech Republic.

High throughput sequencing technologies are currently revolutionizing the cancer research area with rapid improvements in sequencing capacity and time consumption. As a result the most time consuming step has moved from being the sequencing process itself to being the bioinformatic data analysis. RNA sequencing (RNA-Seq) is used in an increasing number of transcriptomic studies. The great advantage of using RNA-Seq is its ability to precisely quantify transcript levels and identify novel transcripts, isoforms, and splice junctions, while further providing information of the mutational landscape down to single base resolution. To ease the hurdles associated with RNA-Seq data analysis there is an increasing demand for tools that are specifically tailored to RNA-Seq data. Here we describe how the newly developed CLC Cancer Research Workbench can be used to analyze and visualize RNA-Seq data with ready-to-use workflows that automatically map, quantify, and annotate transcriptomes. We identify differentially expressed genes and transcripts in Illumina HiSeq transcriptomic data from matched tumor and normal samples from four esophageal adenocarinoma cancer patients, and compare the mutational patterns in the samples with the expression values of the corresponding genes. Results are visualized in a track based genome browser view, which provides the means for quick and easy navigation as well as allowing the user to simultaneously view and annotate multiple samples and different data types (e.g. genes, transcript expression levels, and detected variants).

1392S

A Simple Method of Generating Reproducible NGS Workflows. *M. Mikheev, A. Taffel, G. Ganebnii, D. Leca, A. Lisnik, M. Hohlovich, A. Dobretsov, V. Shakin, M. Gollery.* Biodatomics, Bethesda, MD.

A core tenet of biological study is that research by one scientist or organization must be reproducible -- and thereby verified -- by third parties. The NGS analyses that are now dominating genomic research commonly employ open-source, command line algorithms. While these algorithms are powerful, flexible and readily accessible, they have a drawback when it comes to creating research pipelines that can be identically reproduced and verified by other researchers. Specifically, each algorithm has its own options, variables and formats that govern the manner in which it is executed. Furthermore, the algorithm's executable code itself may go through multiple variations and versions. When these algorithms are then combined into a pipeline, the same variability applies exponentially. Unless all of these factors can be held constant across datasets, locations, and users, research will not be reproducible and verifiable by third parties. Several groups have developed software systems with Graphical User Interfaces designed to ease the manipulation of open source algorithms; however, none embodies a mechanism to enable reproducible research. BioDatomics set out to design and create a mechanism to facilitate reproducible research for its BioDT analytics platform. The solution turned out to be surprisingly simple. Once the original researcher has "locked in" an algorithm or workflow, they can assign it a version number. The version captures all relevant metadata about the algorithm or set of algorithms. When the algorithm or workflow is shared with colleagues, the originator has the option of forcing downstream executions to run as described in the version number. In this way, research can be recreated without regard for who is running the analysis, where it is being run, or even the passage of time. This poster will explain the need for reproducible research, the inherent obstacles within open source algorithms, and will present the simple solution to the problem.

1393M

Clinical whole-exome and whole-genome sequencing in dystonia: a key role for UMD knowledgebases. *M. Miltgen[1,2], A. Blanchard[1,2], L. Barré[1,2], C. Palacin[1,2], C. Guien[1,2], A. Pinard[1,2], A. Roubertie[4,5], C. Béroud[1,2,3], G. Collod-Béroud[1,2].* 1) Aix-Marseille Université, GMGF, 13385, Marseille, France; 2) Inserm, UMR_S910, 13385, Marseille, France; 3) APHM, Hôpital Timone Enfants, Laboratoire de Génétique Moléculaire, 13385, Marseille, France; 4) Département de Neurologie Pédiatrique, Hôpital Gui de Chauliac, 34295, Montpellier, France; 5) Inserm U1051, Institut des Neurosciences de Montpellier, 34091, Montpellier, France.

As the cost of genome sequencing falls, the clinical use of whole exome sequencing (WES) in first line of molecular testing becomes widespread. These approaches can facilitate the work-up of disorders like dystonia that present large clinical variability and clinical overlapping for the different involved genes. WES is also the technique of choice for clinical practice as newly identified genes can be easily added to the bioinformatics pipeline. Nevertheless, because of the huge amount of data generated by such technology, access to reference knowledge base system is central for accurate data analysis and interpretation. To date, several kinds of primary dystonia have been characterized: "pure", "plus", "paroxystic" and "heredodegenerative" covering 23 forms of these diseases and 17 genes have been identified. Our team has developed the UMD (Universal Mutation Database) reference system allowing not only creating locus specific databases, but also analyzing and interpreting data through various algorithms like a pathogenicity of missense and synonymous mutations prediction system (UMD-Predictor®) or algorithms to evaluate the impact of mutations on splicing signals. In order to exhaustively collect, validate and annotate molecular and clinical data for patients presenting with dystonia we have created locus specific databases for genes currently known for being involved in these diseases: *ANO3, ATP1A3, CIZ1, GCDH, GCH1, GNAL, MR1, PRKRA, PRRT2, SGCE, SLC2A1, SPR, TH, THAP1,* and *TUBB4a.* Due to a founder effect for *TAF1* and the lack of genetic heterogeneity for *TOR1A* (four mutations reported to date), they were not included in this project at present. Most of the 15 genes show a wide clinical spectrum that is not restricted to dystonia but is also associated with several other diseases for which data are also included. These databases contain mutations collected from literature and through direct collaborations with diagnostic laboratories representing 1,719 entries to date. The main goals of this work are to provide an exhaustive collection of clinical and molecular data associated with dystonia and to facilitate WES data analysis in the context of dystonia. We believe that these resources could play a pivotal role in such heterogeneous diseases and decrease the diagnostic wavering by facilitating WES integration in clinical practice.

1394T

A high-fidelity simulation validation framework for high-throughput genome sequencing with cancer applications. *J.C. Mu[1,2], M. Mohiyuddin[2], J. Li[2], N. Bani Asadi[2], M.B. Gerstein[3], A. Abyzov[4], W.H. Wong[5,6], H.Y.K. Lam[2].* 1) Department of Electrical Engineering, Stanford University, Stanford, CA; 2) Department of Bioinformatics, Bina Technologies, Redwood City, CA; 3) Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT; 4) Department of Health Sciences Research, Mayo Clinic, MN; 5) Department of Statistics, Stanford University, Stanford, CA; 6) Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA.

High precision simulation and computational validation are essential in characterizing and assessing the accuracy of NGS analysis. We present VarSim, an integrated computational framework that leverages state-of-the-art read simulation and vast annotation databases to generate realistic high-throughput sequencing reads for normal and tumor samples and report detailed accuracy statistics. It also has the ability to isolate the analysis to specific regions of interest, such as disease genes.

For each normal or disease sample, VarSim first generates a phased diploid genome using variants, including various types of structural variations, from existing annotations and novel sites. Next, reads are simulated from this diploid genome using empirical error models. After alignment and variant-calling on the simulated reads, VarSim reports detailed statistics on the accuracy of the results in an HTML document. These statistics include alignment accuracy and variant-calling accuracy for different variant types and sizes, as well as for different categories of genomic regions, e.g., genes and repeats. Since VarSim generates a diploid genome, genotyping accuracy is also reported. For tumor samples, VarSim also uses COSMIC annotations to generate the tumor variant set. To simulate the impurity and heterogeneity of tumor samples, the reads for a tumor sample are constructed by mixing the reads generated from multiple sets of tumor variants with the reads generated from normal variants in a specified proportion. After running somatic mutation callers on the reads, VarSim then generates a report on the accuracy of the somatic variant calls.

To demonstrate its utility and enable rapid validations, we generated three synthetic genomes and their reads at high coverage (100x): one a normal female sample, another a male sample with a rare disease and finally a tumor-normal sample pair. We compared the accuracy statistics generated by VarSim for these genomes on popular secondary analysis tools including aligners, small variant callers, structural variant (SV) callers and somatic mutation callers. One of these tools includes our multi-algorithm SV caller, MetaSV, which improves both sensitivity and specificity of SV calling. No other simulation framework offers such a comprehensive validation of secondary analysis over a variety of whole genome sample types.

1395S

Accurate estimation of transcript isoform expression from RNA-Seq data by improved variational Bayesian inference. *N. Nariai, K. Kojima, T. Mimori, Y. Sato, Y. Kawai, Y. Yamaguchi-Kabata, M. Nagasaki.* Tohoku University, Sendai, Miyagi, Japan.

RNA sequencing (RNA-Seq) technologies have been widely used for quantification and identification of transcripts at single-base resolution. However, in many cases, short reads can be aligned to more than one location of reference sequences, which complicates quantification of transcript isoforms from RNA-Seq data. Previously, graphical models have been proposed to handle isoform abundances as parameters, and unknown alignment states as hidden variables. By using the expectation maximization (EM) algorithm, maximum likelihood estimate can be obtained through an iterative update of parameters. However, the EM algorithm poses a risk of overfitting because of noise, such as sequencing errors, introduced in RNA-Seq data. Also, its rate of convergence is known to be slow. We have previously proposed a Bayesian approach, TIGAR (https://github.com/nariai/tigar), in which model parameters are handled as the posterior distribution and are estimated by variational Bayesian inference in order to avoid the overfitting problem and control the model complexity (the number of transcript isoforms). Here, we have implemented an improved algorithm to eliminate unnecessary calculations in each iteration of variational Bayesian inference, which significantly speeds up the convergence of the iterative computations. We apply our method to both simulation data and real data and evaluate performance in comparison to existing methods. Our method performs better than existing methods for both single-end and paired-end reads.

1396M

Phy-Mer: A novel alignment-free and reference-independent mitochondrial haplogroup classifier. *D. Navarro-Gomez[1], J. Leipzig[2], M. Lott[2], L. Shen[1], A. Stassen[3], D.C. Wallace[2,4], M. van Oven[5], M.J. Falk[4,6], X. Gai[1].* 1) Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA; 2) 2Center for Mitochondrial Medicine and Epigenomic Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Clinical Genetics, Maastricht University, The Netherlands; 4) University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 5) Department of Forensic Molecular Biology, Erasmus MC - University Medical Center Rotterdam, The Netherlands; 6) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Accurate mitochondrial haplogroups determination is critical for population genetics, anthropology, and forensics. It is equally if not more important for clinical genetics as mitochondrial haplogroup associations with human diseases have been identified. Despite this, there is a lack of computational tools to facilitate automated classification starting with mitochondrial sequences. A mitochondrial sequence has to be aligned to a mitochondrial reference genome in order to identify the polymorphisms, which are then compared individually and frequently manually to the haplogroup determining polymorphisms, such as those summarized in Phylotree (http://www.phylotree.org/). What complicates more is that the choice of optimal mitochondrial reference genome remains controversial as some researchers pointed to various limitations with the commonly used Revised Cambridge Reference Sequence (rCRS) and proposed to replace rCRS with Reconstructed Sapiens Reference Sequence (RSRS) instead. There is a strong need therefore for an alignment-free mitochondrial haplogroup classifier that is hence agnostic to the reference sequence of choice. Results: Using a k-mer approach, we developed a Python software package called Phy-Mer that determines the associated haplogroup of a mitochondrial sequence to the highest resolution currently feasible based on Phylotree without having to align it. Comparisons with other available manually curations, and existing haplogroup tools such as HaploGreb and Mitomaster, revealed superb performance of our tool, in terms of accuracy, sensitivity, as well as the ease to use. Phy-Mer has an added functionality such that next-generation sequence data can be used directly as input. Accurate haplogroup call can be obtained in less than 1 minute per sample with FASTA input and about 2 minutes per sample with BAM file of either whole-exome or whole-genome sequencing data sets on a desktop computer. Availability: The source code and sequence library is being made freely available on GitHub (https://github.com/), as well as the websites of the Mitochondrial Disease Sequence Data Resource Consortium (http://mseqdr.org) and PhyloTree.org (http://www.phylotree.org/).

1397T

Application of Machine Learning Techniques to Next Generation Sequencing Quality Control. *S.M. Nicholls[1,2], A. Clare[2], J.C. Randall[1].* 1) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Aberystwyth University, Aberystwyth, Wales, United Kingdom.

Advances in genetic sequencing hardware have introduced the concept of massively parallel DNA sequencing, reducing both the time and cost involved in performing genetic analysis. However, these "next-generation" processes are complicated and open to error, thus quality control is an essential step to ensure confidence in any downstream analyses carried out.

As part of the Sanger Institute's human genetics informatics pipeline, a system called *auto_qc* is responsible for applying quality control to samples within the pipeline. It does this by comparing a subset of quality metrics output by *samtools stats* to hard-coded thresholds, determining whether a particular sample has reached a level that requires a warning, or has exceeded the threshold and failed a test entirely. Whilst this catches most of the very poor quality outputs, a large number of samples are flagged for manual inspection at the warning level; a time consuming task which can invite inefficiency and error.

Using a set of 13,455 lanelets (each containing data from one indexed sample extracted from a single lane of multiplex sequencing) drawn from 4,915 DNA samples we investigate the accuracy of a range of machine learning classifiers for a series of quality control data sets and show it is possible to generate minimal yet accurate models, creating decision trees which closely resemble the behaviour of an already existing manually curated QC system -- using stratified cross-validation, a method which measures the proportion of samples that were correctly classified as indicated by the manually curated training set, our models were able to gain scores of 95% and above. To support this work we have developed *Frontier*, a Python package which provides users with interfaces for the reading, storage and retrieval of *samtools stats* data for the purpose of working with a machine learning framework. *Frontier* is freely available open-source software hosted at https://github.com/SamStudio8/frontier.

1398S

Computational Medicine for Investigation the functional Sterol Regulatory Element Binding Protein-1 gene Polymorphism: a new challenge for Glucose-6-phosphate dehydrogenase deficiency biology. *A. Palasuwan.* Faculty of Allied Health Sciences, Chulalongkorn University, Bangkok, Thailand.

Glucose-6-phosphate dehydrogenase (G6PD) deficiency is the genetically hemolytic disease and commonly presents in Southeast Asia particularly Thailand. The severity and pattern of clinical manifestation vary greatly both among individuals with G6PD deficiency and among people of different ethnic backgrounds. Although the relevance of G6PD variants and clinical manifestations have extensively described in various populations, the association of G6PD variants and enzymatic activity is not clearly understood. Recently the expressions of G6PD have been shown to be regulated by Sterol Regulatory Element Binding Protein-1 (SREBP-1) gene. To uncover the primary causative SNP of SREBP-1 gene related to human phenotypic variation in statistical effectiveness with large sample set, the new inventory technology such as high throughput genome sequencing, DNA chip technology are the promising tools. In order to reduce the cost and time consuming in a large cohort genetic association study. Computational genetics approach facilitates the selection of potentially functional genetic polymorphisms and the plausible biological explanation of genetic association study results. In this study, the in-silico genetic analysis approach has been used for uncovering the potential functional genetic polymorphisms of SREBP-1 gene. The genetic polymorphism mining technique, functional effect prediction, linkage disequilibrium(LD) approach of SREBP-1 gene polymorphisms were extensively studied in reference populations from HAPMAP project. Many candidate single nucleotide polymorphisms (SNP), a nucleotide substitution, localized in regulatory region have been identified as a functional SNP. However nonsynonymous SNPs also demonstrated the possible functional effect. The large haplotype block obtained from LD analysis indicated a high nonrandom association in this region. We therefore proposed the usefulness of computational genetic approach for analyzing SREBP-1 gene and demonstrated the fascinating results of high potentially functional SNPs which can be a good candidate marker for human genetic association study. The utilization of computational genetic approach is the meaningful tool in the field of life science both in pre-and post-experimental study.

1399M

Computational tools for discovery of patterns and associations in genetic and genomic data. *P. Pavlidis, E. Portales-Casamar, G. Charathsandran, E. Chong, C. Kwok, C. McDonald, N. St. Georges, P. Tan, A. Zoubarev, S. Rogic.* Psychiatry / Centre for High-Throughput Biology, University of British Columbia, Vancouver, British Columbia, Canada.

We present three free, open source integrated computational systems designed to aid researchers who are analyzing complex genetic and genomics data sets. They address common challenges in functionally characterizing and prioritizing experimentally obtained gene lists or genetic variants.

Phenocarta (phenocarta.chibi.ubc.ca) is a curated gene-phenotype knowledgebase consolidating information from over 15 data sources. Phenocarta currently holds more than 120,000 lines of evidence linking over 12,000 genes to nearly 3000 different phenotypes. We are introducing a novel evidence scoring scheme that will further enhance the utility of the data. Phenocarta data can be downloaded for use in other tools such as ErmineJ (gene set analysis).

Gemma (gemma.chibi.ubc.ca) is a database, analysis software system and web site for genomics data re-use and meta-analysis. Currently, Gemma contains analyzed data from over 6000 curated and re-analyzed expression profiling studies (microarrays and RNA-seq), yielding hundreds of millions of differential expression results and coexpression patterns (correlated expression) for retrieval and visualization. With optional registration users can save their own data and securely share it with other users. Gemma can be used to identify potential functional relationships among genes and with respect to phenotypes, and provides multiple search, analysis, browsing and visualization tools.

The newest tool, **ASPIREdb** (aspiredb.chibi.ubc.ca), supports the exploration, analysis and mining of complex phenome-genome datasets. The system has a powerful query engine for complex searches of genotype and phenotype characteristics in conjunction with data from external sources, such as DGV and DECIPHER. We will describe analysis tools to mine the data, including the ability to perform enrichment analysis of phenotypes in user-identified subject subgroups. ASPIREdb provides graphical representations, including an interactive ideogram, heat map views and genome browser integration. Through integration with Gemma and Phenocarta, ASPIREdb provides information on gene networks, differential gene expression and gene-phenotype associations.

1400T

Sparse structural equations for joint phenotype-genotype network analysis. *M.L. Rahman, P. Wang, M. Xiong.* Human Genetics Center, University of Texas Health Science center at Houston, Houston, TX.

Multivariate lineal model is the most widely used statistical method for genetic studies of multiple phenotypes. However, multivariate linear model for genetic studies of multiple phenotypes has several serious limitations. First, it ignores the correlations among the phenotypes. In practice, the most clinical phenotypes are often correlated. Overlooking correlation information of phenotypes will compromise identification of genetic architecture of multiple phenotypes. Second, it cannot be used to estimate the causal effects of genetic variants on the phenotypes and causal relationships among phenotypes. To overcome these limitations, we propose sparse structural equations for modeling the phenotype-genotype networks. In other words, the multiple phenotypes are presented as endogenous variables and genetic variants are modeled as exogenous variables in the structural equations. Since the current methods of parameter estimation for structural equations will lead to fully connected networks, we use regularization principal to reduce densely connected networks to sparse networks and develop alternative direction methods of multiplier to estimate parameters in the structural equations and identify the structure of phenotype-genotype networks. We formalize the causal inference in the phenotype-genotype networks using Peal's do operation. The directed acyclic graph (DAG) or its Markov equivalence class corresponding to the structural equation model is estimated from data and total causal effects of genetic variants on the phenotypes will be estimated by the DAG and Peal's back door operation. The proposed method was assessed by large simulations and applied to the NHLBI's Exome Sequencing data set with 12 phenotypes and 7,194 genes which were located in 234 pathways. We constructed phenotype-genotype network for each pathway and estimated their causal relationships. We used test statistics to formally test association of the SNPs in the networks with 12 phenotypes and identified a total of 89 genes significantly associated with at least one phenotype.

1401S

Human Splicing Finder: An invaluable system to annotate the impact of mutations on splicing signals. *G. RaÃ[1,2], D. Salgado[1,2,3], J.P. Desvignes[1,2], A. Blanchard[1,2], M. Miltgen[1,2], G. Blandin[1,2], C. Guien[1,2], A. Pinard[1,2], L. Barré[1,2], S. Olschwang[1,2,4], G. Collod-Béroud[1,2], C. Béroud[1,2,4].* 1) Aix-Marseille Université, GMGF, 13385, Marseille, France; 2) Inserm, UMR_S 910, 13385, Marseille, France; 3) EMBL Australia; Australian Regenerative Medicine Institute (ARMI), Monash University, Building 75, Clayton, Victoria 3800, Australia; 4) APHM, Hôpital TIMONE Enfants , Laboratoire de Génétique Moléculaire, 13385, Marseille, France.

The NGS technologies have dramatically changed our approach to rare human genetic diseases, leading to the identification of many disease-causing genes: the challenge is not anymore data production but data interpretation. It has been reported that many mutations (including intronic and up to 50% of missense) are not pathogenic because of their impact on the protein but rather because of their impact on the correct assembly of the pre-mRNA splicing machinery. Thus it is necessary to integrate this level of information in NGS bioinformatics pipelines. The Human Splicing Finder (HSF) system combines 12 different algorithms to identify and predict mutations' effect on splicing motifs including the acceptor and donor splice sites, the branch point and auxiliary sequences known to either enhance or repress splicing: Exonic Splicing Enhancers (ESE) and Exonic Splicing Silencers (ESS). If the HSF system has quickly become an international reference, the complexity of splicing signals often result in a flow of information preventing correct interpretation by non-specialists. To make this information accessible to all we have created an expert system for data interpretation. This algorithm uses a decision tree based on mutation position: intronic vs. exonic, localization or not in wild type splice sites or branch point. Based on these criteria, context-irrelevant signals are automatically ruled out to focus on relevant ones to evaluate the potential impact of mutations. To evaluate the data-interpretation expert system, we selected two datasets: 200 mutations reported to affect splicing (150 intronic and 50 exonic); 70 mutations (intronic and exonic) reported as having no effect on splicing signals. We then compared predictions with the manually interpreted data and with the in vitro results. The first comparison gives a 100% concordance, while the second demonstrate accuracy >0.96 for donor and acceptor splice sites; 0.95 for branch point but only 0.57 for ESE/ESS. This last result was expected, as the prediction of such signals is still difficult because of insufficient experimental data. Furthermore, to allow implementation of the HSF system in NGS pipelines, we created a dedicated webservice. In the NGS era, the HSF system could now be used both in research and clinical practice by non-experts thanks to its new data-interpretation expert system and its dedicated webservice. It could be combined with other systems such as the UMD-Predictor for WGS.

1402M

Plot: A tool to automatically summarise single variant analyses. *N.W. Rayner[1,2,3], N. Robertson[1,2], M.I. McCarthy[1,2,4].* 1) Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK; 3) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 4) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK.

With the continued drop in the costs of genotyping chips and sequencing, along with the increasingly diverse number of phenotypes collected on cohorts, the number of genome wide association analyses and meta-analyses being performed has increased dramatically in recent years. The analyses are generally amenable to automation via scripting: however it is time consuming and error prone summarising the many output files. To address this problem we have developed a program, Plot, which automates this process providing a quick and easy way to generate a single Portable Document Format (PDF) document of multiple single variant analyses. The program runs as a two stage process, the first stage creates a Manhattan plot, adding chromosome and position information if required, and a Quantile Quantile (QQ) plot as well as a table of all independent variants, currently defined as distance >100kb, with a p value $<10^{-5}$. The table is annotated with predicted variant effect, if any, and whether the variant is within or near genic features. Other user definable metrics such as minor allele frequency, alleles and effect allele can also be included. The top 20 independent variant signals are also plotted using LocusZoom, its standard PDF output is automatically converted to jpeg files for inclusion in the next stage. The second stage compiles the stage one outputs into an eXtensible Markup Language (XML) DocBook document. This is formatted into an eXtensible Stylesheet Language - Formatting Object (XSL-FO) which describes a two page per analysis report comprising the Manhattan and QQ plots on the first page and the top 20 table and the top 4 variant's locusZoom plots on the second. Finally the XSL-FO is rendered using Apache™ Formatting Objects Processor (FOP) into a single PDF. The program can cope with up to two files per analysis and supports many different summary file formats from both sequencing and genotype chip based single variant analyses and meta-analyses, such as SNPtest, METAL, GWAMA and EPACTS. New formats can easily be added to the program via the properties file. To date the program has been used successfully by a number of analysts working within some of the large consortia such as SUMMIT, DIRECT, ENGAGE and UK10K running on up to ~150 different analyses at a time.

1403T

Network Modeling of Transcriptional Response to Influenza Vaccination. *A. Renwick, J. Belmont, C.A. Shaw.* Baylor College of Medicine, Houston, TX.

Although genome wide data has become common in biomedical science, fully exploiting the mutlivariate character of the data has remained a challenge. Instead of using the data more holistically to form and test systems level hypotheses, the data are frequently used as a massively parallel screening tool to identify individual candidates. When more systems level analyses are performed on data, they are often introduced sequentially. In this approach, data are analyzed first in a massively parallel fashion, and then the filtered results (the "hits") from the parallel analyses are introduced to content-based enrichment analyses. A potentially more powerful approach is to analyze the data from the systems level in the primary analysis. A number of methodological tools potentiate this analysis approach. One method is the Gaussian Lasso, which we use to infer an undirected bayesian network among changes in transcript levels. We have performed a comparative analysis of a sequential gene content approach vs the network approach using our primary data for human influenza vaccine response. Human response to vaccination invites a multivariate approach as it is a complex phenomenon involving signals between cell types and migration of cells between tissues. The inferred network identifies expression modules that respond coordinately to vaccination. Our analysis shows the interferon signaling system is extremely important in the first day after vaccination. Our results highlight the similarities and differences as well as the additional power of the multivariate content analysis approach.

1404S

A Novel Family-based Approach for Analysis and Interpretation of Exome Sequencing Data in Pedigrees. *R. Robison, M. Zhu, J. Zhu, L. Dobias, M. Brown, T. Richards, C. London, B. Beebe, T. Lefler, G. Simon, K. Wang.* Tute Genomics Inc. 150 S. 100 W. Provo, UT 84601.

Introduction: The widespread use of exome sequencing has created the need to rapidly understand the functionality of identified variants and evaluate their consequences on human phenotypes. One key way to address this is by using a family-based statistical approach to analyze the data in order to more rapidly identify causal variants for clinical syndromes. Methods: We developed a web application called Tute Genomics for comprehensive and user-friendly annotation and interpretation of genomic data, including a case-control and family-based analyses. Analysis was completed using the following approach: (1) comprehensive variant annotation with over 100 annotation types (2) variant filtering using default or user-specified criteria (3) probabilistic modeling of variants, using a machine-learning algorithm to rank all variants in a genome by their likelihood of having functional consequences with respect to a specific phenotype of interest (known as the Tute Score). Pedigree information and optional phenotype information were incorporated in the analysis as prior information, and variants were identified with labels such as inherited, de novo, compound heterozygous and shared with sibling. This information, along with Tute Score, was then used to calculate the family adjusted variant scores and gene scores in a probabilistic approach. Results: In an exome sequencing study on a pedigree where one subject was discovered to have idiopathic hemolytic anemia (IHA), we identified two heterozygous variants in *PKLR* gene, a gene previously confirmed to be associated with IHA, ranked 7th (c.1022G>C) and 11th (p.R569Q) out of all variants based on Tute Scores. When applying a recessive model where the effect of compound heterozygosity or homozygosity was considered, the *PKLR* gene ranked 3rd based on the Tute Gene Scores. When further applying a phenotype label of 'anemia' in the analysis, the *PKLR* gene ranked 1st based on adjusted Tute Gene Scores. This analysis demonstrates that our pipeline can effectively prioritize disease-causing mutations and genes in pedigrees with exome sequencing data. Conclusion: Tute can rapidly interpret large volumes of sequencing data, including family-based datasets, via a web-based clinical genome interpretation platform, with comprehensive annotation types and user-friendly interface for conducting analyses. Our results demonstrate the effectiveness of this family-based approach to support disease gene discovery.

1405M

Improving Computational Prediction of Clinically Relevant Genome Variation. *A. Rychkova, C. Bustamante.* Genetics Department, Stanford University, Stanford, CA.

Rapid, accurate, and inexpensive genome sequencing promises to transform medical care. A critical hurdle to enabling personalize genomic medicine is predicting the functional impact of novel genome variation. This is a particularly pressing problem at "clinically relevant genes" where some mutations are already known to impact phenotype (such as the BRCA1/2 cancer susceptibility genes), but where we have a very incomplete map of how genotype impacts clinical phenotype. Differentiating "benign" from "pathogenic" genetic variants is far from an exact science and, oftentimes, a doctor is left with the frustrating and inconclusive result that a patient carries a "Variant of Unknown Significance" (VUS). Existing computational approaches to variant classification all suffer from low overall accuracy rates. Their poor performance limits the general utility of these tools in the determination of whether a novel genetic variant at a "clinically relevant gene" is actually related to the disease of interest or whether incidental findings ought to be returned and when. In this project we considered variants located in the coding region of the genome, in particular missense mutations, which cause the amino acid change in the protein. Such mutations might lead to diseases by interrupting protein folding and decreasing its overall stability. Protein stability is a fundamental property affecting function, activity, and regulation of biomolecules. It is measured by the folding free energy, the free energy difference between the folded and unfolded states of proteins. Therefore, estimated folding free energy should give valuable information about the functional consequence of missense mutations. Here we develop a machine learning method that combines the predictor of protein stability with other predictors based on amino acid sequence, structure and evolutionary information. As a first pilot experiment, we focused on the Clinical and Functional TRanslation of CFTR (CFTR2) database, which assembles clinical data and accompanying CFTR variants from individuals with cystic fibrosis (CF). We used variants found in this database along with 1000 Genomes data to train and validate our method. Along with the NCBI ClinVar and NHGRI ClinGen databases we are helping develop, we will use other Locus Specific Data Bases to evaluate functional consequences of variants in clinically actionable genes reported by ACMG, and those used in newborn screening.

1406T

Cancer is a Zero Sum Game between Cells and Cells! *A.R. Salehi Chaleshtori[1], S. Bamohabbat Chafjiri[2].* 1) Medical Genetic Center of Genome, Isfahan, Iran; 2) Information and Security LAB, Sharif University of Technology, Tehran, Iran.

Cancer is the major problem of our age and massive efforts performed to solve this problem. Cancer is very variable and tumor heterogeneity is documented for many characters, including the production of growth factors, one of the hallmarks of cancer. Combined interactions within the tumor can maintain heterogeneity for the production of growth factors and explain short term effectiveness of some therapies like RNAi. Alternative strategies for evolutionarily stable treatments are discussed, but this study propose that these strategies must be completed by mathematical principles like Game theory. Game theory is a social science whose aim is to understand the behavior of interacting decision-makers. We propose that determination of equilibrium point between growth factors and apoptosis inducers can be very effective in cancer treatment as well as cancer prediction. Cancer is recognized by uncontrolled growth, invasion, metastasis and death in the end. This process is initiate by defect in equilibrium point and defeating the whole system finally. Cells that overexpressing oncogenic products elevate their growth and divisions to access the impermanent success but this process led to whole system defeat permanently. Regarding to the game theorem, we propose that tumorigenesis is a zero sum game (or competitive game) in which canceric cells elect the same strategy (further growth and division), controlling the game and finally win the game. If other cells select a definite strategy to play against cancerous cells, this cause to control the game by these healthy cells and winning the game eventually. These selfish canceric cells prefer their success to the system but if these cells possess their balance, don't exceed from equilibrium point through normal expression of oncogenes and apoptosis inducers and select stable strategy against cancerous cells, system can access the permanent success entirely. It is concluded that game theory is an applied issue in cancer treatment and prediction. Normal cells can control the cancer if these normal cells operate stable and choose the right strategy. These cells require to be trained in this field, win the game and controlling the cancer.

1407S

UMD-Predictor: A variant annotation masterpiece for NGS pipelines. *D. Salgado[1,2,3], J.P. Desvignes[1,3], G. Rai[1,3], A. Blanchard[1,3], M. Miltgen[1,3], A. Pinard[1,3], G. Collod-Béroud[1,3], C. Béroud[1,3,4].* 1) Aix-Marseille Université, GMGF, 13385, Marseille, France; 2) EMBL Australia; Australian Regenerative Medicine Institute (ARMI), Monash University, Building 75, Clayton, Victoria 3800, Australia; 3) Inserm, UMR_S 910, 13385, Marseille, France; 4) APHM, Hôpital TIMONE Enfants , Laboratoire de Génétique Moléculaire, 13385, Marseille, France.

Whole Exome Sequencing (WES) technologies are increasingly applied to clinical practice and medical research. Those experiments generate 50 to 90,000 variations per individual from which only one or two are pathogenic mutations responsible for Mendelian diseases. It is thus crucial to differentiate non-pathogenic from pathogenic mutations to limit downstream analysis. To do so, we created the UMD-Predictor system (http://umd-predictor.eu) to annotate all coding variations for their potential pathogenic effect. UMD-Predictor contains all nucleotide substitutions from human protein coding transcripts annotated through a combinatorial approach that aggregates: data related to splicing signals, evolutionary data, biochemical substitutions matrices, mutant position, and frequency in the general population. Scoring values range from 0 to 100 and correspond to 4 classes: <50 = polymorphism; ≥50 <65 = probable polymorphism; ≥65 <74 = probably pathogenic, and ≥74 = pathogenic mutations. To evaluate UMD-Predictor's efficiency we used 4 datasets containing pathogenic (P) and non-pathogenic variations (NP): Varibench with dbSNP (19,335 P; 7,897 NP), Uniprot (20,821 P; 36,825 NP), Clinvar (10,669 P; 1,817 NP) and PredictSNP (24,082 P; 19,800 NP). We then compared our system to the commonly used systems SIFT 5.1.1, PolyPhen 2.2.2, Provean 1.1.3, Mutation Assessor 2, CONDEL 1.5, MutationTaster 2 (MT2) and CADD. Various statistical parameters were calculated including ROC curves and Diagnostic Odds Ratio (DOR) which measures the effectiveness of a diagnostic test and the log(DOR) to study the trade-off between sensitivity and specificity. For all parameters, UMD-Predictor gave the best results. For example, with the VariBench/dbSNP dataset it displayed a: ROC AUC of 0.954 vs. 0.834 for CADD; DOR of 86.6 vs. 12.6 for MT2 and a log(DOR) with a 2-logs increase vs. all others. In the NGS context, UMD-Predictor's webservice analysis time was 45s (± 4s) per dataset vs. 420-9740s for other systems. If all systems accurately annotate the pathogenic mutations, UMD-Predictor gave the shortest list of candidate mutations with an average of 739 versus 2027 for other tested predictors (64% reduction of false positives). We believe that the UMD predictor system will be a masterpiece in NGS pipelines both through its very high specificity (0.95), its availability through webservices and speed. It could be combined with other systems such as Human Splicing Finder for WGS.

1408M

SUGAR: high-resolution refinement of high-throughput sequencing reads considering their spatial organization in flowcells. *Y. Sato, K. Kojima, N. Nariai, Y. Yamaguchi-Kabata, Y. Kawai, M. Nagasaki.* Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan.

To obtain valuable insights from high-throughput sequencing data, it is essential to control low-quality portions of the data affected by technical errors like air bubbles in sequencing fluidics. As a part of quality control pipeline in the Tohoku Medical Megabank Project, we developed a software SUGAR (subtile-based GUI-assisted refiner) which can handle full data of Illumina HiSeq and MiSeq machines with user-friendly GUI interface and interactive analysis capability. The SUGAR generates high-resolution quality heatmaps of the flowcell (maximum 100×100 per tile), enabling users to find signals of technical errors during sequencing. In addition, the sequence read/nucleotide data generated from the error-affected regions of a flowcell can be selectively removed by automated analysis or GUI-assisted operations by the SUGAR. This automated cleaning operation based on the sequencing quality scores (Phred) was applied to a public human genome sequencing data, and we proved that the overall mapping quality (MapQ) was improved (29.5 to 30.1; an average MapQ of the discarded reads was 25.0). Such a high-resolution data cleaning enabled by the SUGAR would improve the subsequent variant analyses that require high-quality sequence and mapping results. Consequently, this software will be especially useful to control the quality of variant calls of the low population cells, *e.g.*, cancers, and improve the analyses of somatic mutations, mitochondrial micro-heteroplasmy, etc.

1409T

RaMWAS: Analysis software for rapid methylome-wide association studies. *A.A. Shabalin, K.A. Aberg, E.J. van den Oord.* Center for Biomarker Research and Personalized Med, Virginia Commonwealth University, Richmond, VA., USA.

DNA methylation plays important role in a variety of biological processes. It is critical in tissue differentiation and development and has been implicated in several diseases including cancer. High-throughput sequencing technologies allow for genome-wide measurement of DNA methylation at high resolution, up to the level of individual CpGs. Methylation measurements from about 30 million CpGs across hundreds to thousands of samples comprise billions of values. Analysis of such data is computationally complex and requires special tools to avoid excessive memory requirements and to allow for parallelization. We present a new R/Bioconductor package, RaMWAS, which has been developed to streamline methylome-wide association studies (MWAS). Starting with aligned reads, RaMWAS can perform all steps to conduct MWAS. First, it obtains the location of target CpGs from the human genome. Next, if not known in advance, the DNA fragment size distribution is estimated using the location of mapped reads around isolated CpGs. The estimate of DNA fragment size distribution is then used to calculate read coverage at the target CpGs. The full set of CpGs can later be reduced by exclusion of non-methylated sites and combining adjacent CpGs with highly correlated coverage estimates. The data reduction step decreases computational complexity of the association study and eases the multiple testing problem. Next, RaMWAS can perform principal component analysis to identify common sources of variation, such as from population substructure. Finally, RaMWAS can perform association testing between methylation coverage estimates and phenotypes of interest accounting for known covariates and the calculated principal components. RaMWAS can also perform eQTL-like analysis by testing methylation coverage estimates against gene expression or genotype measurements. Although RaMWAS is primarily designed for analysis of data from MBD-seq, MeDIP-seq, and similar technologies, it can also be used with whole genome bisulfite sequencing data. We tested RaMWAS on MBD-seq measurements from 1,500 samples with an average of 68 million reads per sample. Using RaMWAS we estimated methylation coverage at about 30 million CpGs, which were then collapsed to about 5 million CpG sites. Using the R package MatrixEQTL, RaMWAS performed methylation QTL analysis with 5 million SNPs, a total of 25 trillion statistical tests, in 1.5 days on a computing cluster with 50 nodes.

1410S

Genome and Transcriptome Free Analysis of RNA-Seq Data (GT-FAR) using cloud computing. *T. Souaiaia[1], K. Vahi[2], R. Mayani[2], J. Herstein[1], O. Evagrafov[1], T. Chen[3], E. Deelman[2], J. Knowles[1].* 1) Zilkha Neurogenetic Institute, USC, Los Angeles, CA., USA; 2) Information Science Institute, USC, Los Angeles, CA., USA; 3) Dept of Molecular and Computation Biology, USC, Los Angeles, CA., USA.

Despite being the current gold-standard for the evaluation of gene expression, alignment studies involving RNA-seq data frequently suffer from low-mapping rates and incorrect estimation of gene models, especially for model organisms with poorly annotated transcriptomes. To overcome these shortcomings, we have developed a cloud-based version of GT-FAR, our transcriptome optional RNA-seq analysis pipeline to bring accurate expression analysis to researchers without significant computational resources. GT-FAR consists of three independent modules which (1) Annotate a reference transcriptome and infer possible splice locations, (2) Filter, trim, and align reads to reference transcripts, introns and novel splice junctions and (3) cluster unmapped or raw reads into a compact sequence preserving index to facilitate to detection of significant sequence differences between samples and provide queries to multiple homology databases. The steps in each of these modules consist of multiple tasks with variable runtime and memory requirements; often local campus resources preclude the availability of nodes capable of providing knowles@med.usc.eduan efficient combination of CPU and memory availability. Additionally, even when resources are available, the need for on-demand computation may overwhelm the necessary computational tasks carried out by many small research groups. To minimize the financial and logistical strain necessary to run a state-of-the-art analysis pipeline we have modeled GT-FAR as a Pegasus workflow, enabling it to be a deployed as a virtual machine in the Amazon EC2 cloud. We are currnelty developing a custom web interface over Pagasus WMS for the virtual machine are developing a virtual machine which will allow users to upload custom datasets, track computational progress, and be alerted when their output is available for download. The Pegasus WMS represents the workflow in an abstract form that is independent of computational resources or file location resulting in executable workflows that can be deployed on local resources, remote clusters, or computing clouds. The Pegasus WMS also provides the additional wherin the event of a fatal error or hardware failure, a rescue workflow is created that performs only the tasks not yet completed.

1411M

Pedigree Reconstruction by PRIMUS using Exome Sequencing Data. *J. Staples[1], M.H. Cho[2], D. Qiao[2], E.K. Silverman[2], U.W. Center for Mendelian Genomics[1], D.A. Nickerson[1], J.E. Below[3].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Brigham and Women's Hospital, Boston, MA; 3) Epidemiology, Human Genetics & Environmental Sciences, University of Texas Health Science Center, Houston, TX.

We have developed a software program known as PRIMUS designed to reconstruct pedigrees from variant genotypes generated by genotyping arrays, exome sequencing, or whole genome sequencing. It quickly verifies that the provided samples match the expected pedigrees and helps to correct inconsistencies. PRIMUS is an efficient algorithm for accurate pedigree reconstruction, and is particularly useful in the analysis of Mendelian diseases. In this report, we validate the performance of PRIMUS on thousands of simulated pedigrees and demonstrate its utility in reconstructing, validating, and correcting expected pedigrees using simulated and clinical pedigrees obtained from the University of Washington Center for Mendelian Genomics (UW CMG). To investigate the performance of PRIMUS on a broad range of known pedigrees, we simulated pedigree structures of varying sizes, structures, genotypes, and combinations of missing data for individuals in the pedigrees. Our analysis revealed that PRIMUS reconstructs the true pedigree in 93.1% of the simulations and a partial pedigree for an additional 6.3%. PRIMUS is an essential tool for quality control within the UW CMG, and is used to reconstruct clinically obtained pedigrees submitted to the UW CMG. One dataset consisted of 49 pedigrees containing 351 individuals with Early Onset Chronic Obstructive Pulmonary Disease (EOCOPD). Using only the variants captured by exome sequencing, we calculated genome-wide identity-by-decent (IBD) estimates with PLINK and reconstruct each pedigree with PRIMUS. These results confirm that 43 of the 49 pedigrees matched the expected pedigree information collected in the study. Among the remaining pedigrees, PRIMUS identified and corrected a combination of non-paternity errors, unintentional sample swaps, and duplicate samples, which are common issues among all large pedigree datasets. Using PRIMUS, we completed validation and error correction of these pedigrees in under a minute of runtime and a few minutes of visual inspection. PRIMUS has detected inconsistencies in other submitted pedigree data that have altered the disease model used to identify the disease causing gene—ultimately, leading to the identification of the causal gene. Using only genotype data from arrays or exome sequencing data to verify and correct expected pedigrees, PRIMUS saves time and effort that would otherwise be spent on manual verification. Funding: NSF, NHGRI, and NHLBI.

1412T

DbGaP Phenotype Quality Control. *A. Sturcke, Y. Jin, S. Pretel, N. Popova, M. Lee, L. Ziyabari, M. Feolo.* National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health, Bethesda, MD.

NCBI's dbGaP (database of Genotypes and Phenotypes) has distributed over 5,800 phenotype datasets constituting ~4.8 billion phenotypic values to more than 2,000 approved investigators since 2007. The phenotype data consists of subject consents, subject-sample mapping (SSM), subject phenotypes, and sample attributes. The SSM links IDs found in phenotype data to IDs found in molecular genetic data, such as SNPs, expression, methylation, and sequence data. The phenotype data submitted to dbGaP often contains inconsistencies or missing information, which may skew subsequent analyses of associations between phenotypes and genotypes. We have created a dbGaP submission guide to standardize the formats of all submitted datasets and data dictionaries. All submitted files are initially checked by QC scripts for formatting inconsistencies and data integrity across subjects and samples, subject consents, and the mapping of subjects and samples. The QC process also identifies conflicts in subject gender, affection status, pedigree structure, and unexpected duplicates. Potential HIPAA violations are identified by QC scripts developed to detect the presence of visit dates, birth dates, zip codes, and ages over 89. The potential HIPAA violations reported by the QC scripts are further checked manually to identify true violations. Most of the HIPAA violations detected are dates. Datasets that meet criteria are then compared to their data dictionary to assure that every variable contains a description, that every encoded data value contains a code meaning, and that the values match the units listed and fall within the logical min/max. After the submitted data have passed the pre-loading QC tests, phenotype data are loaded into dbGaP databases. Variable summaries are generated and value distributions are displayed on a preview website for review by the submitter. Statistical summaries are further manually checked to detect missing code meanings and extreme data values that could potentially be used to identify individual participants and their families. Additional ID and gender checks are conducted between phenotype and molecular genetic data. This presentation will describe the phenotype quality control process designed to ensure the integrity and accuracy of the phenotype data distributed by dbGaP.

1413S

A case study for high throughput analysis of NGS data for translational research using Globus Genomics. *D. Sulakhe[1], A. Rodriguez[1], K. Bhuvaneshwar[2], Y. Gusev[2], R. Madduri[1], L. Lacinski[1], U. Dave[1], I. Foster[1], S. Madhavan[2].* 1) Computation Institute, University of Chicago, Chicago, IL; 2) Innovation Center for Biomedical Informatics, Georgetown University Medical center, Washington, DC.

Next generation sequencing (NGS) generates massive amounts of data and its analysis requires access to powerful computational infrastructure, high quality bioinformatics software, and personnel skilled to operate the tools. We present a practical solution to this data management and analysis challenge by using the Globus Genomics platform that simplifies large-scale data handling and provides advanced tools for NGS data analysis. Globus Genomics uses Globus Transfer to allow seamless data transfer from distributed data endpoints such as sequencing centers into its analytical platform for immediate analysis. It hosts an enhanced Galaxy instance on Amazon Web Service (AWS) resources with hundreds of widely used NGS analysis tools and many pre-defined best practices pipelines for whole genome/exome, RNA-seq, or ChIP-seq data analysis. Unlimited scalability and enabling simultaneous analysis of numerous data sets is possible due to the platform's ability to provision on-demand compute clusters on AWS and submit workflows to that cluster from Galaxy. Globus Genomics allows tool specific provisioning of Amazon EC2 nodes, thus accommodating a wide range of CPU and memory intensive analytical tools that require varying compute capabilities that helps in dramatically reducing execution times and costs. This platform has been prototyped successfully for the Innovation Center for Biomedical Informatics (ICBI) to enhance translational research at Georgetown University Medical Center. We have created exome and whole genome analytical workflows, a RNA-seq workflow, and have demonstrated the analysis at scale for multiple samples. We setup Globus Online data endpoints at ICBI's local storage and S3 based storage to provide easy access to the sequence data and to seamlessly transfer the input data into Globus Genomics for analysis. The analysis of a single 80gb whole genome (paired-end) took ~12 hours using memory intensive and compute intensive AWS cluster instance (cr1.8xlarge). To test scalability, we ran the exome analysis pipeline on a batch of 78 samples from a lung cancer study obtained from EBI's SRA (ERP001575) and it was completed in parallel in 40 hours. We also ran the RNAseq workflow on a batch of 21 samples of TCGA's ovarian cancer samples downloaded from Cancer Genomic Hub (CGhub) archive. The analysis of these RNA samples was completed in 24 hours. We will present detailed results, best practices and common bottlenecks overcome by using Globus Genomics.

1414M

Targeted alignment and end repair elimination increase alignment and methylation measure accuracy for reduced representation bisulfite sequencing. *Z. Sun[1], R. Kanwar[1], S. Baheti[1], M. Goelzenleuchter[2], J.P.A. Kocher[1], A.S. Beutler[1].* 1) Mayo Clinic, 200 1st ST, Rochester, MN 55906; 2) Charité - Universitaetsmedizin Berlin, Berlin, Germany.

Background DNA methylation is an important epigenetic modification involved in many biological processes. Reduced representation bisulfite sequencing (RRBS) is a cost-efficient method for studying DNA methylation at base level on a genomic scale, yet analyzing the RRBS data is challenging. Although several programs/pipelines are available for the data analysis, it is not clear which strategy performs the best and there has been no much attention to the contamination issue from artificial cytosines incorporated during the end repair step of library preparation. To address these, we have developed Targeted Alignment and Artificial Cytosine Elimination for RRBS (TRACE-RRBS), which aligns bisulfite sequence reads to MSP1 digitally digested reference and specifically removes the end repair cytosines before CpG methylation summarization. We have tested and compared the algorithms on a simulated and a real dataset in terms of alignment speed, accuracy, and accurate methylation estimate with 7 other RRBS tools and Illumina 450K microarray platform. **Results** By aligning RRBS reads only to the genomic regions where Msp1 digests, TRACE-RRBS demonstrated as the fastest, most sensitive and specific alignment tool among the 7 compared for a simulated dataset. For the real dataset with ~50 million of reads, TRACE-RRBS took about the same time as RRBSMAP, a third to a sixth of time needed for Bismark and Novoalign. TRACE-RRBS aligned more reads correctly than most of other tools and achieved the highest correlation with 450k microarray data. The end repair artificial cytosine removal increased correlation between nearby CpGs and methylation quantification. **Conclusions** TRACE-RRBS is a fast and accurate alignment and methylation quantification tool for DNA methylation from RRBS. It is implemented using platform independent Java programming language. The package is available for public use (http://bioinformaticstools.mayo.edu).

1415T

Unraveling epistatic causal genes of diseases with hyper-sensitive multiple testing procedure. *A. Terada[1,2], K. Tsuda[3,4,5], J. Sese[1].* 1) Department of Computer Science, Ochanomizu University, Tokyo, Japan; 2) Research Fellow of the Japan Society for the Promotion of Science; 3) Department of Computational Biology, The University of Tokyo, Chiba, Japan; 4) Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan; 5) JST ERATO Minato Discrete Structure Manipulation System Project, Hokkaido, Japan.

Epistatic interactions are key to unraveling the genetic contributions to complex diseases and traits. However, most of the analyses in genome-wide association study (GWAS) perform a statistical test for single SNP or gene. As the reason, the existing multiple testing procedures that are widely used to avoid false discoveries in GWAS, including the Bonferroni correction and the Benjamini-Hochberg procedure, are too conservative to detect such interactions, and finding any statistically significant combinations after applying such corrections is hopeless. We propose a novel multiple testing procedure to identify the epistatic effects of complex diseases, which is called the Limitless Arity Multiple-testing Procedure (LAMP). The LAMP can list any statistically significant combinations without any limit and can be substituted for the Bonferroni correction by rigorously calculating the family-wise error rate (FWER). We demonstrated that the LAMP compute the adjusted P-value by multiplying by a value that is 250 times smaller than the value by the Bonferroni correction or Holm procedure, while maintaining the FWER under the threshold. We also showed that the LAMP could identify significant combinations from existing GWAS data that have the potential to cause diseases, and that were overlooked by the Bonferroni correction. This procedure may contribute to the discovery of new combinatorial effects by the reanalysis of existing data.

1416S

Graphical algorithm for integration of genetic and biological data: Proof of principle using psoriasis as a model. *L.C. Tsoi[1], J.T. Elder[2,3], G.R. Abecasis[1].* 1) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) Department of Dermatology, University of Michigan, Ann Arbor, MI; 3) Ann Arbor Veterans Affairs Hospital, University of Michigan, Ann Arbor, MI.

Pathway analyses to reveal biological mechanisms for genetic association studies have great potential to better understanding of complex traits with major human disease impact. However, current approaches have not been optimized to maximize statistical power to identify enriched functions/pathways, especially when the genetic data derives from studies using platforms (such as Immunochip and Metabochip) customized to focus on markers near previously identified top-ranking loci. We present here a novel approach, called MEAGA (Minimum distance-based Enrichment Analysis for Genetic Association), with the potential to address both of these important concerns. MEAGA performs enrichment analysis using graphical algorithms to identify subgraphs among genes and measure their closeness in an interaction database. It also incorporates a statistic summarizing the numbers and total length of subgraphs, depicting the overlap between observed genetic signals and defined function/pathway gene-sets. MEAGA uses sampling techniques to approximate empirical and multiple testing-corrected p-values. We show in simulation studies that MEAGA has more power than count-based strategies to identify disease-associated functions/pathways, and the increase in power is influenced by the shortest distances among associated genes in the interactome. We applied MEAGA to the results of a meta-analysis of psoriasis using Immunochip datasets, and showed that associated genes are significantly enriched in immune-related functions and closer with each other in the protein-protein interaction network. Count-based strategies failed to identify functional enrichment in the same data.

1417M

A population- and pedigree-aware alignment strategy for Next Generation Sequencing data. *E. Valkanas[1], E. Flynn[1], T. Gall[2], J. Elson[3], A. Brandt[1], P. Pemberton[1], L. Carmichael[4], J. Osman[4], S. Leighton[4], M. Groner[4], D. Adams[1, 5], W. Gahl[5], T. Markello[1].* 1) NIH Undiagnosed Diseases Program Translational Laboratory, Bethesda, MD; 2) Walter Reed National Military Medical Center, Bethesda, MD; 3) Microsoft Research, Redmond, WA; 4) Appistry Inc., St. Louis, MO; 5) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD.

Next Generation Sequencing (NGS) data is conventionally aligned to a standard human reference sequence that provides a common reference for understanding genetic variation. However, since the standard human reference is derived from a limited number of samples, it cannot account for all non-pathogenic variations arising in the human population. This problem confounds alignment and genotyping of NGS data. The NIH Undiagnosed Diseases Program has developed a Diploid Alignment pipeline in order to increase the accuracy of short read alignment and improve variant calling. This strategy applies pedigree information in the form of genome-wide family SNP Chip data and imputes population data using SNP Chip data and HapMap information from the 1000 Genomes Project. The result is a set of three reference sequences for alignment: a maternal, paternal, and concatenated parental reference. The short reads from NGS were then aligned to each of these individualized references and the results were lifted over to the standard reference coordinates. The lift over enables results from the diploid alignment to be annotated with information from common databases. To determine the efficacy of this alignment strategy, we aligned a cohort of 70 individuals using this novel diploid strategy and compared the results to those obtained using a standard alignment pipeline. Using the diploid alignment strategy, there was a significant decrease in gapped alignments as well as in the number of differing variant calls. Results that differed between the two pipelines were Sanger validated; of the interpretable sequences, the diploid alignment strategy correctly called genotypes six times more frequently than the standard alignment strategy. These results suggest that an aligner that is both pedigree and population aware will call genetic variants with greater accuracy than an aligner that relies only on information from the standard human reference sequence.

1418T

My-Forensic-Loci-queries (MyFLq) BaseSpace application for analysis of forensic STR data generated by massive parallel sequencing. *F. Van Nieuwerburgh[1], C. Van Neste[1], D. Van Hoofstat[1], W. Van Criekinge[2], D. Deforce[1].* 1) Lab of Pharmaceutical Biotechology, Ghent University, Ghent, Belgium; 2) Biobix, Ghent University, Ghent, Belgium.

Forensic DNA profiles of short tandem repeat (STR) loci are currently obtained using PCR followed by capillary electrophoresis (CE). Massive parallel sequencing (MPS) technologies do not rely on size separation and thus relieve the limitations on locus multiplexy. MPS therefore creates extra possibilities within forensics for analyzing degraded samples, mixed samples, etc. Data analysis of raw sequencer reads, resulting in a reliable and usable forensic human identification report is still in early development. Only a couple of bioinformatics methods have been published. Unfortunately, all of these methods are command-line based and thus unsuitable for use by forensic analysts without a bioinformatics training. Recently, we published an open-source bioinformatic framework My-Forensic-Loci-queries (MyFLq) for analysis of MPS forensic data in a generic, platform independent way. The MyFLq framework was successfully applied on an Illumina MiSeq dataset generated from a multi-locus STR PCR on both single contributor samples and multiple person mixture samples.

We now present the newest developments on MyFLq. We created an application with an easy to use graphical user interface in which we have incorporated the MyFLq algorithms. Several new features have been added such a electropherogram style visual representation of the results, allowing for backward compatibility with CE results. To our knowledge, we are the first to present an open-source GUI application for forensic MPS data analysis. It can run as a standalone web application, or run embedded in the Illumina BaseSpace environment. In the latter, analysis is only a matter of running the samples on an Illumina machine and starting the MyFLq app to obtain a finished, graphical and interactive human identification report.

1419S

Identification of Somatic Mutations at Single-Cell Resolution. *X. Wang, K. Szulwach, P. Chen, G. Sun, M. Unger, R. Ramakrishnan.* R&D, Fluidigm inc, South San Francisco, CA.

Cancer genomes are extremely complex and diverse. When cells accumulate low-frequency mutations to form polyclonal cell populations, they co-exist with normal cells. It is difficult to identify these somatic mutations and the underlying clonal architecture from bulk cell populations, which mask the underlying heterogeneity of these rare cell types. DNA sequencing that incorporates the C1™Single-Cell Auto Prep System from Fluidigm provides a streamlined, upstream workflow for single-cell capture, lysis, and whole genome amplification. We have applied this workflow to 100 cancer/normal single-cells derived from the same individual [CRL-2338 (HCC1954) ductal carcinoma cells and CRL-2339 (HCC1954BL) normal B-lymphoblasts]. We subsequently performed whole-exome sequencing at an average of 27x coverage per cell. In order to facilitate the interpretation of variants identified in single cells, we developed an analysis workflow to identify somatic mutations and clonal architecture at single-cell resolution. In the analysis, we first determined the false discovery rate (FDR) of single cell genotyping based on high confidence homozygous sites in bulk genomic DNA, and then we applied the FDR to a binomial test (a cumulative distribution function) to determine the probability of observing a variant in a given number of cells among the total number of cells tested. A set of mutation filters, including the minimum number of single cells with mutations, variant allele frequency in tumor single cells, and the p-value of Fisher's Exact test of normal/tumor (and so forth), were used to eliminate false somatic mutations caused by random errors from whole genome amplification and sequencing. Our results demonstrate that the single-cell mutation analysis workflow can accurately detect somatic mutations in single cells and match results from the TCGA/ICGA "somatic mutation" calling benchmark study. The above analysis workflow has been integrated in an R package (SINGuLAR™Analysis Toolset 4.0) for free public use.

1420M

Sparse functional graphical model for joint analysis of RNA-seq and DNA sequencing data. *P. Wang[1,2], S. Guo[1,2], L. Jin[2], M. Xiong[1].* 1) University of Texas School of Public Health, Houston, TX., Select a Country; 2) Fudan University, Shanghai, China.

A focus of systems biology is to combine multiple types of genomic and epigenomic data and generate large networks of intermolecular interactions including physical associations underlying protein-protein and functional association measuring relationships between genes and correlated expression between genes. The growing availability and application of the high-throughput methods based on next-generation sequencing (NGS) stand as key players in systems biology, but the analysis and the interpretation of the vast amount of genomic and epigenomic NGS data poses great conceptual and computational challenges. The current methods for RNA-seq and DNA sequencing data integration are originally designed for microarray platforms. These methods either overlook position level and isoform information in RNA-seq or unable to deal with rare variants in NGS data of DNA. To overcome these limitations and multicolinearity in NGS data, instead of modeling the genome as a collection of separated individual loci, we model the genome as a continuum. To fully capture expression variation at the level of exon, chromosomal position, allele, splicing isoform provided by RNA-seq, and use entire spectrum of allelic information, we develop sparse functional graphical model for joint analysis of RNA-seq and NGS generated genotypes where the position-level read counts of expression or the number of risk alleles within a gene is taken as a function of genomic position. Alternative direction methods of multiplier was used to estimate the parameters in the penalized functional graphical model. Its goal is to simultaneously construct a sparse Gaussian graphical model and detect association of genes with the gene expressions. The proposed method was applied to RNA-seq data of ovarian cancer with 158 tumor samples from TCGA dataset. We jointly analyzed 64 genes in mTOR signaling pathway and 9 genes with somatic mutations. We reconstructed networks of expressions and somatic mutations with 54 edges. We identified association of mutated TP53 with PIK3CG, PCK3B, RRAGB, CAB39 and TSC1 expressions, the mutated TTN with the expressions of genes RICTOR, EIF4E1B and RPS6KA6, and the mutated DST with the expressions of gene PRKCA. In summary, the proposed method will shift the paradigm of systems biology research from multivariate data analysis to large scale penalized and functional data analysis and stimulate both theoretical and practical researches in systems biology with NGS data.

1421T

Swiss: a bioinformatics tool for identifying overlap between novel loci in GWAS scan results and a GWAS catalog. *R.P. Welch, T.M. Teslovich.* Biostatistics, University of Michigan, Ann Arbor, MI.

We present here a bioinformatics tool to scan genome-wide association study (GWAS) results and identify overlap between significant variants and previously identified variants in a GWAS catalog. This is a necessary step to identify novel associations, but to our knowledge there is no publicly-available software to automate the task. To this end, Swiss performs two primary functions. First, it identifies lead variants by clumping GWAS results, either by using measures of linkage disequilibrium (D', r2) to keep only the most significant variant per set of variants in LD, or by keeping the most significant variant within windows defined by physical distance. LD-based clumping is particularly useful for meta-analysis results, when it is not always possible to perform conditional analyses to identify statistically independent signals. The user must supply genotype data (in VCF format) to be used in LD calculations if LD-based clumping is desired; study-specific genotype data or files available from the 1000 Genomes Project FTP server may be used. Second, Swiss will attempt to find overlap between lead variants and a GWAS catalog, which can be generated by the user, or downloaded directly from the NHGRI GWAS Catalog website. Allowing the user to specify a customized catalog is critical, as many research groups maintain internal catalogs tailored to their traits of interest and/or including unpublished associations. Swiss identifies variants in the catalog that are in LD with each lead variant. The user can provide a different VCF for LD calculation in this step, for example when study-specific genotype data used for clumping of initial findings do not include all published GWAS variants. Swiss will also identify GWAS catalog variants within a user-specified distance of each lead variant, regardless of LD, thereby facilitating identification of known associations as well as secondary signals at the same loci. Swiss can be run in parallel on results from multiple traits, which is becoming increasingly important in metabolomics and eQTL studies with hundreds or thousands of traits. In the near future, we plan to integrate Swiss with LocusZoom, providing an easy way to identify significant lead variants from GWAS results and create LocusZoom plots highlighting both novel and previously known associations. We also plan to add built-in support for LD calculations using 1000 Genomes data for multiple ancestry groups.

1422S

Churchill: An Ultra-Fast Analysis Pipeline for the Discovery of Human Genetic Variation in Clinical and Population Scale Genomics. *P. White[1,2], B. Kelly[1], J. Fitch[1], D. Corsmeier[1], H. Kuck[1], A. Naik[1].* 1) The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) The Ohio State University, Columbus, OH.

Next generation sequencing (NGS) has revolutionized genetic research, enabling dramatic increases in the discovery of new functional variants in syndromic and common diseases. This technology has been widely adopted by the research community and is rapidly being implemented clinically, driven by recognition of NGS diagnostic utility and enhancements in quality and speed of data acquisition. Compounded by declining sequencing costs, this exponential growth in data generation has created a computational bottleneck. Current analysis approaches can take weeks to complete, resulting in bioinformatics overheads that exceed sequencing costs and represent a significant limitation. Churchill is a computational approach that overcomes these challenges, fully automating the analytical process required to take raw sequencing data through the complex and computationally intensive processes of alignment, post-alignment processing, local realignment, recalibration and variant discovery. Compared with alternative analysis pipelines, Churchill is simpler, faster, deterministic and 100% reproducible. Through implementation of novel parallelization techniques, Churchill enables computationally efficient whole genome sequencing analysis in less than two hours, ideal for clinical applications where turnaround time is paramount but data quality cannot be sacrificed. The algorithm enables population scale genome analysis to be performed cost-effectively using cloud resources. Churchill is able to effectively utilize distributed clusters of computers to allow the processing of data sets that would have otherwise been too large to process at a reasonable cost in a practical time frame. To demonstrate this we analyzed 1,088 whole genome raw data sets available from the 1000 Genomes Project Consortium (1KG) utilizing AWS resources. Despite the large number of samples, population variant frequencies were produced in less than 7 days for ~$11 per genome. We achieved a high degree of correlation to the 1KG SNP allele frequencies and discovered ~3 million INDELS not reported in the original 1KG analysis. As we look to the future, cloud computing will become indispensable in the analysis of human genome sequencing and the $1000 genome will soon be a reality. Churchill solves the sequence analysis computational bottleneck and through use of cloud computing resources enables rapid analysis of population scale sequencing datasets.

1423M

Neat-optimal whole genome reconstruction by a small set of genomic variants. *M. Xiong, N. Lin, J. Yu, L. Ma, J. Jiang, P. Wang, S. Guo.* Dept Biostatistics, Univ Texas Hlth Sci, Houston, TX.

The emerging genomic technologies will produce so large genomic and epigenomic data that traditional technologies and tools are unable to storage, transfer, manage and analyze them when the number of sequenced genomes exceeds ten thousands. It is urgent to develop novel concepts, theory and computational algorithms to filter out irrelevant and redundant genomic variants and select a set of sufficient genomic variants which can optimally recover whole genome information in the genomic data analysis. To meet these challenges, we propose to formulate genomic and epigenomic data as genomic matrices and selection of an essential set of genomic variants for the whole genome reconstruction as a subset selection problem for matrices. We develop matrix approximation theories and error estimation methods as powerful tools for optimal genomic matrix column and row selection. Both deterministic and randomized algorithms are designed to select genomic variants with provable guarantees that whole genome reconstruction by selected genomic variants can reach prespecified high accuracy. This nice property is due to the fact that matrix information largely depends on the rank of matrix. We show that selection of genomic variants is independent of phenotypes. To evaluate their performance, the proposed genome subset selection algorithms were assessed by large-scale simulations and applied to sparse principal component analysis in population genetics with 1000 genome data, quantitative trait prediction with sequencing data from the NHLBI's Exome Sequencing Project, cluster analysis of RNA-seq data from NIH TCGA project, prediction of a coronary artery disease (CAD) with GWAS data from WTCCC study. Our results show that the most genetic analysis using selected subset of genomic variants can reach the performance as good as using original whole genome datasets. For example, in CAD risk prediction study where 1,929 cases and 2,938 controls were sampled and the total number of SNP markers is 393,473. Using 65 SNPs out of 393,473 SNPs can reach average accuracy of 81.04% in the test datasets by 10 fold cross validation evaluation. However, using original dataset and traditional logistic regression, we can only reach 62.25% accuracy in the test datasets. In summary, the genomic matrix approximation theory and optimal column and row subset selection algorithms will have important implication in big genomic and epigenomic data analysis.

1424T

Detecting Nuclear Receptors Using a Finite Mixture Model. *M. Xu[1], D. Umbach[2], Y. Yao[1], L. Li[2].* 1) Intramural Research Program, National Institute of Menthal Health, Bethesda, MD; 2) Intramural Research Program, National Institute of Environmental Health Sciences, Bethesda, MD.

Motivation: Nuclear receptors are a special BiPartitie structure which binds to DNA sequences to regulate transcription. The typical orientations of the two BiPartitie structure includes direct repeat, reverse direct repeat, inverted repeat, and evert repeat. A large-scaled ChIP-seq data may contain all or some of these orientations of nuclear receptors as well as half-sites. It is a challenging task to identify those nuclear receptors and their binding sites from a large-scaled ChIP-seq data with potential enrichment of all the orientations. Results: In this paper, we develop a tool NRMotif which can be used to 1) estimate the position weight matrix (PWM) of the embedded nuclear receptors and distribution of the gap between the two half-sites of nuclear receptors from a large-scaled ChIP-seq data set; 2) predict the location of the nuclear receptor binding site; 3) estimate the proportion of each orientation of nuclear receptors. We propose a mixture model frame work to model all the orientation of the embedded nuclear receptors and consider Markov model for the background. An EM-based iteration algorithm is used to maximized the complete data likelihood. We conducted simulation study and applied the algorithm to real datasets. Comparing with other available tools, our software has been able to identify all the orientations of the bipartite structured dimers as well as monomers from large-scaled ChIP-seq data set.

1425S

Bayesian inference for tumour heterogeneity using the Hamming Ball Sampler. *C. Yau[1,3], P. Kirk[1], M. Titsias[2].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxon, United Kingdom; 2) Department of Informatics Athens University of Economics and Business (AUEB) Patision 76, 10434, Athens, Greece; 3) Department of Statistics, University of Oxford, 1 South Parks Road, Oxford, United Kingdon.

The characterization of sub-clonal structure from whole genome sequencing of heterogeneous tumors is a computationally challenging problem due to the large number of mutations and unknown number of tumor sub-populations. Many recent approaches have adopted methods based on Bayesian nonparametric models (the Dirichlet or Indian Buffet Process) but standard computational methods for inference for these models can be slow as they rely on inefficient Markov Chain Monte Carlo methods. We have developed a novel Markov Chain Monte Carlo sampling algorithm that allows efficient and exact Bayesian inference for multiple membership models that is scalable to massive datasets. While standard approaches use asymmetric updates that are prone to becoming trapped in local modes. Our algorithm uses locally symmetric modes that allows joint updating of the latent variable structure. We illustrate the utility of the sampling algorithm for tumor deconvolution problems and demonstrate its advantage over standard computational alternatives using in vitro mixtures of tumor DNA.

1426M

LVpicker: picking up true, low-frequency variants for studying cancer heterogeneity. *J. Zhang[1,2], J. Majewski[1,2].* 1) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Genome Quebec Innovation Centre, Montreal, Quebec, Canada.

Low-frequency, somatic mutations due to cancer subclones are common and important for studying the heterogeneity within a tumour mass that shed light on cancer biology and preventing the recurrences and progresses of tumours. Several bioinformatics tools, such as MuTect, are designed to call low frequency variants from paired normal and tumor samples. However, it is difficult to distinguish true, low-frequency variants (LFV) from background noise, etc. sequencing and mapping errors. Here we present an algorithm, LVpicker, to pick up true mutations from LFV. This approach uses a binomial test to calculate the probability of observing certain number of mutant reads in total reads in the test sample, given the site-specific error rate that can be generated from pooled control samples. The program also filters variants if all reads supporting the variant (including controls) have mapping quality lower than 30. Using LVpicker, we firstly studied a patient with the blood sample and saliva sample that underwent exome sequencing. Among 117 low-frequency variants called from the blood and saliva sample comparison, LVpiker rejected ~60% of the calls after Bonferroni correction. This demonstrated the ability of LVpicker to reject false positives from LFV. We then used LVpicker to study five spatially distinct biopsies from a primary glioblastoma multiforme (pGBM) tumour and the matched blood sample. LVpicker picked several true LFVs that potentially contributed to tumour evolution, including a known driver mutation in gene PIK3CA, which showed increased frequency during progression. We also used MiSeq to validate the false positives. Finally we manually decreased the frequencies of some high-frequency driver mutations in ATRX and H3.3 genes and demonstrated that those mutations did not have high site-specific error rate and could be picked by our program. Our work suggests that LVpicker is a useful tool to study cancer heterogeneity and should be applied after raw variant calling.

1427T

Efficient and accurate de novo assembly algorithm for paired-end reads and its application in indel calling. *L. Zhao[1], Y. Guan[1,2,3].* 1) USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine , Houston, TX., US; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX. US; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX., US.

De novo assembly of paired-end reads into long contigs has been demonstrated to have markedly greater power in solving genetic related problems. Unfortunately, existing assemblers either suffer from high time complexity or low repeat tolerance. In this study, a novel assembly algorithm is presented to overcome these difficulties, and the power of this method in indel calling is demonstrated as well. Giving the input of paired-end reads, the model partitions these reads into overlapping groups. Within each group, a Smith-Waterman type of alignment algorithm is designed to align paired-end reads, accounting for the random insert lengths. The alignment forms a temporary assembly object that usually consists of two short contigs. By repeating this procedure, a few temporary assembly objects are assembled. These temporary assembly objects are then aligned together by using a Smith-Waterman type alignment algorithm incorporating seeding to form a larger temporary assembly object. The process continues until a single, gap filled, and long contig is obtained. This bottom-up procedure of assembling contigs from paired-end reads are carried out in parallel to produce many long contigs that span 2000 - 3000 base pairs. Our method assembles haplotypes, but it retains two/multiple alleles caused by various situations at specific regions, such as the region with different paternal and maternal haplotypes, the repetitive region having multiple alleles, and the region containing paralogs. Based on the assembled contigs, we are able to call indels by mapping the long contigs to a reference genome. To evaluate the power of this method, we have simulated 2000 indels with size ranging from 1 base to 5 bases at the major histocompatibility complex (MHC) region of the Homo sapiens. Preliminary experimental results show that the true discovery rate of our method is 0.88 with the false discovery rate under 0.0025, while at the same false discovery rate level the true discovery rates of Fermi and velvet (two well-known assemblers) are only 0.37 and 0.23, respectively. Besides the greater indel calling performance, our method is superior to existing ones in the parallelizability *per se*. These observations indicate that our method has great potential in variants calling and clinical genetic diagnoses.

1428S

PGS: a tool for association study of high-dimensional microRNA expression data with repeated measures. *Y. Zheng[1], Z. Fei[2], W. Zhang[3], J. Starren[4], L. Liu[5], A. Baccarelli[6], Y. Li[2], L. Hou[5,7].* 1) Institute for Public Health and Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL; 2) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 3) Institute of Human Genetics, University of Illinois at Chicago, Chicago, IL; 4) Northwestern University Biomedical Informatics Center, Northwestern University, Chicago, IL; 5) Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL; 6) Department of Environmental Health, Harvard School of Public Health, Boston, MA; 7) The Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University, Chicago, IL.

MicroRNAs (miRNAs) are short single-stranded non-coding molecules that usually function as negative regulators to silence or suppress gene expression. Due to interested in the dynamic nature of the miRNA and reduced microarray and sequencing costs, a growing number of researchers are now measuring high-dimensional miRNAs expression data using repeated or multiple measures in which each individual has more than one sample collected and measured over time. However, the commonly used univariate association testing or the site-by-site (SBS) testing may underutilize the longitudinal feature of the data, leading to underpowered results and less biologically meaningful results. We propose a Penalized regression model incorporating Grid Search method (PGS), for analyzing associations of high-dimensional microRNA expression data with repeated measures. The development of this analytical framework was motivated by a real-world miRNA dataset. Comparisons between PGS and the SBS testing revealed that PGS provided on average 9% smaller phenotype prediction errors and 57% higher enrichment of phenotype-related biological pathways than SBS testing. Our extensive simulations showed that PGS provided more accurate estimates (root-mean-square-error=0.02, compared to 0.18 using SBS testing) and higher sensitivity (sensitivity=0.86, compared to 0.67 using SBS testing) with comparable specificity (>0.95). PGS is suitable for high-dimensional microRNA expression data with repeated measures in that, by exploiting underlying dependent structures, it relies on variable selection in the context of a multiple regression model, which circumvents multiple testing issues, enhances statistical power, and yields more biologically meaningful results. PGS is also applicable to other longitudinally collected high-dimensional quantitative data, such as, epigenomics, mRNA transcriptomics, proteomics, metabolomics, etc. The growing number of studies conducting high-dimensional profiling dataset using different platforms requires a more comprehensive evaluation of PGS in various study settings. (Availability: R source code for PGS tool is available for download at https://github.com/feizhe/PGS).

1429M

FExSeq: A familial exome sequencing discovery pipeline. *C. Chung[1,2], Z. Wang[1,2], X. Zhang[1,2], M. Wang[1,2], W. Luo[1,2], S. Suman[1,2], L. Burdett[1,2], M. Cullen[1,2], B. Hicks[1,2], M. Yeager[1,2], J. Boland[1,2].* 1) Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD; 2) Division of Cancer Epidemiology and Genetics, NCI, NIH, Rockville, MD.

Whole-exome sequencing has become a popular approach for disease-causing and/or disease-predisposing variant discovery. Analysis of massive amount of exome sequence data requires efficient bioinformatic approach. Although a number of state-of-art variant calling algorithms enabled accurate variant callings, a reproducible, streamlined pipeline for discovering biologically significant variants in the phenotype of interest is scarce. We have developed a pipeline called FExSeq discovery pipeline that utilizes familial exome sequencing data to 1) perform effective filtration by multiple parallel comparisons to public datasets; 2) to perform segregation analysis to identify variants that are family/disease group specific and/or shared among disease group; 3) to prioritize functional variants based on multiple annotation scheme. Here, we present the results from analyzing ~1,500 whole-exome sequenced samples from 710 families of 27 different cancer types. Our comprehensive FExSeq discovery pipeline effectively prioritize list of variants for follow-up functional studies.

1430T

A novel approach to methylation-Seq data analysis based on functional principle component analysis (FPCA). *S. Guo[1,2], L. Li[1], J. Jiang[1], N. Lin[1], M. Chen[2], L. Jin[2], M. Xiong[1].* 1) Human Genetics Center, Division of Biostatistics, University of Texas School of Public Health, Houston, TX 77030, USA; 2) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200433, China.

Methyl-CpG binding domain (MBD) protein-enriched genome sequencing (MBD-seq) has been considered as one of the most important approaches for methylome-wide association studies (MWAS). However, the present MBD-seq statistical analysis methods for differential methylated region (DMR), including 1) average reads in certain region or 2) to estimate methylation center (MACS2) are designed for analysis of methylation that is measured by microarray. Such methods have ignored the spatial characteristic in the function of DNA methylation, which would greatly decrease the power to detect the disease associated differential methylation region (DMR). In our present study, we propose a novel method for the analysis to the MBD-seq data in which the methylation profiles were taken as random functions andfunctional principle component analysis (FPCA) was introduced to fit the methylation profile for CpG island, CpG shore and CpG shelves. We develop FPCA-based statistics to test for differential methylation. We used spares hierarchical and K-means clustering to discover biologically meaningful pattern. Genome-wide DNA methylation-seq profile of 24 pairs of colorectal cancer (CRC) and normal tissues were used to validation the performance of the method. Simulation results showed that our proposed method substantially outperforms the traditional methods for testing differential methylation. In real data analysis, our method identified 3,425 hyper-methylation regions (HEMR) and 692 hypo-methylation regions (HOMR) in tumors, in which, 738 HEMR and 224 HOMR cannot be identified by traditional methods, among which the most important CRC associated CpG island methylator phenotypes (CIMP) such as CACNA1G, IGF2, NEUROG1, RUNX3, SOCS1, UCHL1, ADAMTS1 and AOX1 were identified. DMRs identified by FPCA showed significant consistency with the distribution of DMR in tumor and normal tissues. We found more than 39.3% of DMR were distributed in CpG shore and CpG shelves, which suggest the important biological role of CpG shore and CpG shelves in cancer development and progression. In summary, we provided an effective novel method to analyze next-generation MBD-Seq dataset based on Functional PCA (FPCA), which can identify aberrant DMR in colorectal cancer.

1431S

Next-generation sequencing reveals the presence and positions of novel duplications in clinical samples. *M. Kennemer[1], V. Semenysty[2], N. Patil[1], C. Hartshorne[1], M. Rabideau[1], A. Agbarya[2], Y. Kaplan[3], A. Dvir[3], L. Soussan-Gutman[3], K. Jacobs[1], M. Powers[1], J. Paul[1], S. Topper[1].* 1) Invitae Corporation, San Francisco, CA., United States; 2) Rambam Health Care Campus, Haifa, Israel; 3) Teva Pharmaceuticals Industries, Petach-Tikva, Israel.

Germline copy number variants (CNVs) account for a significant fraction of hereditary cancer pathogenic mutations. For example, approximately 5-10% of BRCA1 and BRCA2 pathogenic variants involve large deletions and duplications. Traditional analysis for these rare events was performed by multiplex ligation-dependent probe amplification, quantitative PCR or comparative genomic hybridization. Although germline CNVs can be detected from next-generation sequencing (NGS) data generated using targeted DNA capture technologies (e.g. exomes and other panels), methods for doing so must overcome many technical challenges. Several algorithms have been published to detect CNVs in such data, though they may not yet be adequate for use in diagnostic testing laboratories, particularly for detection of small single-exon CNVs. We have previously described a new method, CNVitae which is designed to detect single-exon CNV as well as larger regions sequenced using NGS with sensitivity and specificity of >99%. We have tested clinical samples where we performed NGS on 29 genes associated with hereditary cancer syndromes where CNVs were detected. One case is a 38 year-old Christian Arab woman diagnosed with invasive ductal carcinoma at 30 years of age. A duplication involving exons 5-11 of BRCA2 was identified from NGS read count data using CNVitae. To understand the impact of this duplication on the BRCA2 protein we analyzed sequence alignments from this region for either split-reads or discordant mate pairs and were able to confirm that this duplication occurred in tandem within the gene. The duplication is predicted to cause a Met to Arg change at codon 1594 followed by a frame-shift that ends with a premature truncation at codon 1597. The truncated protein is expected to result in a loss-of-function; a well-documented mechanism for BRCA2 inherited breast cancer susceptibility. In addition to being a novel duplication, this is the first clinically reported duplication in BRCA2 using these new methods of detecting CNVs. We have found additional examples of clinical cases where CNVs were detected using our methods of NGS and CNVitae software. This demonstrates the ability of NGS to not only detect single nucleotide sequence changes but also CNVs in a cost-effective scalable single assay for clinical genetic testing.

1432M

OncoRep: An n-of-1 reporting tool to support genome-guided treatment for breast cancer patients using RNA-sequencing. *T. Meissner, K.M. Fisch, L. Gioia, A.I. Su.* Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA, USA.

Introduction Breast cancer is the leading cause of cancer among females. Being a heterogeneous disease, it comprises multiple tumor entities associated with distinctive histological patterns, different biological features and clinical behaviors. Next generation sequencing allows us to study this heterogeneous disease in multiple dimensions. This provides in depth insight into tumor pathogenesis on the individual patient level paving the path to individualized medicine. The advent of individualized medicine introduces a shift in how cancer patients will be treated in the future, away from the one drug-one disease paradigm towards the idea of bringing the right drug to the right patient. **Problem** Challenges that arise with this paradigm shift are i) preprocessing and analyzing sequencing data in the n-of-1 setting ii) extracting relevant information from various layers of omics data iii) integrating omics data with drug databases, (iv) presenting the information to the clinician in an understandable manner, and (v) completing these steps in a timely manner to provide clinically relevant and actionable targets to a tumor board or the treating physician. **Results** To address these challenges, we present the extensible and customizable open-source framework OncoRep, an RNA-Seq based n-of-1 reporting tool for patients with cancer. We have applied it to breast cancer, including prospective molecular classification, detection of altered genes, and pathways, identification of gene fusion events, clinically actionable mutations and it reports suitable drugs based on identified actionable targets. It presents integration and visualization of these omics data in an approachable html based interactive report as well as a PDF based summary report, providing the clinician and tumor board with a tool to guide the treatment decision making process.

1433T

Identification of transcriptional regulators associated with breast cancer risk. *K.B. Meyer[1], M.A.A. Castro[2], F. Markowetz[1], B.A.J. Ponder[1].* 1) Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Cambridge CB2 0RE, Cambridgeshire, United Kingdom; 2) Department of Biochemistry, Federal University of Rio Grande do Sul (UFRGS), Rua Ramiro Barcelos, 2600, Anexo, 90035-003 Porto Alegre, Brazil.

Over 70 breast cancer risk loci have been identified at formal genome-wide significance, but there may be many more loci with small effect. The challenge is to understand how these risk loci synergise to influence disease risk, and to find read-outs of the combined risks. We are seeking evidence of common mechanisms by mapping the genes associated with these loci onto breast cancer gene regulatory networks. We have examined the regulatory network in breast cancer for enrichment of genes associated with breast GWAS loci[1]. First we used ARACHNe[2] to assign potential target genes, so-called regulons, to transcription factors (TFs) using gene expression data from Metabric[3]. A second step examined whether these regulons are enriched for genes associated with breast cancer GWAS loci, by applying an extended variant set enrichment[1,4] (EVSE) that uses eQTLs to define the genes associated with the known breast cancer risk loci[5]. We have implemented EVSE on a genome-wide basis and identified 36 TFs significantly associated with risk in both cohorts (995 and 997 samples) of Metabric. Interestingly four master regulators of FGFR2 signalling, previously defined as a risk-associated pathway[1], were found as risk TFs (ESR1, GATA3, FOXA1 and SPDEF). We extensively validated our EVSE results: (1) Negative control GWAS signatures e.g. bone mineral density or random SNPs do not yield a positive association. (2) EVSE with random but size matched regulons does not find a significant association. (3) Lastly, we demonstrate that the eQTL step in the analysis strongly influences the results: eQTLs from ER+ and ER- tumours identify different sets of risk associated TFs. By studying the correlation of gene expression between regulons, we found that risk TFs fall into two distinct, but internally highly correlated groups, each of which may represent a group of TFs relating to a common mechanism. Using the TCGA data set we find that 33% of the risk TFs showed somatic alteration in more than 3% of tumours, while only 0.3% of a set of control TFs were altered to the same extent, providing independent validation of our results and suggesting that similar pathways may be responsible for risk as well as disease progression. 1. Fletcher, M et al. (2013) Nat. Commun. 4, 2464-76. 2. Carro, M.S et al. (2010) Nature 463, 318-325. 3. Curtis, C et al. (2013) Nature 486, 346-352. 4. Cowper-Sal Iari, R et al (2012) Nat. Genet. 44, 1191-1198 5. Michailidou, K et al. (2013) Nat. Genet. 45, 392-398.

1434S

Molecular docking simulations provide insights in the substrate binding sites and possible substrates of the ABCC6 transporter. *O. Vanakker[1], M.J. Hosen[1,2], A. Zubaer[2], S. Thapa[2], B. Khadka[2], A. De Paepe[1].* 1) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 2) Genetic Engineering and Biotechnology, Shahjalal University of Science Technology, Sylhet 3114, Bangladesh.

The human ATP-binding cassette family C member 6 (ABCC6) gene encodes an ABC transporter protein (ABCC6), primarily expressed in liver and kidney. Mutations in the ABCC6 gene cause pseudoxanthoma elasticum (PXE), an autosomal recessive connective tissue disease characterized by ectopic mineralization of the elastic fibers. The pathophysiology underlying PXE is incompletely understood, which can at least partly be explained by the undetermined nature of the ABCC6 substrates as well as the unknown substrate recognition and binding sites. Several compounds, including anionic glutathione conjugates (N-ethylmaleimide; NEM-GS) and leukotriene C4 (LTC4) were shown to be modestly transported in vitro; conversely, vitamin K3 (VK3) was demonstrated not to be transported by ABCC6. To predict the possible substrate binding pockets of the ABCC6 transporter, we modeled a 3D homology model, qualified for molecular docking and virtual screening approaches. By docking 10 reported in vitro substrates in our 3D homology model and using VK3 as a negative control, we were able to predict two substrate binding sites for this protein. Further, virtual screening of 4651 metabolites from the Human Serum Metabolome Database in our model disclosed a prediction of possible substrates for ABCC6, which are mostly lipid and biliary secretion compounds, some of which are found to be involved in mineralization. Virtual screening expands this possibility to explore more compounds that can interact with ABCC6, and may aid in understanding the mechanisms leading to PXE.

1435M

A method for the discovery of long-range genomic interactions from 3C-seq experiments. *T. Yuan, M. Du, R. Dittmar, S. Xia, Y. Guo, L. Wang.* Pathology, Medical College of Wisconsin, Milwaukee, WI.

Abstract: Integration of chromosome conformation capture (3C) with high-throughput sequencing (HTS) (3C-seq) enables us to detect long-range genome-scale chromatin interactions between two genomic regions through counting the ligated fragments co-localized into these regions. However, the signals from real chromatin interactions are often overshadowed by a large number of the signals from short-range interactions or accident interactions. It is difficult to determine specific long-range chromatin interactions by setting a threshold of read counts (RC). Here, we present a new algorithm known as Tscore calculation. The algorithm assumes that the most chromatin interactions detected by sequencing are non-specific and randomly distributed, of which density functions can be fitted. For a specific genomic region (viewpoint) on a chromosome, the cumulative probabilities of the other genomic regions along this chromosome interacted with this viewpoint (cis-interactions) are calculated through the density function of the normal distribution fitted by multiplying the count numbers of cis-localizations and the distance between the viewpoint and the other regions, and the cumulative probabilities of the regions on the other chromosome interacted with the viewpoint (trans-interactions) are calculated through the density function of the exponential distribution fitted by the counting of trans-localizations between the viewpoint and these regions. The scores (Tscore), the conversion of the cumulative probabilities, can be used for measuring frequency of cis-/trans-interactions. WIn addition, by comparing various factors, e found that background noise and counting methods have significant effects on the detection of chromatin interactions. We therefore develop a tool known as 3C-analyzer in order to apply this algorithm into 3C-seq data analysis. 3C-analyzer integrated all analytic work from raw data in FASTQ format to the detection of significant chromatin interactions from 3C-seq experiments. It provides a user-friendly experience by including graphic user interface, sample management, and primary enzyme setup. The ability of parallel processing is optimized for large data analysis. Our case studies show that 3C-analyzer enable us to detect significant genomic interactions across different 3C technologies (Capture-C and 4C).

1436T

Using a reference panel to increase coverage in pooled sequencing experiments. *H. Al-Asadi, M. Stephens.* University of Chicago, Chicago, IL.

Pooled sequencing is an increasingly popular technique for assessing genome-wide population allele frequencies. Increasing the sequencing depth- or coverage - improves the accuracy of these allele frequency estimates but requires additional cost. Reference panels are often available in pooled sequencing applications and provide additional information about a sample -namely SNP correlations- which allows you to borrow information from linked SNPs. In this study, we demonstrate use of an LD-based model which incorporates information from a panel to increase the effective coverage for each SNP. For example, with two perfectly correlated SNPs, our approach doubles the coverage because reads from both SNPs can be used for any one SNP. We test our method on two evolve and re-sequence experiments.

1437S

The DNA Integrity Number: A novel approach for objective integrity classification of genomic DNA samples. *M. Gassmann[1], E. Schmidt[1], A. Inche[2], I. Pechtl[1], R. Salowsky[1], B. McHoull[2].* 1) Agilent Technologies, Waldbronn, Baden-Würtemberg, Germany; 2) Agilent Technologies UK Limited, Edinburgh, United Kingdom.

Genomic DNA (gDNA) is used as starting material in the experimental workflow of many applications in molecular biology. The integrity of the DNA critically affects the success of many downstream experiments like array CGH or sequencing. Initial electrophoretic analysis of the sample is highly recommended as the respective downstream applications can be expensive and time consuming. The Agilent Genomic DNA ScreenTape Assay has been primarily developed for the electrophoretic analysis of genomic DNA samples. A ScreenTape is a pre-packaged microfluidic device designed for performing electrophoretic applications in a microscale format. It is used in combination with the Agilent 2200 TapeStation instrument. Degradation of gDNA is typically a gradual process in which high-molecular weight DNA is fragmented into smaller species. It can occur either enzymatically, chemically or mechanically. Judging the integrity of DNA by visual evaluation of the electropherogram trace is subjective and can be error-prone. In order to standardize this a novel algorithm was developed to score gDNA samples on the 2200 TapeStation. The DNA integrity number (DIN) is calculated from several features obtained from the electrophoretic trace and ranges from 1 to 10. Here we show data demonstrating the reproducibility, scalability and linearity of the DIN. The DIN is independent from instrument, reagent and sample concentration variability and can be used as objective measure for determining the integrity of gDNA.

1438M

Genomic susceptibility for cancer prediction by supervised machine-learning methods on SNP-syntax. *S. Kim[1], M. Kim[2].* 1) Chemistry, University of California, Berkeley, Berkeley, CA., USA; 2) Computer Science, University of California, Davis, CA, USA.

It has been widely assumed that human genomic variations are associated with individual's susceptibility to complex diseases such as cancer. However, extensive genome-wide association studies so far had limited success in that the results have low predictive value of practical utility to individuals. We present a prediction process where two supervised learning methods are applied to two different descriptors of each individual's common genomic variations to predict an individual's susceptibility to each of 8 major cancer traits plus healthy trait. The accuracy of the prediction ranges from 33 to 57% depending on cancer type, significantly better than 11% for a random prediction, with probability estimates that may be useful for making practical health-decisions for individuals or for a population.

1439T

A composite classifier for prioritizing somatic SNVs based on predicted functional impact, protein disorder, and gene expression. *W. Liao, B.J. Chen, K. Wrzeszczynski, K. Arora, T. Bloom, N. Robine, V. Vacic.* Bioinformatics, New York Genome Center, New York, NY.

The majority of observed somatic variants are thought to be passengers and distinguishing them from true functional drivers of cancer remains an ongoing challenge. While great strides have been made in predicting the impact of variants in the context of inherited disease, only recently have tools specifically designed for annotating somatic variants been developed. One clear cause for this delay has been the lack of reliable datasets for training, but the growing availability of cancer genome data is helping to address this problem. Following the strategy utilized in previous work where recurrence and well-characterized gene lists were used to distinguish putative driver mutations from passengers, we composed a collection of positive and negative examples comprised of variants from the COSMIC, ICGC, TCGA, and UniProt repositories. Using these datasets to evaluate several annotation tools, we observed that while cancer-specific methods performed reasonably well, by and large they were outperformed by a composite of predictions from several tools—most of which were intended to assess variants associated with inherited disease, not cancer. Second, because most of current methods rely on sequence conservation as a proxy for function, we found them to be less sensitive in regions of intrinsic protein disorder. Finally, after applying these tools to variants predicted from TCGA glioblastoma samples where RNA data was also available, we found that sample-specific expression context provided valuable information, allowing us to deprioritize unexpressed genes while raising the priority of mutations in genes that were overexpressed or exhibited allele-specific expression—two notable examples being EGFR and TP53. This suggested that RNA-seq can be useful for inferring functional impact of somatic variants. In light of these observations, we developed a novel composite classifier that incorporates several cancer-focused prioritization algorithms in conjunction with well-established methods. Additionally, it considers a disorder prediction score, and examines over- and allele-specific expression at a per sample level. Taken together, our novel cancer-focused somatic variant prioritization method performed better than existing algorithms and represents a tool that can facilitate interpretation of somatic variants and understanding of cancer etiology.

1440S

Barcode-based template identification of KIR region in human genomes. *C. Lo[1], S. Zakov[1], S. Kim[1], B. Halldorsson[2], V. Bafna[1].* 1) Computer Science and Engineering Dept, University of California, San Diego, San Diego, CA; 2) Biomedical Engineering, School of Science and Engineering, Reykjavík University Reykjavik University. Reykjavik, Iceland.

The immunospecific regions of the human genome, such as the region encoding the killer cell immunoglobulin- like receptors (KIRs) on Chromosome 19 and the region encoding the major histocompatibility complex in humans (HLAs) on Chromosome 6, are among the most important regions functionally, but their hypervariability, in gene copy number and mutations, makes it difficult to assemble and characterize these regions. Previous methods characterizing these regions rely on laboratory methods using traditional and quantitative PCR primers and probes. Here, we propose a computational method to type the KIR region and to determine various allele groups of KIR genes directly from whole genome sequencing data. Our method is based on barcoding (deriving signatures) known KIR templates as well as sequenced fragments, and comparing the two sets of signatures.The method is alignment-free and efficient, and easily scales to large populations of individuals. We demonstrate the robustness and accuracy of our method over simulated data, trios from various populations in the 1000 genomes project as well as 298 Icelandic trios. The results indicate that this method can resolve all KIR types up to an equivalence that can only be broken by long range haplotyping or inheritance patterns.

1441M

Network-Augmented Genomic Analysis (NAGA) applied to Cystic Fibrosis studies. *S. Loguercio[1], D. Martino-Roth[2], D. Hutt[2], A. Su[1], W. Balch[2].* 1) Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA; 2) Department of Cell Biology, The Scripps Research Institute, La Jolla, CA.

Cystic fibrosis (CF) is an early onset disease characterized by a defect in the apical chloride channel, CF transmembrane conductance regulator (CFTR). The most common disease causing mutation is a 3 base pair deletion resulting in loss of Phe 508 (F508del), which leads to misfolding, endoplasmic reticulum (ER) retention and efficient ER associated degradation of the protein. To elucidate the molecular networks influencing the folding and function of CFTR in CF, we recently screened CFBE41o- cells containing F508del-CFTR against a siRNA library of 2500 targets known to be involved in protein homeostasis. In parallel, we generated a high confidence CFTR interactome of F508del-CFTR in the same cell line. Given a list of high-scoring siRNA hits for CFTR rescue of function, and a set of CFTR binding proteins, we sought to connect these datasets through an integrated protein-protein interaction network, and use shortest path analysis to uncover the minimal network structure consistent with both the CFTR interactome and siRNA data. The goal of this approach is to prioritize proteins connecting CFTR with siRNA hits that may act as central “hubs” in cellular processes required for CFTR functional rescue. For each protein in the subgraph, it computes the number of distinct siRNA hits that utilize the protein on its shortest path to CFTR. In order to filter out nonspecific protein hubs, this computation is repeated using a random selection of hits from the original siRNA library. The analysis identified several novel candidates for CFTR rescue of function that could be validated through targeted siRNA screens. In summary, we present here a novel network-based method to integrate functional genomics data (e.g. siRNA screens) with interactomics datasets (e.g. AP-MS, MudPIT), useful for prioritizing novel functional targets and for identifying relevant network modules. It leverages publicly available information on protein-protein interactions and thus is readily applicable to many scenarios where a connection between functional and biochemical data is sought.

1442T

Whole-Exon Haplotype Calling for Clinical Next-Generation Sequencing. *J. Maguire, A. Gibiansky, M. Rasmussen, E.A. Evans, I.S. Haque.* Counsyl, 180 Kimball Ave, South San Franciso, California.

In the last two years, NGS has begun to make significant inroads into the clinic via a set of screening and diagnostic applications such as carrier screening for recessive disease, and testing for heritable cancer risk. Research tools such as BWA, GATK, FreeBayes, and others make up the current best practices in sequence analysis, but it is rarely appreciated that they are built with various implicit design decisions that are highly specific to the research regime and not appropriate in a clinical setting. We describe in detail KCall, our whole-exon (155bp-4kb) diploid haplotype caller, which is based on diploid HMM inference over a De Bruijn graph. We describe its performance in various clinical testing applications, using data generated from our own patient stream of thousands of samples per month. KCall was designed to address the differences between the high-volume clinical and research regimes -- Clinical test designs focus on small genomic regions of ~0.5-1Mb targeted at specific medical questions, rather than whole exome/genome coverage in typical research studies; Clinically relevant variants are often “hard” sites for NGS and must be called correctly in every patient; research protocols often optimize the average genome-wide accuracy rather than local accuracy at specific sites; Clinical protocols are unbounded in number of potential samples and have strict turnaround-time requirements; finally, research protocols are funding-limited in sample count and can batch-process all samples at once. KCall leverages these differences to call whole-exon (155bp-4kb) haplotypes using 100bp reads in 100x depth sequencing data. Additionally, KCall replaces heuristic tuning of variant calling parameters with a hands-off calibration procedure using calibration flowcells totaling >30,000x depth in our target area (4 HiSeq flowcells). This eliminates heuristic tuning by modeling error processes site-by-site down to a frequency of 1/1000. The end result of these changes is a tool that behaves with a consistent level of quality across all targets, requires zero subjective tuning steps, and provides a complete and accurate view into the patient’s DNA.

1443S

JADE: A tool for comparative analysis of spatially smooth genomic data. *J. Morrison, D. Witten, N. Simon.* Biostatistics, University of Washington, Seattle, WA.

Multiple quantitative genomic data types such as methylation proportion and copy-number are now commonly collected as part of studies seeking to investigate the genetic architecture of human disease. These data types are often extremely high-dimensional with thousands of measurements per subject and can possess a spatial structure with respect to genome position. The presence of these patterns leads naturally to treating the entire profile as the explanatory variable of interest when comparing groups rather than using measurements at single sites. This task can be challenging because profiles must be estimated from the data and are more complicated data types than un-smoothed site-level measurements. We have developed a penalized likelihood approach, Joint Adaptive Differential Estimation (JADE) for simultaneously preforming smoothing, estimation, and cross-group comparison of genomic profiles for quantitative data types with spatial patterns. The method is motivated by a model in which groups have a shared average profile over some of the genome but to differ in a few biologically relevant regions. JADE utilizes the similarity between groups to improve accuracy of group profile estimation while allowing for regions in which the profiles are different. We are also able to create adaptive clusterings which vary along the genome for joint comparison of more than two groups. By taking a unified approach we avoid many of the ad-hoc decisions made in multi-step techniques as well as the challenges of multiple-testing correction and window-size and region pre-selection. JADE is implemented through an efficient dual gradient descent algorithm. We have applied it to the task of identifying differentially methylated regions between different cell types available in the ENCODE database of reduced representation bisulfite sequencing experiments. In these experiments methylation level is obtained as a binomial proportion at millions of closely spaces sites. These data have previously been modeled using splines and other smoothing techniques allowing us to compare results using JADE and other methods.

1444M

Identifying causal noncoding variants using tissue-specific gene regulatory networks. *K. Tan[1], X. Ma[1], B. He[2], L. Teng[1].* 1) Internal Medicine, University of Iowa, Iowa City, IA; 2) Interdisciplinary Graduate Program in Genetics, University of Iowa, Iowa City, IA.

Characterization of noncoding variants poses a significant challenge in human genetics. Among the different classes of noncoding regulatory sequences, transcriptional enhancers represent the primary basis for differential gene expression, with many human diseases resulting from altered enhancer action. Here, we term SNPs located in enhancers eSNPs. To predict causal eSNPs, current practice has been to overlap various types of functional genomics data (chromatin accessibility, transcription factor binding, epigenetic modification, gene expression) with individual GWAS SNPs. The rationale for such approaches is that causal SNPs should alter TF binding affinity and consequently gene expression. While biologically intuitive, there are a number shortcomings for current approaches. First, most studies attempt to interpret GWAS results agnostically instead of using cell/tissue types that are relevant to the diseases. Cell type-specific information has strong predictive value, and can overcome certain limitations in population-based association studies (e.g. reduced power for weak alleles and low-frequency variants). Second, current approaches do not take into account the complex interactions among the genes affected by regulatory variants. Molecular networks have been used to improve the inference accuracy of causal coding variants. This potential has not been examined for noncoding variants. We have developed a general computational framework for identifying causal noncoding variants that affect a specific trait. By casting the causal inference problem into a subnetwork identification problem, our approach considers all candidate eSNPs simultaneously, thus increasing the power of the inference. Further, our network-based approach naturally provides a pathway content for the resulting causal eSNPs. Compared to existing methods, our method achieves significant improvement. We have applied our novel method to SNPs documented in the NHGRI Catalog of Published Genome-Wide Association Studies.

1445T

ClinSeK: targeted clinical variant identification from high-throughput sequencing data. *W. Zhou, H. Zhao, Z. Chong, A. Eterovic, K. Shaw, F. Meric-Bernstam, G. Mills, K. Chen.* The University of Texas MD Anderson Cancer Center, Houston, TX.

The current paradigm of clinical sequencing data analysis employs a lengthy discovery approach: aligning reads to the human reference assembly, discovering mutations from base to base and identifying mutations that are likely actionable. Although widely practiced, such a paradigm is not optimal for clinical applications, which demand rapid acquisition of clinically relevant, sensitive, and unambiguous molecular profiles. We developed ClinSeK based on a knowledge-driven inversely-operating paradigm that directly tests well-characterized, clinically-relevant variants from high-throughput sequencing data, without exhaustively aligning and comparing sequencing reads to the human reference genome. We overcome challenges in analyzing repetitive regions and duplicated reads under this new paradigm. Applying ClinSeK to characterize the molecular profile of over 600 deeply sequenced cancer samples indicated that this new approach increases sensitivity in detecting low frequency variants with over 50-fold reduction in processing time than existing approaches that perform sequential alignment and variant calling. ClinSeK can test point mutations, indels and structural variations in single or paired samples. It supports the analysis of both DNA and RNA sequencing data. It can also serve as a quick and independent cross-validation in complement to existing variant discovery pipelines.

1446S

Detection, Characterization, and Biological Analysis of Long Tandem Repeats Detected in Human Genomes Using Nanochannel Technology. *S. Chan, X. Zhou, Z. Dzakula, A. Hastie, H. Cao.* Research and Development, BioNano Genomics, San Diego, CA.

A large portion of the human genome is known to be composed of long tandem repeats that can span several hundred kilobases to multiple megabases. Although the repeat motifs can be sequenced and the amount of repeat material can be approximated by conventional sequencing technologies, it is often difficult or impossible to assemble them into long contigs, so the exact locations and copy numbers of these repeats often remain elusive, especially when the unit length exceeds the read lengths. Without knowing the genomic context of these repeats or the amount of repeat material, it is difficult to attach any biological relevance to them. Thanks to the recent developments in nanochannel technology which allow us to image intact megabase-scale molecules of DNA, repeat regions can be more accurately characterized and put into biological context. Using Irys® technology and novel algorithms designed specifically to investigate long repeat arrays, we were able to find and characterize previously enigmatic repeat regions in the human genome. The genomic positions of these repeats were found by aligning non-repetitive portions of repeat-containing molecules or consensus *de novo* assemblies to the reference genome. The genomic context then provides insight into the biological significance of the repeats. Here, we have found that molecules containing a prominent 5.6-kilobase tandem repeat are derived from the Lipoprotein(a) (Lp(a)) coding region on chromosome 6, which is linked to atherosclerotic diseases such as heart disease and stroke. Furthermore, since the array is contained on single molecules, we can measure allele array lengths, showing that Irys® technology has potential for aiding in quick, accurate, and cost-efficient prognosis of these and other genetic diseases which are influenced by copy number variations.

1447M

novoBreak: comprehensively characterizing somatic structural breakpoints in cancer genomes. *Z. Chong[1], M. Gao[2], J. Ruan[3], X. Fan[1], W. Zhou[1], T. Chen[1], H. Zafar[1], J. Chen[2], G. Mills[4], K. Chen[1].* 1) Department of Bioinformatics & Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX; 2) Department of Experimental Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX; 3) Agricultural Genomes Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China; 4) Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX.

Somatic structural variations (SVs) are a major source of genomic variations in cancer genomes and may play a driving role in cancer genome evolution. However, current algorithms for detecting SVs using Next-Generation Sequencing (NGS) data could not achieve a high sensitivity and specificity. This is mainly due to the challenges in aligning and interpreting short reads spanning SV breakpoints. We have developed a novel algorithm, novoBreak, which minimizes the alignment issues to achieve a high recall rate and a low false discovery rate. novoBreak can characterize almost all types of SV breakpoints including deletions, duplications, inversions, insertions and translocations at base-pair resolution. It first clusters reads around potential breakpoints and then locally assembles the reads associated with each breakpoint into contigs. After aligning the contigs to the reference, it can identify the precise breakpoints and infer the types of SVs. novoBreak performs substantively better than other widely applied solutions, such as BreakDancer, CREST, etc., based both on simulation data and on the whole genome sequencing data in the tumor genome atlas (TCGA) and the 1000 Genomes project. In the recent NCI ICGC-TCGA DREAM 8.5 somatic mutation calling challenge, novoBreak ranks as one of the best-performing methods. The higher sensitivity of novoBreak makes it possible to uncover more novel and rare SVs, which could delineate a more comprehensive landscape of variations in cancer genomes and shed light on cancer genome evolution. The algorithm of novoBreak can be easily extended to germline SV discovery, which could be used to characterize genome evolution in a long-term view.

1448T

SG-ADVISER: CNV annotation pipeline. *G. Erikson, N. Deshpande, A. Torkamani.* Scripps Translational Science Institute, La Jolla, CA.

CNVs have been associated with a diverse array of diseases, especially cancer, autism, schizophrenia, and developmental delay. A large number of methods are available for the annotation and interpretation of SNPs and small indels, yet little is available for interpretation of CNVs. The Scripps Genome Annotation and Distributed Variant Interpretation Server (SG-ADVISER) CNV pipeline aims to fill this gap. The SG-ADVISER CNV pipeline is a web server developed at The Scripps Translational Science Institute for the annotation of CNVs. Annotation execution proceeds in highly parallel fashion and includes classes of variant annotations that are entirely independent of one another, serially dependent annotations whose execution are dependent upon the completion and status of prior annotations, and synthetic annotations that generate new information through the combination of multiple annotation outputs. The annotated file includes details regarding location, impact on the coding portion of genes, allele frequency information including allele frequencies from the Scripps Wellderly cohort, and overlap information with other reference datasets including ClinVar, DGV, known syndromes etc. The CNV pipeline accepts variant files in CNVnator, Complete Genomics, or plain tab delimited file formats. A variant classification is produced (ADVISER score) based on the American College of Medical Genetics (ACMG) scoring guidelines with categories 1-5. Variants of category 1 are most clinically relevant. The performance of ADVISER classification schema was evaluated using data from International Collaboration for Clinical Genomics (ICCG). We annotated both pathogenic and benign variants and the ADVISER classification schema showed accuracy of 89% specificity and 99% sensitivity. To facilitate the interpretation of the SG-ADVISER CNV pipeline output we added new functionality to the existing SG-ADVISER UI. The SG-ADVISER UI is a visualization tool that allows scientists with little or no programming experience to easily and quickly view, manipulate, sort, and filter the SG-ADVISER CNV output file. For example it is possible to sort or filter CNVs based on Known Disease, Coding Impact or Chromosome Position etc. At any point during the process the tool can provide summary statistics, and output the filtered results to a new file.

1449S

Challenges to CNV Detection in the Clinic using Targeted High Throughput Sequencing Data. *S. Sadedin, A. Oshlack.* Murdoch Childrens Research Institute, Parkville, Victoria, Australia.

High Throughput Sequencing (HTS) is rapidly gaining clinical acceptance as a cost effective solution for diagnosis of highly penetrant but genetically heterogeneous diseases. Despite this success, the analysis methods applied are usually limited to detection of point mutations and small insertions and deletions. Detection of larger structural variations, most notably deletions, is often omitted despite such variants being equally deleterious and common causes of many disorders. While many tools have been published in the literature, in practice, few laboratories are implementing them in their clinical sequencing pipelines. We speculate that this is due to a perceived lack of specificity and robustness. In this work we apply a novel simulation method to explore the reasons for this gap using real data spliced from X chromosomes of male samples into female samples to simulate single copy deletions. We find that real world performance of these methods appears to be highly sensitive to a range of factors including: method-specific tuning parameters, exact sequencing technology and methodology, variability in the quality of data, and the number and type of samples sequenced in a batch. We conclude that in order for HTS CNV detection to become clinically accepted, methods must be developed that can work in a highly robust, self-calibrating fashion and that are well tuned to popular sequencing platforms. Along these lines, we show results from our own method, Angel, which is specialized to achieve high accuracy and robustness on the HaloPlex targeted sequencing platform.

1450M

SAAS-CNV: A joint segmentation approach on aggregated and allele specific signals for the identification of somatic copy number alterations with next-generation sequencing data. *Z. Zhang[1,2], K. Hao[1,2].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY.

Motivation: Cancer genomes exhibit profound somatic copy number alterations (SCNAs). Studying tumor SCNAs using massively parallel sequencing provides unprecedented resolution and meanwhile gives rise to new challenges in data analysis, complicated by tumor aneuploidy and heterogeneity as well as normal cell contamination. While the majority of existing methods utilize total sequencing depth alone for SCNA inference, the allele specific signals are undervalued. We proposed a joint segmentation and inference approach using both signal dimensions to address some of the challenges. **Methods:** Our method consists of four major steps: 1) extracting read depth supporting reference and alternative alleles at each SNP/Indel loci and comparing the total read depth and alternative allele proportion between tumor and matched normal sample; 2) performing joint segmentation on the two signal dimensions; 3) correcting the copy number baseline from which the SCNA state is determined; 4) calling copy number variation (CNV) and copy-neutral loss of heterozygosity (LOH) for each segment based on both signal dimensions. The method is applicable to whole exome/genome sequencing (WES/WGS) data in a tumor-control study, and is readily extended to SNP array data and tumor-only setting. **Results:** We applied the method to a data set containing no SCNAs to test the false positive rate, created by pairing sequencing replicates of a single HapMap sample as normal/tumor pairs, as well as a large-scale WGS data set consisting of 88 hepatocellular carcinoma (HCC) samples along with matched normal samples. SNP array data is also available for the 88 samples to serve as benchmark. Compared with representative methods, such as ExomeCNV and CNAnorm, our method demonstrated high and consistent accuracy, scalability to large cancer studies, capability in handling a variety of platforms, and potential in improving the estimation of tumor ploidy and purity. **Availability:** An R package called saasCNV is available at zhangz05.u.hpc.mssm.edu/saasCNV/.

1451T

The new European Variation Archive Resource at EMBL-EBI. *I. Medina, D. Spalding, CY. Gonzalez, G. Saunders, J. Kandasamy, S. Ur-Rehman, V. Kumanduri, I. Lappalainen, J. Paschall.* European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Cambridge, United Kingdom, CB10 1SD.

The European Variation Archive (EVA; www.ebi.ac.uk/eva) is a new 'one-stop-shop' for all freely available variation data, for all types of variation, from all species. EVA will work with partner databases to guarantee free global access to genetic variation data. The EVA data model focuses on archiving and making available the results of large scale variant discovery and re-sequencing efforts, providing granular access to population specific allele frequencies paired with study meta-data. Data submitted to EVA in VCF format is closely linked with supporting BAM alignments where available in the ENA or EGA databases at the EMBL-EBI. EVA currently contains over 1.7 billion submitted variants, from a range of more than 15 large-scale projects including 1000 Genomes (phase 1 and 3), Exome Variant Server, Genome of the Netherlands and UK10K. We also mirror the data stored at both dbSNP and ClinVar permitting our users easy access to this data. As an example of our commitment to non-human species, EVA includes data from the NextGen project that has compiled more than 220 million livestock variants including rich meta-data on sub-strain and environment. With regards structural variation, EVA builds upon the existing Database of Genomic Variation Archive (DGVA; www.ebi.ac.uk/dgva) that, in collaboration with dbVar at NCBI, currently contains in excess of 10 million Supporting Structural Variants (SSVs) in 120 studies, covering a range of 28 species. The web portal of EVA aims to provide a dynamic and visually interactive set of queries and filters based on modern web technologies such as HTML5. Users can browse and explore our study or variation catalogue, visualize variations or search for a gene. This front-end is made modular and scalable through use of a RESTful web service interface to the backend data store, allowing EVA data to be accessible programmatically for a variety of applications such as annotation pipelines. To provide a high-performance and scalable solution EVA is based on NoSQL databases such as MongoDB and is designed to accommodate data derived from tens of thousands of whole genomes. Data mining and visualisation tools allow direct access to variation data at the granularity of a given submitted study. Complex queries can be built based upon entry points including gene, study, genomic location, variation type and consequence type. All software is released as open source. Please direct questions and submissions to eva-helpdesk@ebi.ac.uk.

1452S

PON-P2, PON-Diso and PON-MMR: tools for prediction of variation pathogenicity. *M. Vihinen, A. Niroula, S. Urolagin.* Department of Experimental Medical Science, Lund University, Lund, Sweden.

Reliable prediction methods are needed to analyze NGS datasets. Due to the large volume of identified variations only computational approaches can handle the datasets. Genomes contain millions of variants and typically over 10 000 of them lead to amino acid substitution. The applied methods have to be both reliable and fast. We have developed a completely new machine learning-based method PON-P2, which fulfills these criteria. It has excellent performance (accuracy 0.87, MCC 0.77) and it can predict very large datasets in reasonable time. It is more reliable and faster than our previous predictor, PON-P [1] and competing methods. PON-P2 is for amino acid substitutions. The method is based on extensive feature selection and training with a large benchmark dataset from VariBench [2]. The method is implemented with random forest. PON-Diso is a method for predicting effects of amino acid substitutions on order/disorder status of proteins [3]. Several proteins contain disordered regions or are completely disordered, i.e. without regularly ordered structure. Changes to the structure in these regions can be related to diseases. We tested the performance of a large number of existing disorder prediction methods, but they were found unsuitable for this task. We developed a novel tool that has a success rate on 70% in cross validation. A related tool was trained for mismatch repair (MMR) gene variations [4]. PON-MMR has accuracy of 0.87. When applied to 758 unclassified variants in InSiGHT database, it could classify 248 cases as pathogenic or benign. InSiGHT variation interpretation committee [5] classified 1370 variants out of 2360 investigated. 46 of those we had predicted to be either benign or pathogenic. 44/46 of them were correct (96%). These tools have been tested with independent benchmark datasets and they show the highest performance currently available. Available at http://structure.bmc.lu.se/ . References: 1. Olatubosun A, Väliaho J, Härkönen J, Thusberg J, Vihinen M. (2012) Hum Mutat 33, 1166-1174. 2. Nair PS, Vihinen M. (2013) Hum Mutat 34, 42-49. 3. Ali HS, Olatubosun A, Vihinen M. (2012) Hum Mutat 33, 642-650. 4. Ali HS, Urolagin S, Gurarslan Ö, Vihinen M. Hum Mutat (in press). 5. Thompson BA et al. (2014) Nat Genet 46, 107-115.

1453M

HLA-Genotyper Prediction of HLA Genotypes from Next Generation Sequencing Data. *J. Farrell[1], G. Jun[1], L.A. Farrer[1], A. DeStefano[2], P. Sebastiani[2].* 1) Biomedical Genetic- Evans 218, Boston Univ Med, Boston, MA; 2) School of Public Health, Boston University, Boston, MA.

Background: HLA genotyping of next generation sequencing data would be useful for testing the association of HLA alleles with adverse drug reactions, auto-immune diseases, infectious diseases and age-related diseases. **Methods:** We have developed HLA-Genotyper for the prediction of HLA genotypes from next generation sequencing data. HLA-Genotyper is a software tool which performs 4-digit HLA allele prediction using a novel Naïve Bayes algorithm. The evaluation of the novel algorithm for MHC I and II Classical HLA loci (A, B, C, DRB1, DQA1, and DQB1) was conducted using whole genome, whole exome and RNA-Seq data from 51 European and 50 Yoruba samples from the 1000 Genomes Project. To validate the predictions, the precision and recall of the predicted HLA genotypes were compared to a "gold standard" based on SSOP and SBT HLA typing. **Results:** For 560 HLA alleles predicted in 51 low coverage (3-7x) European samples, the precision was 0.96 with a recall of 0.59. For the 597 HLA alleles predicted in 50 low coverage Yoruba samples, the precision was 0.90 with a recall of 0.49. For the 51 European samples with whole exome sequencing (50-100x coverage), the precision of the HLA genotype predictions increased to 0.97 with a recall of 0.96. For the 50 Yoruba samples with whole exome sequencing, precision increased to 0.90 with a recall of 0.90. In 45 European samples, the predictions of 518 HLA alleles from RNA-Seq data were near complete concordance with the "gold standard" with a precision of 1.0 and recall of 0.99. The RNA-Seq validation results of 453 assayed alleles from 38 Yoruba samples also had very high precision and recall rates of 0.98. **Conclusion:** With modest coverage and read lengths between 75 and 101bp, the HLA-Genotyper software accurately predicted HLA genotypes from next generation sequencing data commonly used for research. The predictions were best from the RNA-Seq data with near concordance with the gold standard and better than PCR-SSOP accuracy. Using HLA-Genotyper, researchers may readily unravel the association of HLA alleles with many diseases from next generation sequencing experiments without the expensive and laborious HLA typing of thousands of subjects in diverse ethnic populations. This will improve researchers ability to understand biological roles of HLA loci in infection, inflammation, auto-immunity, aging, mental illness and adverse drug reactions.

1454T

A Statistical Approach that Simultaneously Perform Variant Calling and Local Haplotyping Based on Phase-Informative Reads. *K. Kojima, N. Nariai, T. Mimori, Y. Yamaguchi-Kabata, Y. Sato, Y. Kawai, M. Nagasaki.* Integrative Genomics, Tohoku University, Sendai, Japan.

Next generation sequencing (NGS) technologies enable the detection of novel rare variants in genome wide scale. Haplotype phasing of these variants is important for rare variant association studies because variants are often grouped into exon or gene level. Effects of multiple variants and phase-dependent interactions such as compound heterozygosity and cis-effect are considered for the analysis. In addition, phased haplotype information is required for estimating population genetic parameters by considering a coalescent tree or an ancestral recombination graph based inference. In the process of haplotype phasing, various approaches have been proposed by considering the linkage disequilibrium between variant sites. These types of approaches can provide accurate phasing results for common variants, but their accuracies for low-frequency variants or variants around recombination hotspots tend to be low. The other type of approach uses NGS reads spanning multiple heterozygous variant positions (called phase-informative reads). Although this approach is promising for phasing low-frequency variants, the applicable length of the estimated haplotypes is limited due to the current length of NGS reads (around hundred bases). However, since length of sequence reads is growing rapidly, the rate of heterozygous sites phased by phase-informative reads is expected to increase. The quality of estimated variants is crucial in haplotype phasing based on NGS reads, while accurate variant calling is still challenging due to errors on sequencing and read mapping. Since some of such errors can be corrected by considering haplotype phasing, simultaneous estimation of variants and haplotypes is important. Here, we introduce a statistically unified approach for variant calling and haplotype phasing named HapMonster, in which haplotype phasing information from phase-informative reads is used for improving the accuracy of variant calling and the improved variant calls are used for more accurate haplotype phasing. From the comparison with other existing methods on simulation and real sequencing data, we confirm the effectiveness of HapMonster in both variant calling and haplotype phasing.

1455S

Kragle: a new local *de novo* assembler and genotype caller for short tandem repeats and other complex variations. *K. Konvicka, K.B. Jacobs, K. Nykamp, N. Patil, M. Cargill, C. Kautzer, S.A. McCalmon.* Invitae Corporation, San Francisco, CA.

BACKGROUND: Structural variation in the genome can have a profound effect on gene function. In particular, short tandem repeat (STR) expansions are implicated in a wide array of genetic disorders, such as Huntington Disease. We developed software able to call genotypes of complex variations from next generation sequencing (NGS) with short paired-end reads that current software packages fail to assemble and call. These failures have two causes: 1) most *de novo* sequencing algorithms utilize De Bruijn graphs and are not well-suited to assembling low-complexity sequences due to k-mer limitation, and 2) *de novo* sequencing algorithms often utilize sequence error correction techniques that remove information from low-complexity sequences and interfere with STR sequence assembly. **METHODS**: We developed Kragle, software that combines and adapts three known algorithms: paired-end aware read recruitment, consensus-overlap *de novo* sequence assembly and a statistical framework for genotypes. These components were combined to take full advantage of the reads and assemble low-complexity sequences with repetitive content as long as the read length. The statistical framework then allows us to call diploid genotypes from the assembled haplotype sequences and assign confidence based on alternative genotype hypotheses. **RESULTS**: We processed and analyzed 31 positive control samples representing 11 distinct STR-related loci and inherited conditions in technical duplicate. We used Kragle to genotype the samples at 9 of these 11 loci (2 loci failed PCR) and at 10 additional STR loci. The resulting matrix of 31 samples x 19 genotypes showed complete agreement with the reported genotypes, giving only one false positive result (in 1 of 2 replicates). This translates to 100% sensitivity and 99.2% specificity. The reproducibility of genotype calls between technical replicates was 97% when alleles were allowed to differ by +/-1 repeat. These genotyped loci include STR's in the following genes: *AR*, *ATN1*, *ATXN1*, *ATNX2*, *ATXN3*, *ATXN7*, *DMPK*, *FXN*, and *HTT*. In another experiment we were able to call a heterozygous large deletion in *BRCA1* gene, and confirm it with PCR. **CONCLUSIONS**: The above experiments confirmed that Kragle is capable of assembling and genotyping STR sequences up to the length of an NGS read, and assembling junctions resulting from deletions, insertions and chromosomal rearrangements. This will enable clinical diagnosis of many genetic conditions using short-read NGS data.

1456M

Pipeline and Varant annotation tool for identifying causal variants in inherited rare disorders. *K. Kundu[1], S. Rana[1], A. Chellappan[1], U. Sunderam[1], J.M. Puck[2], S.E. Brenner[3], R. Srinivasan[1].* 1) Innovation Labs, Tata Consultancy Services, Berkeley, CA; 2) Department of Pediatrics, University of California, San Francisco, Box 0519, CA, USA; 3) University of California, Berkeley, CA 94720, USA.

We have developed a pipeline for the analysis of genomic variant data, whose distinctive features enabled solving numerous clinical cases related to SCID (Severe Combined Immunodeficiency) and related diseases. The first several steps of the pipeline employ standard tools for mapping and processing of the mapped data in preparation for variant calling, but integrate three different carefully-tuned callers, to yield high quality sets of variants. Quality metrics for mapping, gene coverage, and called variants are generated throughout the run to ensure confidence in downstream analyses. Our Varant tool (available under an open source license) provides extensive variant annotations, such as genes and transcripts affected by the variant, the type of effect, possible functional consequences including pathways that may be affected by the variant, known disease associations from a variety of data sources such as ClinVar, OMIM, GAD, the mouse phenotype database. We have carefully benchmarked Varant to ensure that it not only has all the features present in other widespread tools, such as Annovar and snpEff, but also overcomes the errors made by other tools and is more liberally licensed. For cases where pedigree or phasing information is available the Varant tool classifies variants as following a compound heterozygous model, *de novo*, uniploidal disomy, or X-linked recessive with potential clinical importance. The annotations are combined with predictions from programs such as ToppGene and Endeavour to further prioritize short-listed variants.

Our pipeline was able to identify likely causal variants in several cases where routine protocols would have been expected to fail. In one example of a patient with abnormal T-cell receptor excision circles (TREC) we were able to identify and later confirm mutations in the *NBN* (nibrin) gene that led to a diagnosis of Nijmegen breakage syndrome. In another example, literature-based haploinsufficieny annotations of the *BCL11B* gene allowed us to propose a role for a heterozygous *de novo* variant in this gene in the case of a child with low T-cell counts and absent corpus callosum. Finally, we identified causal compound heterozygous variants in the *PRKDC* gene in a patient with SCID; other tools missed a key variant because the reference genome incorporates a frameshift mutation.

1457T

SVSI: A Fast and Powerful Set-Valued System Identification Approach to Identifying Rare Variants in Sequencing Studies for Ordered Categorical Traits. *W. Bi[1], G. Kang[2], Y. Zhao[1], Y. Cui[3], S. Yan[4], Y. Li[4,5], C. Hartford[6], W. Leung[6,7], J. Zhang[1].* 1) Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences; 2) Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, Tennessee; 3) Department of Statistics and Probability, Michigan State University, East Lansing, Michigan; 4) Department of Genetics, Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 5) Department of Computer Science, University of North Carolina, Chapel Hill, NC; 6) Department of Bone Marrow Transplantation and Cellular Therapy, St. Jude Children'sResearch Hospital, Memphis, Tennessee; 7) Department of Pediatrics, University of Tennessee Health Science Center, Memphis, Tennessee.

For phenotype-genotype association studies that involve a phenotype with ordered multiple response categories, we usually either regroup multiple categories of the phenotype into two categories ("cases" and "controls") and then apply the standard logistic regression(LG) model, or apply non-parametric method of spearman rank correlation or parametric method of ordered logistic (oLG) regression model which accounts for the ordinal nature of the phenotype. However, these approaches may lose statistical power or may not control type I error rate if the underlying genetic variants are rare or sample size is limited.Here we propose a set-valued (SV) system model, which assumes that the underlying continuous phenotype follows a normal distribution, to identify genetic variants associated with an ordinal categorical phenotype. We couple this model with a set-valued system identificationmethod to identify all the key underlying system parameters. Simulation studies show that SVwell controlled the Type I error rate. In the comparison among LG, SV and oLG methods, LG had significantly lower power than SV and oLG due to disregarding of the ordinal nature of the phenotype, and SV had similar or higher power than oLG. Additionally, the SV association parameter estimate was 2.7-28.7 fold less variable compared to the oLG association parameter estimate. Less variability in the association parameter estimate translates to greater power and robustness across the spectrum of minor allele frequencies. These advantages are most pronounced for rare variants or even common variants when sample size is small. For instance, in a simulation with data generated from an additive oLG model with odds ratio of 7.4 for a phenotype with three categories, a single nucleotide polymorphism with minor allele frequency of 0.75% and sample size of 999 (333 per category), the power of SV, oLG and LG models were 70%, 40% and <1%,respectively, at a significance level of 10-6.When applied to one real data, the testing p-values forSV are smaller than those forLG. Thus, SV can be a competitive alternative to LG or oLG in genetic association studies for ordered categorical phenotype.

1458S

A comprehensive empirical evaluation of linear mixed models for GWAS. *D. Heckerman[1], C. Widmer[1], O. Weissbrod[2], N. Fusi[1], C. Kadie[3], R. Davidson[3], J. Listgarten[1], C. Lippert[1].* 1) Microsoft Research, Los Angeles, CA; 2) Computer Science Department, Technion, Israel; 3) Microsoft Research, Redmond, WA.

The linear mixed model (LMM) has recently emerged as the model of choice to correct for confounding structure, including population structure and family relatedness, in GWAS. At their core, mixed models rely on the estimation of a genetic similarity matrix (GSM), which encodes the pairwise similarity between every two individuals in a cohort. These similarities are estimated from SNPs or other genetic variants. Traditionally, the GSM for an LMM has been estimated from all available variants. Recently, however, it has been proposed that only a subset of variants should be used. The motivation behind this suggestion is based on a mathematical equivalence between the LMM and linear regression. In particular, an LMM is equivalent to a form of linear regression in which the variants that determine the GSM in the LMM view are covariates in the linear-regression view. Thus, taking the linear-regression view, one should only estimate the GSM using variants that are relevant to the phenotype. When all variants have an influence on the phenotype, this understanding leads to the traditional approach of including of all variants in the GSM. When some variants have absolutely no influence on the phenotype, then they should be omitted. But what happens when some variants are only minimally relevant, either having an extremely weak causal effect on the phenotype or an extremely weak association with the phenotype through confounding factors? We investigated variant selection with an empirical study of synthetic data sets spanning a wide range of population structure, family structure, and polygenicity, measuring performance in terms of control of type I error and power. Generally, we find that, when population and/or family structure is present, variant selection alone does not perform well. In particular, when population structure is present, the inclusion of principal components (PCs) as covariates to the model improves performance. When family structure is present, variant selection improves performance only when it is used to create a second GSM in addition to a GSM based on all variants, yielding a dual GSM. When population structure and family structure are both present, a combination of PC covariates and dual GSM is best. Finally, we find that, when population and/or family structure is present, use of variant selection is only beneficial when there are a relatively small number of causal SNPs with relatively large effect sizes.

1459M

Accelerating curation of the catalog of GWAS by automatic text mining. *C. Hsu.* UC San Diego, La Jolla, CA.

The Catalog of GWAS is an important resource containing published association between SNPs and phenotypes identified by Genome-Wide Association Studies (GWAS), a well-defined study approach. GWAS studies have been successfully producing discovery and replication of many new disease loci. The number of GWAS is growing rapidly. There is a need for a database that allows researchers to easily query and search for previous results. Such database has been created and maintained by the National Human Genome Research Institute (NHGRI), called "A Catalog of Published Genome-Wide Association Studies" (Catalog of GWAS). The catalog has led to interesting characterization of previous results in GWAS and NHGRI has continued to update and curate the catalog regularly. However, curation of the catalog is current performed by expert curators. Though this will ensure the quality but new publications in GWAS really outpace any human curation team can possibly handle. The goal of the project GWASTool is to develop a new tool to automatically extract the information from research articles for the curation of the catalog of GWAS. Our proposal is to use the curated data currently available from NHGRI as the training examples and apply novel machine-learning algorithms to train information extractors to allow accurate automatic extraction. Machine learning is particularly suitable for this project because a sufficiently large set of training examples is avail-able from the Catalog of GWAS. The idea is to match the records to the text of the article to mark where they were mentioned and should be extracted from. Moreover, data fields for GWAS studies are relatively well defined, unlike other biological text mining tasks, such as extracting protein-protein interaction. The information extractors to be developed will consist of a tagging module to tag the information in full-text of GWAS research articles and a link module to link the tagged values into entries. A special tagger, based on template matching, will be developed to tag and link the information given in tables. We will demonstrate how the automatic information extraction tool works through a Web-based online tool and present remarkable statistics of the GWAS studies from the extracted data.

1460T

GACT: A Tool for Predicting and Converting Genome Build and Allele Definition during Imputation and Meta-analysis of SNP Genotype Data. *A. Sulovari[1,2], D. Li[1,3,4].* 1) Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, VT; 2) Cell, Molecular and Biomedical Sciences Graduate Program, University of Vermont, Burlington, VT; 3) Department of Computer Science, University of Vermont, Burlington, VT; 4) Neuroscience, Behavior and Health Initiative, University of Vermont, Burlington, VT.

Meta-analysis is one of the commonly used methods in genetic studies of complex diseases as it can increase the statistical power to identify new disease-associated variants by combining genotypes from multiple GWAS. This approach requires same allele definition and genome build among individual studies. Similarly, genotype imputation, commonly used prior to meta-analysis, requires the same consistency. However, the genotypes from various GWAS are generated using different genotyping platforms, arrays or SNP-calling approaches, resulting in use of different genome builds and allele definitions. Incorrect assumptions regarding allele definition among combined GWAS will lead to a large portion of discarded genotypes or incorrect association findings. We have developed a tool that predicts and inter-converts between any of the common SNP allele definitions and the major genome builds. In addition, we measured the effect of several factors on imputation quality, and our results indicated that inclusion of singletons in the reference had detrimental effects while ambiguous SNPs had no measurable effect. Unexpectedly, exclusion of genotypes with missing rate > 0.001 (40% of our study SNPs) showed no significant decrease of imputation quality (even significantly higher when compared to the imputation with singletons in the reference), especially for rare SNPs. Based on our GWAS data, an approximate number of 600 thousand well-typed SNPs are likely to be sufficient for high quality genome-wide imputation of rare SNPs (high quality assayed SNPs may compensate for low true-genotype density). In conclusion, we have developed a new, powerful, and user-friendly tool with both command-line and interactive online versions that can accurately predict, and convert between any of the common allele definitions and between genome builds for genome-wide meta-analysis and imputation of genotypes from SNP-arrays or deep-sequencing. Availability: GACT (http://www.uvm.edu/genomics/software/gact).

1461S

Applying compressed sensing to genome-wide association studies. *S. Vattikuti[1], J.J. Lee[2,4], C.C. Chang[3,4], S.D.H. Hsu[4,5], C.C. Chow[1].* 1) Mathematical Biology Section, Laboratory of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD; 2) Department of Psychology, University of Minnesota Twin Cities, 75 East River Parkway, Minneapolis, MN 55455, USA; 3) BGI Hong Kong, 16 Dai Fu Street, Tai Po Industrial Estate, Tai Po, Hong Kong; 4) Cognitive Genomics Lab, BGI Shenzhen, Yantian District, Shenzhen, China; 5) Office of the Vice President for Research and Graduate Studies, Michigan State University, 426 Auditorium Road, East Lansing, MI 48824, USA.

The aim of a genome-wide association study (GWAS) is to isolate DNA markers for variants affecting phenotypes of interest. This is constrained by the fact that the number of markers often far exceeds the number of samples. Compressed sensing (CS) is a body of theory regarding signal recovery when the number of predictor variables (i.e., genotyped markers) exceeds the sample size [1]. Its applicability to GWAS has not been investigated. Using CS theory, we show that all markers with nonzero coefficients can be identified (selected) using an efficient algorithm, provided that they are sufficiently few in number (sparse) relative to sample size. For heritability equal to one ($h^2 = 1$), there is a sharp phase transition from poor performance to complete selection as the sample size is increased. For heritability below one, complete selection still occurs, but the transition is smoothed. We find for $h^2 \sim 0.5$ that a sample size of approximately thirty times the number of markers with nonzero coefficients is sufficient for full selection. This boundary is only weakly dependent on the number of genotyped markers. Practical measures of signal recovery are robust to linkage disequilibrium between a true causal variant and markers residing in the same genomic region. Given a limited sample size, it is possible to discover a phase transition by increasing the penalization; in this case a subset of the support may be recovered. Applying this approach to the GWAS analysis of height and using approximately 12,000 subjects, we show that 70-100% of the 20 selected markers are strongly correlated with height-associated markers identified by the GIANT Consortium [2].

1. Candes, E.J., Plan, Y.: A probabilistic and RIPless theory of compressed sensing. IEEE Trans Inform Theory 57(11), 7235-7254 (2011)

2. Vattikuti, S., Lee, J.J., Chang, C.C., Hsu, S., Chow, C.C.: Application of compressed sensing to genome-wide association studies. GigaScience (in press).

1462M

Cross-Phenotype Analysis of GWAS (CPAG): A powerful tool for detecting shared genetic architecture among human traits and underlying shared pathways. *L.Y. Wang[1], D.C. Ko[1,2,3].* 1) Duke University, Department of Molecular Genetics & Microbiology, Durham, NC; 2) Department of Medicine, Duke University, Durham, NC; 3) Center for Human Genome Variation, Duke University, Durham, NC.

Recent large-scale meta-analyses of GWAS have demonstrated that cross-phenotypic associations and pleiotropy are abundant among GWAS hits for various diseases and traits. However, the degree of disease similarity as well as the underlying genetic basis driving such connections remains poorly understood. By integrating GWAS results, we developed a novel approach to estimate the degree of disease similarity, identify informative disease clusters, and highlight the potential underpinning pathways. A probability-based method was used to estimate similarity of traits by correcting unobserved shared genetic components and is therefore less sensitive to small and/or uneven shared similarities. Comparing nearly a dozen similarity methods, we determined which was most optimal and robust. Unlike previous methods such as polygenic score, linear mixed-effect model and network model, our approach is based on analyses of summary statistics without requiring more genotype/phenotype data. Using data from the NHGRI GWAS Catalog, we constructed human disease trees with more than 600 traits by using similarity matrices based on either SNPs, SNPs corrected for linkage disequilibrium, and implicated genes among pairwise diseases. For pairwise comparisons that showed significant similarity, we implemented gene set enrichment analysis to identify possible associated pathways. Our method categorized related diseases into disease clusters well consistent with pre-defined groups but also revealed novel and unexpected disease connections. We validated the cross-phenotype analysis of GWAS (CPAG) method computationally using two different approaches by comparing the CPAG generated categorization to pre-defined categories. All of this functionality for disease similarity, visualizing clustering, and pathway analysis has been implemented into a flexible stand-alone software program, named CPAG. CPAG also accepts user-defined datasets to predict potential disease connections and shared pathways. Our approach will become increasingly powerful with the ever-expanding identification of novel causal genetic variation attributed to different traits, and lead to further understanding of the genetic architecture of complex traits and additional insights into the pathophysiology of diseases. Indeed, we are currently experimentally testing novel human trait connections discovered from our analysis using zebrafish.

1463T

GeneHeal: An Online Resource for Mutations and Associated Phenotypes in Deafness. *A. Mehta[1], G. Bademci[1], A. Lakum[1], G. Beecham[1], S. Blanton[1], X. liu[2], M. Tekin[1].* 1) Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL 33136, USA; 2) Department of Otolaryngology, Miller School of Medicine, University of Miami, Miami, FL 33136, USA.

The success of genomic medicine depends on many factors, including the availability of well characterized genotype-phenotype correlations. We have developed a web-based resource GeneHeal, which links auditory phenotypic information of the deaf individuals enrolled in our study to the respective pathogenic genomic variants. To display the relationship between deafness gene variants (genotypes) and their respective phenotype, diagnostic categories for each participant (based on age at onset, progression and severity, including average hearing levels, audiogram shape, and associated findings) are stored in the Hearing Loss (HL) Phenotype Database. Causative mutations in deafness genes for the enrolled individuals are stored in the HL Genotype Database. Both databases are accessible through the Miami Genetics of HL Resource (GeneHeal) at http://hihg.med.miami.edu/deafness/geneheal and are open for researchers who would like to contribute their data. It provides search capabilities to end users making it an efficient data repository and query tool for the researchers around the world. The front end of the website is an HTML form running on PHP scripts. Predefined scripts are provided to query the database and relay information seamlessly. Both variant and phenotype oriented queries are available. An output table is generated that contains all the information listed in both databases for the individuals identified with the query. Both databases are updated on a regular basis. Quality assurance and data integrity checks are performed on a regular basis by HIHG personnel as well as daily data backups. Our tool links molecular variation with its detailed auditory phenotypic features for a comprehensive approach to the diagnosis, management, and risk assessment of individuals with HL.

1464S

A fast and accurate method for detection of IBD shared haplotypes in genomewide SNP data. *D. Bjelland[1], U. Lingala[1], N. Lapinski[1], M. Jones[2], M. Keller[1,2].* 1) Institute for Behavioral Genetics University of Colorado, Boulder, CO; 2) Department of Psychology University of Colorado, Boulder, CO.

Several programs have been developed to detect shared identical-by-descent haplotypes between individuals using SNP data. For detecting short shared haplotypes (e.g., 2-3 cM) in large datasets (e.g., > 10,000 individuals), the performance of these programs suffers either in terms of accuracy or in computational time. With this in mind, we developed a program, FISHR, that utilizes initial shared haplotype output from GERMLINE, quickly post-processes this data, and more accurately discovers the true shared haplotypes. FISHR was compared to GERMLINE and BEAGLE Refined IBD by simulating 15 megabases of genotypic data for 8,000 individuals with realistic levels of SNP and phasing errors. GERMLINE was the fastest program analyzed, with times of 9.0, 34.7, 152.3, and 601.0 seconds for samples of 1,000, 2,000, 4,000 and 8,000 individuals, respectively. Run times on equivalent sample sizes were 13.0, 48.3, 189.7, and 753.0 seconds for FISHR (including the initial GERMLINE runs) and 400.3, 2064.0, 29,840.0, and 921,602.7 seconds (or 10.6 days) for BEAGLE Refined IBD. When discovering shared haplotypes with lengths between 2 and 3 cM, FISHR consistently had ~ 8% higher positive predictive value for matched sensitivity when compared to GERMLINE and ~18% higher positive predictive value for matched sensitivity when compared to BEAGLE Refined IBD. Similar results occurred when analyzing shared haplotypes between 3 and 4 cM. Thus, FISHR is more accurate than both GERMLINE and Refined IBD, and is fast enough to be usable for detection of identical-by-descent shared haplotypes in very large genome-wide datasets.

1465M

HapFerret: A flexible haplotype inference program, determining blocks of haplotype inferability. *G. Nelson, S. Limou, C. Winkler, M. Carrington.* Basic Research Program, Leidos Biomedical Corp/FNLCR, Frederick, MD.

The Expectation-Maximization (EM) algorithm for haplotype inference has been superseded for accuracy by newer programs, e.g. PHASE and SHAPEIT, but remains useful for rapid analysis. We have developed an EM implementation, "HapFerret", that is characterized by flexibility and ease of use, notable its use of a natural format for input genotypes and output haplotypes. Genotypes may be input as lines of comma separated alleles, where the alleles may be any alphanumeric; i.e. for HLA data, for the loci MogCA—DQA1—rs9273349—DQB1—rs1894407—TAP2—TAP1, a hypothetical genotype input line: 9,12 101,301 C,T 302,503 A,A 3.2,4.2 3.1,3.1 might yield an output line of haplotype calls 12-301—C—302—A—4.2-3.1; 9-101—T—503—A—3.2—3.1. Input can also be by one locus genotype per line, with the same allele format. Critically for the inference of HLA haplotypes, an arbitrary number of alleles are allowed at each locus.

A special feature is the ability to infer haplotypes for all subsequences with a specified range of lengths. This has the particular purpose of finding blocks in which haplotypes are inferred with relatively low ambiguity. The program measures the ambiguity of haplotype inference with an entropy measure, where a completely unambiguous inference has entropy 0. We propose this measure as an alternate, pragmatic criterion for haplotype blocks, that predicts the effectiveness of inferred haplotypes for association analysis. Since no coalescence calculation is used (being incompatible with the EM algorithm) this inferability is purely a function of the genotype/haplotype combinatorics, and the underlying LD. In general inferability is a function of higher order LD; we conjecture that a necessary and sufficient condition for unambiguous haplotype inference for n loci is that D' = 1 for LD of some order ≤ n. A limitation is that the EM algorithm may give an unambiguous prediction even where the inference is uncertain. To catch such errors, HapFerret contains a bootstrapping procedure; comparing inference between successive bootstrap replicates catches some cases of spurious precision. We show with data with known haplotypes that bootstrapped inference from HapFerret has an accuracy intermediate between standard EM and PHASE. HapFerret may be downloaded from: https://ccrod.cancer.gov/confluence/display/BCGC/BCGC+Software.

1466T

A novel approach to craniofacial gene discovery: *SysCLFT* (Systems tool for Cleft lip/palate gene discovery). *I. Saadi[1], D. Anand[2], D. Djordjevic[3,4], J.W.K. Ho[3,4], S.A. Lachke[2].* 1) Anatomy & Cell Biol, Univ Kansas Med Ctr, Kansas City, KS; 2) Biological Sciences, Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE; 3) Victor Chang Cardiac Research Institute, Sydney, NSW, Australia; 4) Faculty of Medicine, The University of New South Wales, Sydney, NSW, Australia.

Although study of mammalian craniofacial (CF) development has resulted in a detailed functional understanding of several genes associated with orofacial clefts (OFCs), many additional loci remain unsolved. We recently developed an innovative systems approach that is highly effective in identifying genes critical to the development and/or homeostasis of specific tissues. Successful application of this approach identified several novel genes with functional significance in eye tissue and associated defects (Lachke *Science* 2011 331:1571). Here, we expand this approach to develop *SysCLFT* (Systems tool for Cleft lip/palate gene discovery), a new interactive web resource for efficient prioritization of candidate genes associated with OFCs and construction of CF developmental GRN. We analyzed 67 microarray gene expression datasets of various CF tissues, available publicly in FaceBase and NCBI-GEO, using *in silico* subtraction with embryonic whole body (WB) reference datasets. This approach generates systematic gene ranks based on CF-tissue enrichment - instead of simply high expression - and offers a new method of prioritization of candidate genes associated with OFCs. Indeed with WB subtraction, the top 500 highly ranked genes from the palate, frontonasal, mandible or maxillary process datasets are highly enriched in gene ontology (GO) categories for palate development, cell adhesion, cell proliferation, epithelium development and skeletal system development. In contrast, without WB subtraction, *i.e.* data as currently available on FaceBase and NCBI-GEO, the most highly enriched GO categories among the top 500 candidates mainly comprise of housekeeping genes (translation, glycolysis, nucleic acid biosynthetic process, *etc.*). WB-subtraction identifies 41 of 45 (91%) OFC-linked genes at higher expression ranks, compared to only 4 of 45 (9%) genes without WB-subtraction. Major genes with established role in CF development, including *Irf6, Pax7, Pax9, Pbx2, Pbx3, Msx1, Runx2, Satb2, Tbx22*, are highly enriched in WB-subtracted datasets from different facial tissues. Most significantly, we show that the *in silico* subtraction method successfully identifies a majority (85%, *n*=45) of known OFC genes within the top 5 minRank genes in mean chromosomal intervals of 13Mb, each containing 106 genes on average. Thus, *SysCLFT* is a new, highly effective approach to facilitate prediction of genes and prioritization of variants in GWAS and exome-based CF gene discovery efforts.

1467S

Phenolyzer: prioritizing candidate genes from disease/phenotype descriptors. *H. YANG, K. WANG.* Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA.

Whole-genome/exome sequencing and copy number variation studies can generate many candidate genes for specific diseases or phenotypes. However, to best prioritize these candidate genes, traditional approaches typically do not consider the prior knowledge on associations between genes and specific phenotypes. To address this problem, we developed an integrative tool named Phenolyzer, which takes as input a discrete list of phenotype terms and generate a list of candidate genes weighted by the chance of being associated with the phenotype, even in the absence of any genotype data. Phenolyzer works by following a biologist's logic - 1) search databases in CTD's (Comparative Toxicogenomics Database) disease vocabulary and Disease Ontology for a certain disease/phenotype term, interpret the term into multiple specific disease names, find all the associated genes and related information in OMIM (Online Mendelian Inheritance in Man), Orphanet (a journal for rare disease), NCBI's ClinVar, GeneReviews (an expert-authored, peer-reviewed disease descriptions) and GWAS (Genome Wide Association Studies) databases, then generate the seed gene set with conditional probability as scores, 2) grow the seed gene set in the HPRD (Human Protein Reference Database) protein interaction, NCBI's Biosystem, HGNC (HUGO Gene Nomenclature Committee) gene family, and HTRI (Human Transcriptional Regulation Interactions) databases and retrieve an augmented gene set, 3) integrating all the information to score genes. The input can be one or several disease/phenotype terms, while users can optionally supply a gene list or genomic region to further trim down candidate genes. Additionally, Phenolyzer generates an intuitive visualization of gene-phenotype network and gene-gene network on the fly. Furthermore, combined with ANNOVAR, users can prioritize disease variants from whole-genome/exome sequencing data, significantly expediting discovery of disease causal genes. The efficiency and accuracy of Phenolyzer were validated by extensive benchmarking studies using phenotype terms associated with several Mendelian diseases, and also with test datasets on complex diseases. The great performance of Phenolyzer was also compared with other similar tools like Phevor and PosMed. In addition to a command line software tool, we implemented Phenolyzer as a user-friendly web server to facilitate easy access for users without informatics skills. The server is available at http://phenolyzer.usc.edu.

1468M

Determining the Number of Contributors using Forensically Relevant STRs: Effects of Template Mass and Complexity on the Ability to Correctly Identify the Number of Contributors. *L. Alfonse[1], H. Swaminathan[2], D.S. Lun[2], M. Medard[3], C.M. Grgicak[1].* 1) Biomedical Forensic Sciences, Boston University School of Medicine, Boston, MA; 2) Department of Computer Science, Rutgers University-Camden, Camden, NJ; 3) Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

The objective of this project was to evaluate NOCIt - a software that can assess the number of contributors (NOC) to DNA samples. The likelihood that n unrelated individuals contributed to a sample is computed via a Monte Carlo process during which genotypes for the n contributors are randomly chosen based on the allele frequencies, and the ratio of the n contributors to the mixture is randomly selected from a uniform distribution. Modeling of baseline noise, reverse and forward stutter proportions, dropout rates, and allele heights is performed. The likelihood of observing the peak heights at each locus given the genotypes of the contributors, the mixture ratio, and the template DNA mass is computed using single source calibration samples of known genotypes; likelihood values are 'averaged' between iterations and multiplied across loci to give the overall likelihood given n contributors. The output is presented as the probability distribution over 0 to 5 contributors. The performance of NOCIt was tested on 1-, 2-, 3-, 4- and 5-person mixtures of various contributor ratios, amplified with the AmpFlstr® Identifiler® Plus Amplification kit (29 cycles). Samples were amplified at target masses ranging from 0.25 to 0.008 ng and analyzed at 1 RFU. Artifacts such as -A, dissociated dye, bleed through, and spikes were manually removed. Allele frequencies from the Caucasian population provided in the AmpFlstr® Identifiler® Plus manual were utilized. The accuracy of NOCIt, computed by comparing the software's output to the known NOC, was dependent upon 1) the true NOC in the sample and 2) the total mass of DNA in the sample. For example, the accuracy decreased from 100% to 20% for the 1- to 5-person mixtures, respectively. However, when sufficient levels of DNA were amplified, i.e. 0.25 ng, the accuracy was 100% for 1- to 4-person mixtures and 67% for the mixtures containing DNA from 5 persons. The minimum from any one contributor required to correctly estimate the NOC was ~10 cells. These preliminary results suggest that NOCIt is a valuable analysis tool and results in high accuracy rates for samples with sufficient DNA quantities. Furthermore, these data indicate that samples containing low-target quantities may need to be interpreted using multiple or different assumptions on the number of contributors. Future studies will test NOCIt's accuracy with an additional 360 samples and provide data regarding the number of samples needed to calibrate NOCIt.

1469T

Detection of autozygous segments in exomes of inbred individuals. *M. Vigeland[1], K.S. Gjøtterud[2].* 1) Dept of Medical Genetics, Oslo University Hospital, Oslo, Norway; 2) Dept of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway.

We present AutEx, a new algorithm for detecting autozygous segments directly from whole-exome variant files. Our approach is based on the hidden Markov model developed by Leutenegger et al. (2003), taking as input the level of inbreeding, allele frequencies, variant distances and sequencing error probabilities. Using both simulated data and real inbred exomes we compare our method to existing software detecting runs of homozygosity, including PLINK. Overall AutEx performs better, in particular when the level of inbreeding is accurately known. The AutEx algorithm is implemented as part of FILTUS, a lightweight GUI program aimed at non-bioinformaticians, providing a framework for filtering, exploring and statistical analysis of exome data. Combined with these features, AutEx becomes an effective and easy-to-use tool for narrowing down the search for causal variants in exomes of inbred individuals with recessive disorders.

1470S

A systems biology approach for enriching genetic association studies of metabolite profiles with pathway knowledge. *K. Willems van Dijk[1], H.K. Dharuri[1], P. Henneman[2], A. Demirkan[3], A. Isaacs[3], C.M. van Duijn[3], P.A.C. 't Hoen[1], J.B. van Klinken[1].* 1) Human Genetics, Leiden University Medical Center, Leiden, Zuid Holland, Netherlands; 2) Clinical Genetics, Academic Medical Center, Amsterdam, Netherlands; 3) Genetic Epidemiology Unit, Epidemiology, Erasmus Medical Center, Rotterdam, Netherlands.

Plasma metabolite levels have been proposed to reflect the interaction of the genome with the environment and may thus provide insight into the etiology of complex metabolic disorders. Genome-Wide Association Studies (GWAS) on plasma metabolite levels have led to the discovery of many novel loci, but they have been relatively unsuccessful in providing insight into the etiology of complex metabolic disorders. Here, we propose Elementary Flux Mode (EFM) analysis as an approach to assist in the functional interpretation of GWAS results on metabolomics data and to explain the mechanistic link between genotype and metabotype. EFM analysis consists of enumerating the complete set of steady state fluxes that are possible in a metabolic network, thus providing a comprehensive picture of the network's functional capacity. We assessed the potential of EFM analysis to explain a set of genetically determined metabotypes (GDMs) concerning amino acid metabolism in humans that were reported in the OMIM and GWAS catalogue database. We performed EFM analysis on a condensed version of the genome-scale stoichiometric model (GSMM) of the hepatocyte developed by Gille et al. (2010). The resulting model yielded 8.5•10^9 EFMs that were involved in the degradation, synthesis or conversion of amino acids. Subsequently, for each reaction and amino acid, we determined the essentiality of that reaction in the metabolism of the amino acid by counting the number of EFMs that both contained the reaction and were involved in the conversion of the amino acid. Comparing these results with the selected GDMs, we found that the known genotype-metabotype relationships could be predicted with a high degree of accuracy (ROC curve AUC = 0.93). In contrast, predictions that were made based on co-occurrence of the gene and metabolite in KEGG pathway gene sets or on the shortest distance between the gene and the metabolite in the metabolic network had a low degree of specificity and were less accurate (AUC = 0.79 and 0.74 resp.). In conclusion, we present a novel strategy for analysing metabolomics data in a genome-wide setting using a Systems Biology approach. A workflow has been developed to 1) compute the EFMs in a GSMM, 2) map GWAS results to the EFMs, and 3) visualise enriched EFMs. Our approach integrates genetic association analysis with mathematical models of human metabolism and has the potential to provide new mechanistic insights into the causes of metabolic disease.

1471M

Integration of transcriptome, bioinformatics and model organism studies to gain insight into microRNA function. *S. Banfi[1, 2], I. Conte[1], S. Merella[3], J. Garcia-Mantega[3], C. Migliore[4], D. Lazarevic[3], S. Carrella[1], R. Avellino[1], N. Davidson[1], E. Stupka[3].* 1) Telethon Institute of Genetics and Medicine (TIGEM), Naples, Italy; 2) Department of Biochemistry, Biophysics and General Pathology, Second University of Naples, Naples; 3) Center For Translational Genomics and Bioinformatics, San Raffaele Scientific Institute, Milan, Italy; 4) CBM Scrl, c/o Area Science Park, Trieste, Italy.

Vertebrate organogenesis is critically sensitive to gene dosage and even subtle variations in the expression levels of key genes may result in a variety of tissue anomalies. MicroRNAs (miRNAs) are fundamental regulators of gene expression and their role in vertebrate tissue patterning is just beginning to be elucidated. To gain further insight into this issue, we analysed the transcriptomic consequences of manipulating the expression of miR-204 in the Medaka fish model system. We used RNA-Seq and an innovative bioinformatics approach, which combines conventional differential expression analysis with the behaviour expected by miR-204 targets after its overexpression and knockdown. With this approach combined with a correlative analysis of the putative targets, we identified a wider set of mir-204 target genes belonging to different pathways. Together, these approaches confirmed that miR-204 has a key role in eye development and further highlighted its putative function in neural differentiation processes, including axon guidance as supported by in vivo functional studies. Together, our results demonstrate the advantage of integrating next-generation sequencing and bioinformatics approaches to investigate miRNA biology and provide new important information on the role of miRNAs in the control of axon guidance and more broadly in nervous system development.

1472T

Enhanced statistical methods to detect cross-population heterogeneity at GWAS risk loci. *M. Roytman[1], B. Pasaniuc[1, 2, 3].* 1) Bioinformatics Interdepartmental Program, UCLA; 2) Dept of Human Genetics, David Geffen School of Medicine, UCLA; 3) Dept of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA.

GWAS have identified thousands of disease-associated variants to date, many of which have successfully been used to identify biological pathways for drug targets. However, a large majority of these risk alleles were identified in individuals of European ancestry making their transferability to other ethnic groups uncertain. Here we present methods for quantifying how well the known GWAS risk variants extend to individuals of non-European ancestry. We introduce methods that quantify the difference in effect size of a particular risk variant across two populations and propose a log likelihood ratio test for assessing the presence of heterogeneity at a given risk locus. We investigate the power of our approach at different levels of heterogeneity and show that our methods has high power of detecting realistic differences in effect size, while maintaining well-calibrated statistics, closely following a chi-squared distribution with one degree of freedom under null data simulations (where effect size does not differ between the populations). We extend our approach to admixed populations (e.g. African American) and show how locus-specific can be integrated to detect heterogeneity in the allelic effect across different ancestries.

1473S

Regulatory network constructed from the epigenome of normal cells reveals functional connections between disease genes. *R.F. Lowdon[1], G. Elliott[1], B. Zhang[1], J.B. Cheng[2], S.J. Fisher[3], J.F. Costello[4], T. Wang[1].* 1) Genetics, Washington University School of Medicine, Saint Louis, MO. 63108; 2) Department of Dermatology, University of California San Francisco, CA 94143; 3) Department of Obstetrics, Gynecology & Reproductive Sciences, School of Medicine, University of California San Francisco, San Francisco, CA, USA; 4) Dept. of Neurological Surgery, Helen Diller Family Comprehensive Cancer Center, UC-San Francisco, CA 94143.

It is a daunting but critical challenge to understand the functional connection between human disease genes and the function of disease-associated non-coding variants. Uncovering the connections between these genes and variants are crucial for understanding normal and aberrant regulation of human disease genes. We hypothesized that cell type-specific gene regulatory instructions for a transcription factor and its target genes are encoded in the epigenomic landscape around the genes. To test this hypothesis, we developed a regulatory analysis pipeline to construct gene regulatory networks based on a cell's epigenome. Our regulatory analysis approach has the following components: first, by profiling and contrasting epigenomes of different cell types, we defined cell type-specific regulatory elements. Then, we identified enrichment of binding site motifs of specific transcription factors in these regulatory elements. Next we used custom region-gene association tools to identify candidate genes for each element. We also used co-variation of transcription factor expression and target gene expression to increase the confidence of our prediction. Finally, we queried publicly available TF-target gene interaction data to quantify the connections among our TFs and target genes. This approach allowed us to construct a gene regulatory network of a cell type of interest. We successfully captured many known connections in our reconstructed networks, as well as predicted novel connections. Surprisingly, this approach revealed functional connections across disease genes, predicted potential key genes that are not yet implicated in disease, and explained functional consequences of some non-coding, disease associated variants. In our first test, we constructed a gene regulatory network of the embryonic surface ectoderm using epigenomes of cell types of shared developmental origin. We confirmed co-regulation of six genes that encode hemidesmosomes, a structural complex in the epidermis of the skin. Mutations occur in any one of five of these genes in the inherited blistering skin disease, epidermolysis bullosa. In our second test, we constructed a gene regulatory network of the placenta cytotrophoblast cells. The network revealed genes highly connected to the placental disease pre-eclampsia and predicted a master regulator of these genes. Thus our analysis pipeline was robust across biological systems and connected non-coding regulatory elements in human disease.

1474M

Dynamic changes in RNA modifications (epitranscriptome) localization drives new regulation of cancer cells. *Y. Saletore[1,2,3], T. Fernando[4], A. Melnick[4], L. Cerchietti[4], C. Mason[1,2].* 1) Department of Physiology and Biophysics, Weill Medical College, Cornell University, New York, NY 10065, USA; 2) The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY 10065, USA; 3) Tri-Institutional Training Program in Computational Biology and Medicine, New York, NY 10065, USA; 4) Department of Hematology and Oncology, Weill Cornell Medical College, New York, NY, USA.

The epitranscriptome is a very nascent, exciting field that has revealed a complex middle layer of cellular regulation that is mediated by modified RNA bases. Methylated RNA ImmunoPrecipitation-Seq (MeRIP-Seq) utilizes an antibody specific for the RNA modification methyl-6-adenosine (m^6A) to enrich for RNA fragments that contain the epitope. Combined with next-generation sequencing (NGS), these immunoprecipitated fragments (IP) are sequenced with RNA-Seq controls to identify peaks: genomic regions of interest where putative m^6A sites exist. However, the original MeRIP peak-finding methods (like MeRIPPeR) utilized the non-parametric Fisher's exact test to find peak regions that were statistically significant in enrichment for the IP relative to the control. While this was successful in identifying thousands of peaks in both human and mouse samples, m^6A is known to be a dynamic modification and m^6A sites and methylation levels change in response to cell stimuli. Here we introduce MeRIPPeR 2.0, which uses the negative binomial distribution to better model the high dispersion in the RNA-Seq data and applies a probabilistic graphical model to elucidate peak methylation from the IP efficiency. This not only enables better identification of epitranscriptomic peaks, but also more accurate characterization of how those peaks change between conditions.

We then applied MeRIP-Seq and MeRIPPeR 2.0 to Ly1 diffuse large B cell lymphoma samples to further examine the dynamic nature of m^6A in the context of cancer, using normal cells, proliferating cells, and those cells after heat shock, as well as investigating nuclear and cytosolic fractions separately. Heat shock has previously been demonstrated to have an impact on m^6A sites and also simulates B-cell activation. Previous epitranscriptomic studies have demonstrated an enrichment of peaks near the stop codon, but here, we find that some samples exhibit a striking enrichment in the 5' UTR. In concordance with these results, we identified thousands of m^6A peaks throughout the transcriptome and an enrichment near the stop codon. However, we also discovered a small set of m^6A peaks in the 5' UTR that are more highly regulated more dynamic than those found enriched at the stop codon. This indicates that not only the degree of RNA methylation, but also the coordinates of foci within genes can shift, allowing a completely different regulatory avenue for m^6A that has not been seen before.

1475T

An integrated method to predict functional impact of non-synonymous SNVs in human genome. *M. Wang[1], L. Wei[1,2].* 1) Center for Bioinformatics, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, P.R. China; 2) National Institute of Biological Sciences, Beijing, P.R. China.

Correct interpretation of large scale genome variants functional impact and implications in human disease is essential for modern biomedical research and clinical genetics. In silico methods predicting whether a non-synonymous SNV is deleterious or neutral facilitate the discovery of disease causal mutations and classification for SNVs of unknown clinical significance. Most of these tools are built upon machine learning approaches, which classify the two types of SNVs mainly based on conservation information, sequence features and protein structures. However, correct assessment of disease relevance for a SNV requires informative and comprehensive evaluation, which should take various sources of evidence into consideration. We proposed a method to predict functional impact of non-synonymous SNVs in human genome by integrating multiple sources of evidences. Genetic evidence including co-segregation of a SNP with disease phenotype in a family and population case-control associations provide direct evidence linking SNPs to phenotype. Such heterogeneous factors were integrated in our tool by a unified Bayes model, which gives a combined posterior probability that the SNV is pathogenic given all the evidences. Performance evaluation on benchmark data sets showed the integrated method outperforms those widely employed tools. We also demonstrated its utility in disease causal gene discovery from next-generation whole exome sequencing. This method would facilitate rapid annotation and reliable classification for SNPs discovered by NGS in a large scale.

1476S

Identifying the master regulators of complex autoimmune disease susceptibility in Alopecia Areata with reverse-engineered regulatory networks. *J.C. Chen[1,2], A.M. Christiano[2,3].* 1) Systems Biology, Columbia University, New York, NY; 2) Genetics and Development, Columbia University, New York, NY; 3) Dermatology, Columbia University, New York, NY.

Alopecia areata (AA) is an autoimmune disease affecting the hair follicle in which T-cell infiltrates aberrantly localize to and destroy the follicle, resulting in hair loss. While most research has focused on the immunologic etiology of the disease, little has been done to investigate the contributing genetics of the affected end organ e.g. the skin and hair in AA, yet this is an indispensable driver to the etiology of this autoimmune disease. Patients with autoimmune disease do not uniformly present with symptoms across multiple tissues, indicating that genetic components in the end organ specific to a disease are required for autoimmune response. We have taken a systems biology approach to the study of acquired autoimmunity in AA, and modeled it as a genetic, regulatory disease phenotype using reverse-engineered cellular networks. The goal of this study was to take a global, unbiased approach to understanding the regulation of key molecular programs mediating the AA autoimmune response by microarray profiling of scalp biopsies from 60 individuals across three categories of AA clinical presentation and unaffected controls. Through integrating our network-based analysis, with Gene Set Enrichment analysis and unsupervised clustering, we have identified that varying presentations of AA are mediated by the unique expression of six key regulators, with increased severity in presentation coinciding with the involvement of a higher fraction of these regulators. The manipulation of these master regulators is sufficient to induce cytotoxicity in cultured human dermal papillae cells and keratinocytes. This approach has given us a novel means to study the mechanistic regulation behind the complex gene expression networks associated with aberrant autoimmune response in the context of AA.

1477M

Global profiling of condition specific transcription factor binding with ATAC-seq. *R. Pique-Regi, D. Watza, M. Estill, F. Luca.* Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI.

Deciphering the regulatory sequences which control gene transcription is a critical step in understanding both cellular and condition-specific regulatory programs encoded in the human genome. Transcriptional response is typically regulated by transcription factors (TFs) which are known to bind specific regulatory sequence motifs. Profiling the environmental binding activity of these factors can be quickly accomplished at a genome-wide scale with the recently developed technique ATAC-seq, which utilizes the Tn5 transposase to fragment and tag accessible DNA. When coupled with an advanced computational method such as CENTIPEDE binding models for TFs with known motifs can be generated across the genome. To date, there are no methods that efficiently incorporate the information provided by paired-end sequencing which allows both the identification of the library fragment length as well as the two cleavage locations that generated the fragment.

First, we have extended CENTIPEDE to utilize fragment length information, and we are working on extending the model to include the joint statistics of cleavage pairs. Our results indicate that paired-end sequencing provides a more informative footprint model for ATAC-seq libraries which leads to greater accuracy in predicting TF binding. These results were validated with ChIP-seq data (ENCODE Project) for multiple factors including CTCF, NRSF, NRF-1, and NFkB.

We then assayed TF activity in lymphoblastoid cell-lines (LCLs) across multiple treatments (selenium, copper, retinoic acid and iron) for which we previously determined significant differences in gene expression levels. From our initial sequencing results we were able to resolve 383 actively bound motifs (Zscore > 5) across all conditions. We were also able to characterize 5236 regions that have significantly changed chromatin accessibility (FDR < 10%) in response to both copper and selenium. No major change in accessibility was detected for retinoic acid but condition-specific differential binding was detected in a small subset of motif instances. Ongoing analyses focus in integrating the changes in TF binding together with the transcriptional response. These results demonstrate that ATAC-seq together with an improved footprint model are excellent tools for rapid profiling of transcription binding factor activity to study cellular regulatory response to the environment.

1478T

A comprehensive and highly accurate RNA-Seq pipeline using a hybrid sequencing and algorithm approach. *P.T. Afshar[1], G. Gibeling[2], M. Mohiyuddin[3], R. Sebra[4], A. Sethi[6], L.T. Fang[3], J.C. Mu[1], H. Tilgner[5], N. Bani Asadi[2,3], M.B. Gerstein[6], M. Snyder[5], E. Schadt[4], W.H. Wong[7,8], K.F. Au[9], H.Y.K Lam[3].* 1) Department of Electrical Engineering, Stanford University, Stanford, CA; 2) Department of Engineering, Bina Technologies, Redwood City, CA; 3) Department of Bioinformatics, Bina Technologies, Redwood City, CA; 4) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 6) Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT; 7) Department of Statistics, Stanford University, Stanford, CA; 8) Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA; 9) Department of Internal Medicine, University of Iowa, Iowa City, IA.

Emergence of next generation sequencing has brought a new era in transcriptome analysis with RNA-Seq. Although Second Generation Sequencing (SGS) is a high-throughput and high-fidelity technology, full-length isoform analysis is still a challenging problem due to the relatively short read length of SGS, the imperfect performance of existing tools, and the complexity of alternative splicing. Recent studies have suggested to use Third Generation Sequencing (TGS) to improve isoform detection, but it is missing the essence of the analyses enabled by SGS.

Here we present a comprehensive transcriptome profiling pipeline, the RNA-Seq Cocktail pipeline, integrating both SGS and TGS RNA-seq data along with a number of best-of-breed algorithms. TopHat2, SpliceMap, and STAR are used to align RNA-Seq reads from SGS and detect splice junctions. A novel algorithm we developed is used to merge alignments and junctions from multiple mappers and assign confidence level to the merged junctions. It also uses Cufflinks2 for isoform assembly, Cuffdiff for differential expression analysis, and GATK3 for variant detection. To address the high error rate of current TGS, it uses LSC for error correction followed by IDP for RNA-Seq analysis, including fusion gene detection.

We processed 7 human embryonic stem cell (hESC) datasets from the ENCODE project with our pipeline. The merged junction set from our algorithm has up to 33% more detected junctions compared to the junctions from a single tool. In addition, 94% of junctions reported by all the three tools were validated by the EST dataset compared to 25% validation rate for junctions reported by only one tool. We also compared the assembled isoforms from Cufflinks2 based on ~50K TopHat2 alignments to RefSeq and Ensembl. We observed that a higher percentage of singletons was reported by Cufflinks2 (51%) compared to the aforementioned databases (<10%), suggesting limitations in the use of short read SGS data for detecting long isoforms. Our previous study has already shown that at the same specificity (5% FPR), IDP had a much higher sensitivity (62% TP) than Cufflinks (20% TP). We have also applied our pipeline to other RNA-Seq data, including the MCF-7 breast cancer cell line. Our RNA-Seq Cocktail pipeline is by far the most comprehensive RNA-Seq pipeline with very high accuracy.

1479S

Typing of PRDM9 in childhood cancers from Next Generation Sequencing data. *A. Ang Houle[1,2], P. Awadalla[2].* 1) Bioinformatics Dept, University of Montreal, Montreal, Quebec, Canada; 2) Sainte-Justine Research Center, Montreal, Quebec, Canada.

Childhood cancers differ remarkably from those seen in the adult population, and their unique biology suggests that the genetic factors underlying the malignant transformation in the developing tissues differ from those generally identified in adults. The identification of these genetic factors may promote the development of biomarkers contributing to early cancer detection. Prior work from our research group suggests that patients affected with childhood acute lymphoblastic leukaemia show a higher frequency of the rare C allele of PRDM9, which is found in only ~1% of the European population. PRDM9 protein are composed of a repetitive zinc finger array at their C terminal tail, which binds to specific motifs of the DNA during meiosis, and determines the location of double-stranded breaks during recombination. This type of repetitive structure confers instability to the genetic sequence and has been shown to evolve rapidly, explaining partially the extreme polymorphism of the PRDM9 gene in the human population. Moreover, this structure poses important challenges in typing PRDM9 alleles from Next Generation Sequencing (NGS) data, and current computational typing methods are unsuccessful at identifying novel alleles. Here, we present a computational method to accurately type PRDM9 alleles from NGS data, and therefore make possible the identification of alleles in childhood cancer NGS patients. This study provides a method for the challenge of typing PRDM9 alleles from NGS data. Furthermore, the association of rare PRDM9 alleles with particular childhood cancer types may shed light on the unique molecular mechanisms underlying these diseases.

1480M

Integrated Variant Comparison Using Three Different DNA-seq Analysis Methods. *H. Beale, D. Kearns, B. Lee, D. Bailey, T. Lowe, P. Chan.* Maverix Biomics, Inc., San Mateo, CA.

Technology advancement and increased adoption of next generation sequencing has introduced a large amount of raw data that requires extensive analysis to obtain relevant biological information. However, the costs and duration of the subsequent bioinformatic steps including read mapping, variant calling and contextualization can rarely be pre-determined and often cannot be accurately assessed until a substantial portion of the analysis is complete. Here we demonstrate an approach with pre-defined costs and extensive scalability that uses the most highly considered, peer-reviewed, open source techniques and generates complete documentation of the process. The new DNA sequencing analysis pipeline on the Maverix Analytic Platform incorporates three best-in-class algorithms for variant analysis: GATK, FreeBayes, and SAMTools. We illustrate the few steps required to upload FASTQ files and to configure and launch an analysis. All data are extensively protected with security standards that are HIPAA compliant. The key results are summarized in graphs, while individual variants are annotated and presented in an interactive tabular format that is linked to a private version of the UCSC Genome Browser which can display statistics or supporting reads for a selected variant call. The common and distinct variants calls by different analysis methods are visually displayed, and the resulting variants can be extensively filtered on many characteristics. Results can also be downloaded in the tabular format or exported in VCF format for use in other downstream applications. The intermediate results, processing logs, and command lines as executed are available for those who wish to review each step of the pipeline in detail. Use of this approach accelerates the availability of the findings contained in the sequencing results without a loss of quality or control.

1481T

Predicting splicing mutations by information theory-based analysis in rare and common diseases: performance and best practices. *N.G. Caminsky[1], E.J. Mucaki[1], P.K. Rogan[1,2].* 1) Biochemistry, Western University, London, Ontario, Canada; 2) Cytognomix, London, Ontario, Canada.

Disease-causing mutations that affect pre-mRNA splicing are common. We review the literature citing information theory-based mRNA splicing mutation analyses (n=376). This type of analysis has been applied by many groups to study both common Mendelian conditions, and rare syndromes. While the Shannon entropy framework remains the foundation of these analyses, the software for performing these studies has rapidly progressed. Currently, the Automated Splice Site and Exon Definition Analysis server (splice.cytognomix.com) is used to comprehensively analyze single splicing and regulatory mutations. The Shannon pipeline (shannonpipeline.cytognomix.com) analyzes mutations in exomes, genomes, and targeted gene sets. These resources compute individual or total exon information contents (R_i, or $R_{i,\text{total}}$, in bits) of normal and mutant sequences, which correspond to thermodynamic differences that account for splicing effects. Primary applications include interpretation of novel variants, distinction of benign nucleotide alterations, deleterious alleles (including leaky, inactivating, and cryptic sites), and quantitative evaluation of changes in R_i, or $R_{i,\text{total}}$ that account for phenotypic variability and disease severity. Our results, based on a sample of these studies (n=106) report 157 natural splice sites, 19 activate cryptic sites, and 9 are regulatory splicing mutations. Of those affecting natural splice sites, 20 cause leaky and 50 inactivate splicing. Canonical dinucleotide mutations in splice sites comprised 28% but adjacent positions occurred at 41% and 31% within introns and exons, respectively. 82.7% of the studies (n=29) that validated predicted mutations with *in vitro* or *in vivo* expression assays confirmed the information theory-based analyses. Genome-scale information analyses of complete cancer genomes (3 cell lines) detected 17 to 31 splicing mutations, of which 6 to 17 abolished natural splicing, 1-5 were leaky, and 9-13 activated cryptic sites. In 447 tumors of breast cancer patients, on average, 11.8 splicing mutations were present per exome, where 3.1 were inactivating, 6.2 were leaky and 2.6 activated cryptic splicing. Splicing mutations in these tumors have been validated with RNA-Seq data by automating this process (veridical.org). In summary, we propose guidelines for detection of splicing mutations, distinction of polymorphisms, interpretation of constitutive and regulatory binding site mutations, and exon definition of splice isoforms.

1482S

PedigreeAnnotator: a GATK walker to annotate variants based on pedigree information. *B.K. Cornes, J. Patel, T. Bloom, A.V. Abhyankar.* New York Genome Center, New York, NY.

Family-based exome/genome analyses have been effectively used to identify genetic defects in rare Mendelian disorders. Several available tools, with their own advantages/disadvantages, have the ability to either integrate pedigree information or add genomic annotations but no single tool combines all the steps required to make biologically/clinically relevant variant prioritization. Previously we have described a set of R-scripts and a workflow to perform pedigree-aware analyses of next-generation sequencing data. To improve the accuracy and performance of our existing method we reimplemented our R code and workflow with the open-source version of GATK. PedigreeAnnotator accepts, as input, a PLINK format pedigree file and a VCF file containing multiple samples from a single or several families. The pedigree information is used to predict the mode of inheritance, recalibrate genotypes, compute Mendelian error probabilities & inbreeding coefficients and classify variants accordingly (e.g. denovo). Compound heterozygosity and loss of heterozygosity (LOH) annotations are also computed. All available families can be analyzed simultaneously in a computationally efficient way. Additionally, detailed QC is performed including confirmation of family relationships, gender check and cross-sample contamination. Family data is then integrated with publicly available annotation resources (e.g. OMIM, ENCODE). Furthermore, to aid variant prioritization, a scoring scheme is implemented, taking into account different attributes of the annotated variant. A summary report is generated to include QC, variant statistics and a short list of variants based on the prioritization score. PedigreeAnnotator is designed to perform extended variant analysis accurately and efficiently while being user-friendly.

1483M

GenAP workbench: aiding variant classification in clinical diagnostic settings. *M.C. Eike[1], E. Skorve[2], T. Håndstad[1], H. Fontenelle[1], K.S. Gjøtterud[1], J. Børsting[2], M. Aanestad[2], A.L. Culén[2], T. Grünfeld[1], D.E. Undlien[1].* 1) Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; 2) Department of Informatics, University of Oslo, Oslo, Norway.

Classifying genetic variants in terms of clinical significance is a complex process where multiple sources of information of variable quality are reviewed, weighted and combined. With the introduction of high-throughput sequencing technologies into the clinic, this process can quickly become a daunting task without the help of supporting tools. We have therefore developed a workbench for rules-based, semi-automated analysis of sequence variation data for diagnostic use, as part of the project *Norwegian clinical genetic Analysis Platform* (genAP).

The underlying system allows rich annotation of sequence variants from external and internal resources, and the workbench displays this information in a series of structured tabs, highlighting key information and guiding the analyst through an analysis workflow based on standard operating procedures at our department. This includes frequency data, mutation databases and prediction tools for missense variants, as well as tools for splice sites. In addition, we have developed a literature evaluation module, where relevant references are classified by the analyst according to quality and conclusions. For the final report, the system uses the annotation information and analyst input to generate suggested classifications for each variant. Approved results may be exported and added to the in-house database.

We are currently testing a prototype of the workbench, with the aim to introduce it for full scale use by the end of 2014. This presentation will demonstrate the logic of the system, with specific examples and some of our experiences from user testing.

1484T

Hadoop Acceleration of Bioinformatics Algorithms. *M. Gollery, A. Taffel, G. Ganebnii, D. Leca, A. Lisnik, M. Hohlovich, A. Dobretsov, V. Shakin, M. Mikheev.* BioDatomics, Bethesda, MD.

Genomic data production now has the pace, volume, and variety to be considered 'Big Data'. As a result, the processing of genomic data can be improved by applying lessons and techniques from other industries that have worked successfully with Big Data. One of these techniques, now in common use within multiple industries--Energy, Financial Services, Online Commerce, and Telecommunications, for example--is the use of a Hadoop-based analysis infrastructure. Hadoop is an open source software framework whose roots lie in the original Google search engine. The framework is a collection of software modules that manage the storage and processing of large datasets on clusters of commodity hardware. Because Hadoop segments and distributes data across a theoretically infinite number of compute nodes, the framework has the potential for virtually unlimited scalability and a speed increase of multiple orders of magnitude compared to legacy approaches. These developments make it hypothetically possible to accelerate the analytical processing of NGS data. The execution speed of NGS data analysis directly impacts the speed at which research findings become available. Currently, analysis of a large dataset via a complex workflow can take days--in some cases weeks. This represents a significant impediment to research productivity. The goal of this project was to reduce these protracted analysis times by optimizing bioinformatics algorithms for Hadoop and executing them on Hadoop-compatible software and a Hadoop-optimized cluster. At project initiation, BioDatomics optimized several standard bioinformatics analysis tools to take advantage of Hadoop's parallel processing capability. The resulting pipelines were then executed via the BioDT software platform, which is built on Hadoop, running on a Hadoop-optimized hardware cluster. The effect on pipeline run times, compared to standard-version tools running on a comparably-sized non-Hadoop platform, was dramatic. This poster will explain the basics of a Hadoop system, the mechanisms used to optimize bioinformatics algorithms and bioanalytic software for Hadoop, and will present data comparing Hadoop vs. legacy clusters as applied to bioanalytics.

1485S

Benchmarking of Strand NGS variant caller using a whole genome sample NA12878 and data from Genome in a bottle consortium. *R. Hariharan, R. Gupta, P. Gupta, A. Narayanan, S. Aditya, S. Katragadda, V. Veeramachaneni.* Strand Life Sciences, Bangalore, Karnataka, India.

Background and Objectives: To realize the potential and promise of Next Generation Sequencing (NGS) technology towards research and clinical applications, computational approaches are essential to call variants and translate them into actionable knowledge. Many algorithms are developed for variant calling, however often times they differ on multiple variant call predictions. In this work, we'll present Strand NGS (formerly Avadis NGS) variant calling approach and benchmarking results on a 1000 genomes CEU female sample, NA12878. Strand NGS is capable of calling different types of variants including SNPs, InDels, structural variations (SVs) and copy number variations (CNVs). We compare our variant calls (SNPs and InDels) with those from GATK UnifiedGenotyper and highly confident variant call set from NIST - Genome in a Bottle Consortium. **Results:** Strand NGS and GATK identified a total of 6,393,054 and 6,105,466 variants respectively with very similar Het/Hom and Ti/Tv ratios. We observed a high overlap (93%) in these variant calls. When variant calls from Strand NGS are filtered using quality metrics like % supporting reads, variant score, read coverage, stand bias and other PV4 biases, the overlap between the variant calls increases to ~98%, making Strand NGS and GATK very similar. We also compared the Strand NGS variant call set to highly confident variants from Genome in a bottle consortium and found a significant overlap between them. We observed that application of the filters based on the above mentioned quality metrics helps to selectively reduce false positive variant calls, there by increasing both sensitivity and specificity. **Conclusions:** Due to several issues with the raw sequencing data, variant calling is still a challenging problem. Although numerous open-source and proprietary algorithms are available, assessing how well these algorithms perform on different data sets remains a challenge. Assessment of different quality metrics like supporting reads %, variant score, read coverage, strand bias and other PV4 biases, provides a useful way to filter the likely false positive variant calls. The benchmarking results presented in this paper suggest that Strand NGS variant caller is a powerful and flexible approach to call variants and provides a visually appealing way to assess their quality using a variety of quality metrics.

1486M

Accurate detection of low-representation alleles in tumor DNA through augmented exome and transcriptome sequencing. *E. Helman, M. Clark, S. Boyle, A. Patwardhan, D. Church, M. Pratt, S. Luo, N. Leng, C. Haudenschild, R. Chen, J. West.* Personalis, Inc, Menlo Park, CA.

Somatic mutations present at a low allelic fraction have been implicated in tumor progression, recurrence, metastasis and drug resistance. The ability to detect these mutations via next-generation sequencing is crucial but often impaired by small sample size, low cellularity, and tumor heterogeneity. In order for sequencing to aid in directing personalized cancer therapies, these low-representation alleles must be accurately identified and interpreted. We sought to determine the sequencing depth and analytical parameters required to detect small variants in cancer by completing exome and transcriptome sequencing of four cancer cell lines. These lines contain a large number of known mutations in *EGFR*, *KRAS*, *BRAF*, and more than 20 other cancer genes at well defined levels of allelic representation, with some mutations at fractions as low as 1%. We completed high-depth exome sequencing with our ACE exome assay, which augments and improves coverage of over 1200 cancer genes, increasing our sensitivity for cancer variants. We utilized modified bioinformatics approaches and integrated whole-transcriptome sequencing to validate our findings. We identified all 36 known cancer mutations present within these cell line mixtures with our ACE exome and somatic pipeline. To gauge our limit of detection, we computationally down-sampled the aligned sequence files to varying depths of coverage. We find that our method is highly sensitive at moderate depths, detecting mutations present at 1% allelic fraction, and show how the limit of detection changes as depth of coverage changes. We also find high concordance between our estimated allelic fractions and known values. Notably, we recapitulate canonical *EGFR* mutations, such as T790M, which has been shown to confer acquired resistance to treatment with *EGFR* TKIs. Finally, we confirm the expression of coding mutations as well as the predicted effect of regulatory mutations by integrating results from our transciptome sequencing pipeline. Important cancer mutations are often present at small allelic fractions because of tumor heterogeneity and low purity samples. For effective cancer exome analysis, accurate detection of these variants is of paramount importance. Existing assays report limits of detection in the 5-10% range. Here, we use mixture experiments to quantitatively determine the detection limits of our enhanced exome protocol and analysis pipeline and report its sensitivity to low-representation cancer mutations.

1487T

Genomic Analysis of Blood-mediated Disorders in African Americans. *L. Jackson[1], F.L.C. Jackson[2].* 1) Biomedical Engineering, Science, and Health System, Drexel University, Philadelphia, PA; 2) Department of Biology, Howard University, Washington DC.

African-American populations are disproportionately affected by stroke, hypertension, skeletal diseases and renal disorders. This increased disease prevalence is thought to arise, in part, from population specific genetic variation at genes involved in disease phenotypes. We have developed a method of identifying the significant underlying genetic polymorphisms curated from NCBI gene sets. The resulting literature curated gene sets can be used to identify modules and pathways mediating the intersection of stroke, hypertension, renal disease, and skeletal diseases. Stroke, hypertension, renal disease, and skeletal diseases gene sets(N= 2521 genes) were obtained from NCBI Gene and then projected onto the genome, annotated with their gene ontology categories and cellular pathways to draw a bioinformatics portrayal of the overlap between these diseases with disproportionate effect in African American populations. Hotspot regions were then identified in the Human genome diversity panel populations and further characterized in HapMap populations. Mapping addiction genes onto the human genome resulted in eight gene clusters, with at least 15 disease genes (Range: 15-31 genes, pV <0.005) within a 1.5Mb contiguous distance along the genome. Hotspot genes were involved in blood circulation, cell migration, and regulation of phosphorus. Analysis of Human genome diversity panel populations with Sub-Saharan African ancestry showed population differences to non-African populations. We found eight regions of the genome that are strongly involved in complex disorders overrepresented in African American populations. Functional annotation of these hotspots identifies new candidate genes previously uncharacterized in the literature. Finally comparisons of polymorphism data point to 10 strong candidate variants that merit further characterization.

1488S

In silico prediction of splice-altering single nucleotide variants in human genome. *X. Jian[1], X. Liu[1,2].* 1) Division of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; 2) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX.

Purpose *In silico* tools have been developed to predict mutations that may have an impact on pre-mRNA splicing. The major problem that prohibits the use of these tools in research and clinical practice is the difficulty in interpreting the output, partially due to lack of large-scale evaluation analysis of these tools. We aimed to compare some of these tools on a much larger scale and to provide a more accurate and directly interpretable prediction score for splicing mutations. **Methods** Two groups of single nucleotide variants (SNVs) within splicing consensus regions (-3 ~ +8 at the donor site and -12 ~ +2 at the acceptor site) that have known impact on splicing from public databases were used to evaluate eight *in silico* tools by receiver operating characteristic (ROC) analysis with ten-fold cross-validation. Two ensemble learning methods, adaptive boosting and random forests, were used to construct new models that take advantage of individual tools and to compute an ensemble prediction score for a given mutation. An additional test set was used to validate the new models. **Results** We collected 2,959 SNVs, among which 1,164 were positive and 1,795 were negative. Four tools were excluded from ROC analysis due to high missing rate. Among the remaining four tools, the Position Weight Matrix model and MaxEntScan outperformed other tools. We combined scores of these four tools as well as four conservation scores using two ensemble methods to construct two new models. Both models significantly improved the predictive performance (area under the curve > 0.97, accuracy > 93%) and were further validated by an additional test set with 45 labeled splicing mutations. We pre-calculated our ensemble scores for all possible SNVs within splicing consensus regions across the human genome and applied them to the mutations from the Catalogue of Somatic Mutations in Cancer database. Analysis showed that predicted splice-altering variants are enriched in both recurrent mutations and known cancer genes. **Conclusions** Our results demonstrated that some *in silico* methods are powerful tools in predicting the impact of splicing mutations, and ensemble learning methods can further improve prediction. Providing a directly interpretable prediction score for all possible SNVs within splicing consensus regions across the human genome shall significantly facilitate splicing defect prediction and detection, and thus contribute to providing new targets for gene therapy and newborn screening.

1489M

High-accuracy imputation for HLA class I and II genes based on genome-wide SNP data of population-specific references. *S. Khor[1], W. Yang[2], M. Kawashima[1,3], S. Kamitsuji[2], X. Zheng[4], N. Nishida[1,5], H. Sawai[1], H. Toyoda[1], T. Miyagawa[1], N. Kamatani[2], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; 2) StaGen Co., Ltd. Statistical Genetics Analysis Division. Tokyo, Japan; 3) National Bioscience Database Center, Japan Science and Technology Agency, Tokyo, Japan; 4) Department of Biostatistics, University of Washington. Washington, United States; 5) The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan.

Statistical imputation for genotypes of human leukocyte antigen (HLA) genes is becoming an indispensable tool in the fine-mappings of disease association signals from case-control genome-wide association study (GWAS). However, most of the HLA imputation tools available are based on European descendents and not suitable for the direct application to non-European populations. Among the HLA imputation tools, HIBAG R package is a flexible tool equipped with a wide range of population based classifiers and accommodate for the building of custom classifiers. Using Japanese healthy control datasets, Japanese-based classifiers were built. HLA imputation accuracy in five HLA genes (*HLA-A, B, DRB1, DQB1, DPB1*) increased considerably to 95.1-99.5%, as compared to 90.5-98.7% when HIBAG Asian reference and 82.5-98.8% when HIBAG multi-ethnic reference were used. We also compared data from different platforms (Illumina Omni 2.5, Affymetrix 6.0, Affymetrix Axiom ASI), but there was little difference in the results. Call threshold of 0.4 is recommended for Japanese classifier in contrast with 0.5 recommended by HIBAG. Our results showed that population matched reference panel is essential to impute common HLA alleles and especially important for specific HLA alleles in individual populations.

1490T

LNCScore: a machine-learning approach for novel lncRNA discovery from RNA-Seq data. *J.H. Kim[1], O. Evgrafov[1,2], J. Knowles[1,2,3], K. Wang[1,2,3].* 1) Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA; 2) Department of Psychiatry and Behavioral Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) Division of Bioinformatics, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA.

RNA-Seq techniques are widely used in biomedical research to investigate the transcriptional landscape of different tissues or conditions on human and other species. Although most studies focused on the analysis of protein coding genes, the importance of non-coding transcripts, especially long non-coding RNAs (lncRNAs) has been increasingly recognized as important regulators of transcriptional activity and the category of them could be substantially larger than previous thought. Despite the increased interests on lncRNA, our understanding on ncRNA is still at rudimentary and studying known lncRNA and identifying novel lncRNA are challenging problems without ideal solutions yet. Typical features of lncRNA include the longer size of transcript length >200bp, the likely genomic location between genes, and the lack of complete ORF, but these features alone cannot be used to infer lncRNA confidently. To address this issue, we developed a called LNCScore for predicting novel lncRNA directly from alignment files from RNA-Seq data sets. Our strategy is based on a scoring system using multiple features for lncRNA, and use well known/annotated lncRNA to train a statistical model to identify optimized parameters to separate candidate lncRNA from random noises. We incorporated most of the commonly used features used for lncRNA prediction and added more criteria to score the likelihood of a predicted transcript to be a genuine lncRNA, such as phylogenetic modeling and comparison with related species, the use of SNPs that may result in new open reading frames, the epigenetic markers from ChIP-Seq experiments where available, the gene family information of neighboring genes, and comparison to well established database like NONCODE and LNCipedia. This can be used on poorly studied tissues from model organisms with rich annotation data and also can be applied to newly sequenced species with limited functional annotations on genes. The input of our software is the alignment files in BAM format generated with reference-guided assembly, and the output is a list of possible novel lncRNA, each with LNCScore (a confidence score for the likelihood of being a genuine lncRNA) as well as FPKM values and read counts, to facilitate downstream analysis such as differential expression or splicing. Our method will help researchers identify novel lncRNAs and help determine the potential functional significance of these novel non-coding RNAs.

1491S

Biomarker Discovery From RNA-seq Data Using a Biologist-Friendly Analysis Platform. *B. Lee, D. Kearns, D. Bailey, T.M. Lowe, P. Chan.* Maverix Biomics, Inc., San Mateo, CA.

The analysis of large-scale next-generation sequencing data requires both increasing amounts of computing power and exponential growth in storage. In many cases, the data may be generated by multiple laboratories at distant locations, which complicates the management, processing, and collaboration required to interpret the data. To enable hands-on exploration of RNA-seq data sets by a much larger community of biologists, we have developed the Maverix Analytic Platform. This cloud-based platform, that is built for the biologist, performs quality assessment, leverages best-in-class analysis methods, and provides an integrated UCSC-genome browser endpoint to enable visualization and interpretation of results. In this case study, we will present an automated RNA-seq pipeline that helps improve gene expression analysis by providing results from two common algorithms, Cufflinks and DESeq. This automated pipeline provides multiple sample sorting and filtering capabilities to maximize analysis flexibility in interpreting results. A comparison between the two methods is displayed in a Venn diagram for quick assessment of highly concordant differentially expressed genes in addition to downloadable gene lists and expression results. With the genome browser, we will show a comparison to microarray data through customizable data tracks. The results include examples of potential biomarkers identified by using this integrated analysis approach that illustrates the untapped opportunities for discovery in large-scale RNA-seq analysis.

1492M

PARADIGM-SHIFT predicts the functional impact of 'driver modules' in multiple cancers using pathway impact analysis. *S. Ng[1], C. Benz[2], D. Haussler[1], J. Stuart[1].* 1) Biomolecular Engineering, UC Santa Cruz, Santa Cruz, CA; 2) Buck Institute for Research on Aging, Novato, CA.

The major mechanism by which cancer arises is through genomic alterations. These alterations can lead to changes in gene regulation, protein structure, and function. Individual tumors can contain hundreds to thousands of alterations. It is critical to distinguish alterations that have an important role defining the cancer - drivers - from alterations that are unimportant to the tumor - passengers. Driver genes can lead to significant changes in their pathways; however, alterations in a single gene may not be sufficient to explain all the pathway perturbations across patients. Additional alterations could range from DNA copy number changes, gene-fusions, or even lesser understood non-coding mutations. Identifying these 'driver modules' is essential for understanding cancer disease mechanisms, which can help guide treatment decisions as well as identify novel targets for treatment.

We have developed a functional impact prediction method called PARADIGM-SHIFT based on integrated pathway analysis to discriminate loss-of-function, neutral, and gain-of-function alterations. Utilizing the set of regulatory interactions annotated for a given gene, we can detect a shift in the downstream effects of an altered gene compared to what is expected from its upstream influences. Additionally, since these shifts in pathway signal can be detected for all samples, PARADIGM-SHIFT can be used to identify additional genomic alterations that lead to similar changes to the altered pathway to form 'driver modules.' Application of our method to the TCGA Pan-Cancer cohort identifies many genes with significant alterations that lead to loss- and gain- of function. PARADIGM-SHIFT then identifies several additional genomic alterations, including gene-fusions and non-coding mutations, which are significantly implicated in these pathway changes. This analysis offers insight into the mechanism of gene-fusions and non-coding mutations that cannot be assessed by most conventional methods.

1493T

Geneious R7: a bioinformatics platform for biologists. *C. Olsen[1], K. Qaadri[1], R. Moir[2], M. Kearse[2], S. Buxton[2], M. Cheung[2], B. Milicevic[2], W. Hengjie[2], J. Kuhn[2], S. Stones-Havas[2], C. Duran[2].* 1) Biomatters, Inc. Newark, NJ; 2) Biomatters, Ltd. Auckland, New Zealand.

Biomatters' Geneious R7 is a bioinformatics software platform that allows researchers the command of industry-leading algorithms and tools for their genomic and protein sequence analyses. Using a glass-box approach for software design, Geneious R7 offers a comprehensive suite of peer-reviewed tools that enable researchers to be more efficient with their bioinformatic workflows. Researchers at all levels can easily manage, analyze, and share their sequence data via a single intuitive software application. R7 provides tools for next-generation sequence analysis, sequence alignment, molecular cloning, chromatogram assembly, and phylogenetics. New features for this major version release include tools for Gibson & TALEN assembly, TOPO cloning along with algorithms for RAxML, FastTree, Garli, LastZ, Bowtie2, as well as a number of new plug-ins. R7 affords real-time dynamic interaction with sequence data and empowers biologists to produce stunning publication quality images to increase the impact of their research. By utilizing Geneious R7, biologists can easily improve their sequence analysis workflow efficiencies to free up more time for their research. This poster aims to demonstrate the new features and benefits of the highly integrated Geneious R7 tool-suite.

1494S

Assembly Hubs and Genome Browser in a Box Makes Viewing Private Annotations and Custom Sequences Easy. *B.J. Raney, M. Haeussler, A.S. Hinrichs, A.S. Zweig, D. Karolchik, H. Clawson, J. Casper, M. Speir, B. Lee, T. Dreszer, G. Barber, P. Fujita, D. Haussler, W.J. Kent.* University of California at Santa Cruz, Center for Biomolecular Science and Engineering, School of Engineering, Santa Cruz, CA, USA.

The UCSC Genome Browser is one of the most popular resources in genomics, allowing users to view a large collection of annotations on human, mouse and many other vertebrate assemblies released by NCBI. The Genome Browser also allows users to upload their own annotations as custom tracks, or more recently, to create stand-alone Track Hubs which can contain thousands of tracks of custom annotations.

Two new features of the UCSC Genome Browser facilitate its use in situations where privacy is a priority, and where the sequence being visualized is not part of UCSC's standard set of assembly browsers.

The first technology is called Genome Browser in a Box (GBiB) which uses virtual machine technology to allow users to easily mirror the genome browser on a protected network, or in situations when private annotation cannot be uploaded to UCSC for privacy reasons. To install GBiB one need only download VirtualBox, which is freely available for all major operating systems, and the GBiB virtual image which is a pre-installed copy of the UCSC Genome Browser. The Genome Browser is open source and free for non-commercial users.

The second new feature of the UCSC Genome Browser is Assembly Hubs, an extension of the Track Hub mechanism introduced in 2012. Track Hubs allow users to create their own annotation on assemblies supported at UCSC. Assembly Track Hubs allow users to provide their own DNA sequence on which the Track Hub annotation can then be viewed. Assembly Track Hubs require only a network accessible location where the sequence and annotations are stored in a flat-file directory hierarchy. Currently Track Hubs support the following annotation formats: BigBed, BigWig, BAM, HAL, and VCF/tabix.

These two technologies combined make it possible for an end-user to create a UCSC browser instance with custom sequence and annotations without passing sensitive data over the Internet. Using GBiB and Track Hubs, users can view their own annotations aligned with other annotations curated at UCSC.

1495M

Comparative transcriptome analysis reveals a proangiogenic compensatory mechanism for increased placental vascularization in women with reduced vasodilation. *L.M. RODRIGUEZ[1], L. RISHISHWAR[2], L. WANG[2], A.C. AGUILAR[1], K. JORDAN[2], A. CASTILLO[1].* 1) UNIVERSIDAD DEL VALLE, CALI, VALLE DEL CAUCA, COLOMBIA; 2) GEORGIA INSTITUTE OF TECHNOLOGY, ATL, GA, USA.

Purpose: The activation of angiogenic pathways is essential for placental development in pregnant women. Flow mediated dilation (FMD) is a metric commonly used to measure endothelial response to the increased blood flow characteristic of placental vascularization during pregnancy. We are taking a number of approaches to compare cohorts of high- versus low-FMD pregnant women in an effort to understand how endothelial response to vasodilation relates to the process of placental angiogenesis and vascularization during pregnancy. For this study, we used a genome-wide expression (i.e. transcriptome) analysis approach to identify genes that may help to mediate the link between endothelial response and the construction of the placental vascular network during pregnancy. Methodology: RNA-seq was used to characterize genome-wide expression levels from placental tissue for cohorts of pregnant women that had either high versus low-FMD in order to identify differentially expressed genes that may be involved in angiogenesis pathways. We focused our differential expression analysis on down-regulated targets of proangiogenic miRNAs previously shown to be up-regulated in low-FMD individuals. Results: We applied a novel transcriptome analysis approach (GFOLD) that allows for characterization of differentially expressed genes using low sample sizes; in this case, we had sample sizes of n=2 each for the high and low-FMD cohorts. Using this approach, placenta samples from women with low-FMD showed significantly down-regulated expression for the igfbp1 (GFOLD score = -1.164) and lep (GFOLD score = -0.079) genes, which encode for the Insulin-like growth factor (IGF)-binding proteins (IGFBP1) and Leptin (LEP) respectively. These genes are both expressed in placenta and participate in the activation of the proangiogenic Insulin Receptor Signaling (IRS) pathway. Conclusion: These preliminary results are consistent with a role for IGFBP1 and LEP in a process of placental hypervascularization that may help compensate for reduced endothelial response to vasodilation in low-FMD women. This would account for the fact that the low-FMD women in our cohort do not show any apparent pregnancy complications related to placental vascularization. However, the down-regulation of these genes, and the effect on their pathways, may lead to subsequent problems with predisposition to diabetes and macrovascular disease in low-FMD women.

1496T

Pilot data from the Virtual Genomics Clinic (VGC). *J.V. Thakuria[1, 2, 3].* 1) Massachusetts General Hospital, Boston, MA; 2) Theopolis, Cambridge, MA; 3) Veritas Genetics, Cambridge, MA.

While the number of published reports demonstrating medical utility of exome sequencing in patients suspected of monogenic disease continues to grow, data are still limited on the medical utility of exome sequencing in healthy, asymptomatic individuals. Over the past year, ACMG recommendations to return results on a panel of 56 genes in both adult and pediatric cases whenever clinical exome or whole genome sequencing is performed, highlights the importance of gathering this data. Initial reports on the number needed to screen (NNS) in the general population before a suspected pathogenic mutation is identified varies widely from <1% to >20%, but is usually reported in the 1.5%-5% range (excluding potential reproductive utility from knowledge of carrier status for recessive disorders) - and is highly dependent on variant filtering stringencies. Preliminary pilot phase data from over 80 exomes with ~60X average on target coverage from the Virtual Genomics Clinic (VGC) study parsed against the 56 ACMG IF (incidental findings) genes and undergoing secondary individualized analyses based on voluntarily provided personal and family medical histories, is reported. The pilot cohort was primarily recruited from the Young Presidents' Organization (YPO). Study consent, pre- and post-test genomic counseling, as well as return of results, are generally performed remotely via tele- or videoconferencing. All study participants are given the option to receive an easily portable, compressed copy of their exome data on thumbnail drive and may opt in for indefinite biobanking of their samples for ongoing research and clinically focused reanalysis of data on a yearly basis on average and guided by evolving medical genomics knowledge. Alternatively, participants may opt to have their samples discarded and backup data deleted after receiving their compressed data and initial medical interpretation of results. Data privacy is enhanced using several "off the grid" features such as hybrid systems utilizing onsite bioinformatics infrastructure for storage and processing while leveraging cloud-based analyses tools on pooled, de-identified data. Future efforts, including the use of physical decryption keys such as Giambattista della Porta modified Alberti cipher disks for off-site collaborators (eg, for access to corresponding phenotype data), will add additional layers of data security in the VGC.

1497S

Mega2: data reformatting for facilitating genetic linkage and association analyses. *D.E. Weeks[1, 2], R.V. Baron[1], C. Kollar[1], N. Mukhopadhyay[3].* 1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 3) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA.

A comprehensive genetic study of a complex disease invariably requires reformatting the data into the precise input format required by each of many analysis programs used. Our Mega2 software facilitates this process by automating common data reformatting tasks, thereby markedly reducing the chance of data reformatting errors, making data setup much less error-prone, and saving the user the time of writing, debugging, and maintaining data conversion scripts. We have recently extended the capabilities of Mega2 in a number of ways. In addition to accepting LINKAGE-format input files, it now accepts VCF, BCF, and compressed VCF files and can apply common 'VCFtools' filters. While continuing to support conversion to commonly-used formats like Merlin, Mendel, SimWalk2, and SOLAR, PLINK, Cranefoot, IQLS, FBAT, MORGAN, BEAGLE, Eigenstrat, and Structure, Mega2 has now been extended to support data conversion to PLINK/SEQ format. Mega2 can now pass through non-numeric alleles to analysis programs that will accept them, instead of recoding them to numeric alleles as was done previously. For some output options, Mega2 can generate high-quality plots of the results using our nplplot R package, as well as generate custom track files for use within the UCSC genome browser. When controlled by a batch file, Mega2 can be used in an automated manner within data analysis pipelines. Mega2 also supports organisms other than humans. Mega2 is open source and freely available and has extensive documentation and a tutorial, available on our http://watson.hgen.pitt.edu/register web site. This work was supported by NIH grant R01 GM076667 (PI: Weeks). Earlier contributions to our initial code base were made by Lee Almasy, Mark Schroeder, and William P. Mulvihill.

1498M

Single Cell RNA-Seq analysis of Tumor Composition. *I. Ragoussis[1,4], Y-C. Wang[1,4], E. Iacucci[1,4], L. Letourneau[4], P. Savage[2], A. Monast[2], N. Bertos[2], A. Omeroglou[3], M. Park[2].* 1) Department of Human Genetics & McGill University and Genome Quebec Innovation Centre, McGill University, Montreal, Quebec, Canada; 2) The Rosalind and Morris Goodman Cancer Research Centre, McGill University, Montreal, Quebec, Canada; 3) Department of Pathology, McGill University, Montreal, Quebec, Canada; 4) McGill University and Genome Quebec Innovation Centre, McGill University, Montreal, Quebec, Canada.

Single-cell RNA-Seq experimentation and analysis represents an advanced and powerful tool for cutting-edge human genetics research as we can now monitor the whole transcriptome at the single-cell level. The ability to interrogate control and disease states of single cells allows for specific and unadulterated characterization of these conditions. Here we examine the issues surrounding the analysis of individual and bulk breast cancer cells derived from a tumor biopsy. In brief, live single-cell mRNA to cDNA libraries were prepared using the Fluidigm C1 Single-Cell microfluidic system. cDNA libraries were then converted to sequencing ready libraries using the Illumina Nextera XT kit and sequenced on the HiSeq2500 using Rapid Paired End 150bp mode. In parallel, control cDNA libraries were prepared from extracted total RNA and 200 cells serving as benchmarks to systematically evaluate the sensitivity and accuracy of our single-cell RNA-seq approach. We sequenced > 100 cells from a Her2 + ve breast tumor and detected between 4501 and 7168 genes in the single cell samples and between 8782 and 13512 genes in the bulk samples. The analysis was carried out via a process which included filtering, clustering, as well as various comparative and pathway analysis. Clustering (based on read counts and FPKMs) revealed three distinct and state-related expression profile groups (clusters). Statistical analysis between these groups of cells revealed several significant genes, of which five were known breast cancer stem cell markers (CD44, TGFBR2, MUC1, KRT5, and ITGA6). Pathway analysis provided greater detail into the mechanisms (such as cell-cycle checkpoint and apoptosis) believed to be responsible for the cellular states of the groups as mediated by known cancer associated genes (BCAS2, BRIX1, BRE, and TRIAP1). These approaches demonstrate our ability to classify single cell samples, identify potential biomarkers, and elucidate which pathways were involved in the aberrant cell growth. Complementary to this, single-cell samples showed cell state specific clustering leading to the conclusion that this approach allows for the characterization of cellular heterogeneity, both important steps in the understanding of tumor biology. Within this setting, we have been able to move towards distinguishing between different cell populations in the tumor sample, characterizing the tumor purity, and identifying distinct expression patterns.

1499T

Standardized phenotyping enables rapid and accurate prioritization of disease-associated and previously unreported sequence variants. *W.P. Bone[1], D.R. Adams[1], M.J. Davis[2], D. Draper[1], E.D. Flynn[1], R.A. Godfrey[1], C. Groden[1], M. Haendel[6], E. Lee[1], A.E. Links[1], T. Markello[1], C. Mungall[4], M. Nehrebecky[1], P.N. Robinson[5], M. Sican[1], D. Smedley[3], C.J. Tifft[1], C. Toro[1], E. Valkanas[1], C. Wahl[1], N.L. Washington[4], L. Wolfe[1], C.F. Boerkoel[1], W.A. Gahl[1].* 1) Undiagnosed Diseases Program, National Human Genome Research Institute, Bethesda, MD; 2) National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA; 3) Mouse Informatics group, Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, United Kingdom; 4) Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA; 5) Institute for Medical and Human Genetics, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany; 6) University Library and Department of Medical Informatics and Epidemiology, Oregon Health and Sciences University, Portland, Oregon 97239, USA.

The NIH Undiagnosed Diseases Program (UDP) uses exome analysis for generating hypotheses of potential genetic causes of disease. This approach, which incorporates Mendelian segregation and population frequency filters and prioritizes on rarity and predicted deleteriousness, has enabled several diagnoses. However, this method is laborious, even when the variants are in known disease-causing genes. Therefore, we hypothesized that use of patient phenotype data recorded in Human Phenotype Ontology (HPO) terms would enhance the efficiency of diagnosing known diseases. To this end, we used the Exomiser 2.0 algorithm, which queries human, mouse, and zebrafish phenotypic data as well as protein-protein association data. The algorithm prioritizes variants based on variant allele frequency, predicted deleteriousness, and phenotypic relevance. For eleven solved UDP cases, Exomiser 2.0 ranked all eleven diagnostic variants within the top ten candidates; five ranked as the top candidate. When applied to a cohort of 23 undiagnosed UDP patients, Exomiser 2.0 aided in the diagnosis of three cases by prioritizing variants for two known diseases and for one new disease. We conclude that use of standardized phenotyping in combination with purely genomic exome analysis provides a rapid and effective method of screening for variants in known disease causing genes and possibly for identification of new disorders.

1500S

Discovery and validation of mechanistic underpinnings of *cis*-regulatory variants underlying *FTO* association with type 2 diabetes and obesity risk. *M. Claussnitzer[1, 2,3], LD. Ward[1], YH. Hsu[2], H. Hauner[3], M. Kellis[1].* 1) Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA; 2) Institute for Aging Research, Harvard Medical School, Boston MA; 3) Institute for Nutritional Medicine, Technical University München, Munich Germany.

Genome-wide association studies (GWAS) have revealed numerous risk loci associated with type 2 diabetes (T2D) and obesity. However, GWAS signals are rarely traced to the disease-causing variants, given the still incomplete annotation of the human genome, especially for non-coding variants that account for the majority of disease associations. This is particularly complex given the uncertainty on the specific causal variant (given numerous variants in linkage disequilibrium, LD), the uncertainty on the cell type of action, the target genes and upstream regulators of non-coding variants. Here, we overcome these challenges using a combined computational and experimental approach. We use regulatory annotations across 127 tissue/cell types to prioritize functional non-coding variants in regulatory regions, and a comparative genomics across multiple species to recognize likely driver regulatory motifs. We next use diverse experimental assays to validate the cell-type specific regulatory activity for 16 predicted variants and their binding regulators. In particular, for the *FTO* obesity/T2D risk locus, we report multiple lines of evidence supporting that the intronic variant rs1421085 is causal for the FTO association, acts in adipocytes, affects ARID5B binding, and alters *IRX5* gene regulation. Briefly: 1) rs1421085 harbors a functional regulatory ARID5B motif based on conservation of a ARID5B-CART-GATA motif cluster across species; 2) rs1421085 is located in a candidate adipocyte enhancer based on chromatin state analysis across 127 reference epigenomes; 3) rs1421085 shows allele-specific binding of ARID5B using electrophoretic mobility shift assays; 4) rs1421085 shows cell type-specific enhancer effects for SGBS adipocytes using luciferase assays; 5) the rs1421085 risk allele increases *IRX5* expression in human adipocytes, based on qRT-PCR; 6) Lastly, increase in *IRX5* mRNA levels depend on the risk allele and regulation by ARID5B, based on *ARID5B* siRNA knockdowns. This suggests that a non-coding variant acting in adipocytes underlies the FTO association which was previously thought to be mainly driven by the brain. More broadly, our results suggest a general method for the computational discovery and experimental dissection of disease variants and have important implications on the study complex traits, which can help bridge the genotype-to-phenotype gap between genetic variants, molecular mechanisms, and cellular and organismal phenotypes.

1501M

Fast and Accurate Site Frequency Spectrum Estimation from Low Coverage Sequence Data. *E. Han[1], J. Sinsheimer[1,2], J. Novembre[3,4].* 1) Biostatistics, UCLA, Los Angeles, CA; 2) Human Genetics and Biomathematics, UCLA, Los Angeles, CA; 3) Ecology and Evolutionary Biology, UCLA, Los Angeles, CA; 4) Human Genetics, University of Chicago, Chicago, IL.

The distribution of allele frequencies across polymorphic sites, also known as the site frequency spectrum (SFS), is of primary interest in population genetics. It is a complete summary of sequence variation at unlinked sites and more generally, its shape reflects underlying population genetic processes. One practical challenge is that inferring the SFS from low coverage sequencing data in a straightforward manner by using genotype calls can lead to significant bias. To reduce bias, previous studies have used a statistical method that directly estimates the SFS from sequencing data by first computing site likelihood vectors (i.e. the likelihood a site has a each possible allele frequency conditional on observed sequence reads) using a dynamic programming (DP) algorithm. Although this method produces an accurate SFS, computing the site likelihood vector is quadratic in the number of samples sequenced. To overcome this computational challenge, we propose an algorithm we call the "adaptive K-restricted" algorithm, which is linear in the number of genomes to compute the site likelihood vector. This algorithm works because in a lower triangular matrix that arises in the DP algorithm, all non-negligible values of the site likelihood vector are concentrated on a few cells around the best-guess allele counts. We show that our adaptive K-restricted algorithm has comparable accuracy but is faster than the original DP algorithm. This speed improvement makes SFS estimation practical when using low coverage NGS data from a large number of individuals.

1502T

Parallelization of genome-wide local ancestry inference. *R. Johnson[1], G. Nelson[1], C. Winkler[2].* 1) BSP CCR Genetics Core, Frederick National Laboratory, Frederick, MD; 2) Basic Research Laboratory, Frederick National Laboratory, Frederick, MD.

Admixture linkage analysis has proven to be a powerful method to map disease genes and other phenotypic traits. The ALDsuite package in R uses a hidden Markov model (HMM) to infer local ancestry in both sparse and dense marker data, includes statistical tools to map disease genes in admixed populations and generates graphical output for visualizing results. One common drawback of HMM frameworks is their computational inefficiency, relative to other algorithms. While ALDsuite has been designed to maintain computational efficiency with marker sets of increasing density, the use of multiple Markov chains in parallel processes can offer significant time savings. During burn-in iterations, each independent chain's parameter state is combined and a global parameter state is maintained. At the completion of each iteration, each local parameter state is updated to be more like the global parameter state, using a weighted sum of the two. Early in the burn-in phase the global parameter state is more heavily weighted to push all chains to a more likely region of the parameter space. Later in the burn-in phase the local parameter state is more heavily weighted to allow more independent starting points for follow-on iterations. Each chain is then allowed to independently sample the parameter space until the desired number of samples has been reached. ALDsuite was developed to provide geneticists with easy to use software to augment admixture mapping studies with powerful insights gained from an analysis of admixture linkage disequilibrium in two-way and poly admixed populations. This parallelization scheme allows for a near-linear increase in speed per additional processor.

1503S

CliniCall - Bridging the Gap From High-Throughput DNA Sequencing to Actionable Variants. *S. McGee, G. Jimenez, T. Kolar, M.O. Dorschner, J.D. Smith, D.A. Nickerson.* Genome Sciences, University of Washington, Seattle, WA.

CliniCall is a stand-alone application that bridges the information gap between patient DNA sequencing data and clinical relevance. With the explosion of clinical DNA sequencing tests available there is a critical need for analysis tools to help filter and prioritize clinically actionable variants. This has created a significant challenge for clinical laboratories levering these new technologies, ranging from weighing the importance of rare variation to cross-referencing multiple annotation databases. To facilitate analysis, interpretation and reporting, we have developed a user-friendly tool to view sequence alignments, quality metrics and variant annotation. A combination of BAM and VCF files - or simply a GVCF file - can be easily imported for variant identification within user-defined gene sets. With CliniCall's new multi-sample display, variants can be easily compared across samples and filtered on a user-defined set of metrics such as allele frequency, conservation score, presence/absence in specific databases and annotation functionality or clinical-significance. The CliniCall platform clearly summarizes data quality for the sequenced samples, and simultaneously links to existing databases. In addition, CliniCall now not only graphically summarizes the data, but allows for filtering on a constantly-updated set of ClinVar variants. In addition it provides an up-to-date set of Targeted Gene Panels - such as the ACMG 56 Incidental Genes - that can be automatically pre-loaded into the analysis scheme for full reporting. This all is designed to make quantitative decisions regarding the validity and categorization of the identified variants easy for the clinical geneticist. This is done by utilizing an extensive set of exome sequences to report a statistical measure of confidence for each newly discovered rare variant based on the quality of all previous calls. The CliniCall tool provides an easy and efficient way to integrate a myriad of data into a single viewable format for clinical genetics.

1504M

A Comparison of Genomes and Exomes and the Impact on the Incidentalome. *E.G. Farrow[1], L.K. Willig[1], C.J. Saunders[1], K.J. Barger[2], G.P. Twist[1], N.A. Miller[1], S.F. Kingsmore[1].* 1) Center for Pediatric Genomic Medicine, Children's Mercy Hospitals & Clinics, Kansas City, MO; 2) Endocrinology, Children's Mercy Hospitals & Clinics, Kansas City, MO 64108.

As the cost of genomic sequencing continues to decline, the comparison of data generated from whole genome versus whole exome sequencing becomes increasingly important. Adding another layer of complexity, the bioinfomatics tools utilized for analysis can have a large impact on the number and quality of variants. In order to establish the best practices within our center, data from a set of genomes and exomes was analyzed utilizing different filtering metrics, both with and without Variant Quality Score Recalibration (VQSR), notated as '.raw'. The comparison set included approximately 75 genomes and 1200 exomes. Genome samples were prepared for sequencing using the Illumina TruSeq PCR free sample kit. Exome samples were prepared utilizing Illumina's Expanded Exome kit with TruSeq library sample preparation. In order to minimize differences in variant calling, the entire set of exomes underwent variant calling as a single batch. As expected, removing filtering applied by VQSR led to an increase in the total number of variants detected; approximately 40% in genomes and 8% in the exome data set. When the analysis was restricted to rare ACMG category 1-3 variants, a similar increase was seen: approximately 30% in genomes and 5% in the exome data set; indicating that the increase in variants in the .raw files is not limited to non-coding regions. Surprisingly, although the total number of variants was higher overall in genomes, the average number of rare variants in ACMG categories 1-3 was only slightly different between genomes and exomes, both .raw and with VQSR (genome 936 and 687 variants, exome 904 and 858, p=.04 and .09 respectively). The incidentalome is also an important component of analysis for genomic data. When the analysis was further restricted to include only ACMG category 1 and 2 variants with a MAF<1% in the genes recommended for the incidentalome, the average number of variants per sample was not statistically different between genomes and exomes, .raw and filtered files (genome 1.6 and 1.3 variants, exome 1.3 and 1.3 variants, respectively). Taken together, our results demonstrate that eliminating VQSR results in higher sensitivity without a significant decrease in specificity and that whole genome analysis detects slightly more rare variants per sample than whole exome analysis. Further, our results indicate that the burden of the incidentalome, in both genome and exome, in our data set is approximately one variant per person.

1505T

Diploid Alignment of Whole Human Genome Data. *P.J. Pemberton[1], E. Valkanas[1], W.P. Bone[1], C.J. Markello[1], E.D. Flynn[1], A.E. Links[1], C.F. Boerkoel[1], D.R. Adams[1,2], W.A. Gahl[1,2,3], T.C. Markello[1,2], NIH Intramural Sequencing Center, NIH Genomics Core, and UDP Clinicial Team.* 1) Undiagnosed Diseases Program, NHGRI/NIH, Bethesda, MD; 2) Medical Genetics Branch, NHGRI/NIH Bethesda, MD; 3) Clinical Director, NHGRI/NIH Bethesda, MD.

A major problem in the analysis of human genome sequences is the miss-mapping of short read NextGen sequences to nearly equivalent gene loci, i.e. Loci homologous to the sequence miss-mapped. This is a function of the number and density of the difference between any one human sequence and the human reference genome used as the alignment template. In order for the alignment template to be a closer match to the sequence that the short reads were derived from, we customized the canonical reference sequence with high confidence differences from the canonical sequence that were derived from SNP chip hybridization results, BEAGLE imputed 1000 genome haplotypes, and results from one iteration cycle of the short read data results of genotyped calls by Haplotype Caller according to Broad Institutes' best practice pipeline. Diploid alignment exome data has already been performed in the UDP and is being run on a commercial platform (Appistry Inc.). We now describe this technique as applied to the whole genome sequencing, using low amplification 105bp Illumina HiSeq2000 short read data. This data is approximately 20x larger than exome data, and presents challenges to the current NIH Biowulf super cluster computing environment, and to the current generation of analytic and data storage resources of the UDP. We present our current results on the first 12 genomes and compare these to the exome data from the same family. With this further progress we plan to extend our analytic search for rare unknown disease causal variants beyond the exomic part of the human genome.

1506S

Evaluation of INDEL Callers for Next-Generation DNA Sequencing Data. *R.L. Goldfeder, E.A. Ashley.* Biomedical Informatics, Stanford University, Stanford, CA.

Small insertions and deletions (INDELs) in the human genome play a significant role in disease and genetic variation. In both research and clinical settings, the ability to detect INDELs is critical for disease understanding, diagnosis, and treatment. However, detecting INDELs from next-generation DNA sequencing data is still a major challenge. In order to better understand the current state of the art, we evaluated five commonly used INDEL-calling pipelines: Pindel, Dindel, SOAPindel, UnifiedGenotyper and HaplotypeCaller. To evaluate each pipeline, we generated a synthetic genome that contains INDELs of varying sizes. We created paired-end artificial sequencing reads for this genome, incorporating an error profile similar to that of a HiSeq2000 using ART. Finally, we processed the reads with each of the five pipelines using “best practice” parameters and filters and calculated recall and precision. We found greater variability in recall than precision across the pipelines. We also evaluated how robust each pipeline was to changes in alignment algorithms, read length, coverage, and variant type (homozygous vs heterozygous, with SNPs present or absent). In general, the pipelines perform better with Novoalign, longer reads, higher coverage, and homozygous INDELs. Our results suggest that longer reads and high coverage are necessary for clinical-grade INDEL detection.

1507M

Relationship detection with high-density SNP genotypes obtained from sub-nanogram amounts of fragmented DNA. *D.J. Witherspoon[1], J.C. Tackney[1,2], T.J. Parsons[3], L.B. Jorde[1].* 1) Dept. of Human Genetics, University of Utah, Salt Lake City, UT; 2) Dept. of Anthropology, University of Utah, Salt Lake City, UT; 3) International Commission for Missing Persons, Sarajevo, Bosnia and Herzegovina.

High-density SNP genotypes collected from DNA samples can be used to detect relationships between individuals with high power and accuracy. Such information is also valuable for identifying forensic DNA samples, such as those from crime scenes and from unidentified victims of disasters. High-density SNP genotyping platforms use short probes to quickly and cost-effectively collect genotypes at up to several million loci with ≥98% accuracy. However, they require substantial amounts (≥100 ng) of clean, unfragmented DNA (≥2kb length), which is then massively amplified by isothermal whole-genome amplification (WGA) prior to hybridization and genotyping. Forensic DNA samples are often severely fragmented and of limited amounts. Although microarray probe sizes (≤50 bp) are ideal for genotyping fragmented DNA, WGA fails for short DNA fragments, leaving most forensic DNA samples inaccessible to microarrays. To overcome this limitation, we used a high-throughput sequencing library preparation method developed for ancient DNA (Gansauge and Meyer Nature Protocols 2013). Biotinylated adapter oligonucleotides are ligated to single-stranded heat-denatured DNA, ligated to a dsDNA adapter after complementary strand synthesis, and further amplified by PCR to generate the sequencing library. However, we continue by ligating these library fragments to each other, creating concatemers long enough to amplify by isothermal WGA. The WGA product is then genotyped using the standard Illumina Infinium LCG protocol. To test this protocol, we acoustically fragmented human DNA to 100 bp length and removed fragments longer than 200 bp using SPRI beads. From 200 pg of this DNA, the library preparation and ligation generated 16 ng of product, which we used as input for genotyping on the Illumina HumanOmni2.5 platform (2.3 million SNPs). This yielded 1.3 million called SNP genotypes (57% call rate) of which 1.1 million genotypes were correct (80%, by comparison with genotypes from high-quality DNA of the same donor). Despite the 20% error rate, geographic ancestry of the sample donor is accurately estimated. Simulations demonstrate that even 10,000 SNP genotypes with a 20% error rate suffice to conclusively link a sample to first and second degree relatives with >95% power. Methods that explicitly model high SNP genotype error rates and use more SNPs should allow more distant relationships to be reliably detected. This research was supported by DOJ award 2012-DN-BX-K037 to LBJ.

1508T

Quantifying mitochondrial copy number using next-generation sequencing data. *P. Billing-Ross[1], K. Ye[2], A. Keinan[3], Z. Gu[2].* 1) Molecular Biology & Genetics, Cornell University, Ithaca, NY; 2) Department of Nutritional Sciences, Cornell University, Ithaca, NY; 3) Department of Biological Statistics & Computational Biology, Cornell University, Ithaca, NY.

The abundance of mitochondrial DNA (mtDNA) differs between cell types and changes in mtDNA copy number (mtDNA CN) have been associated with complex diseases such as cancer. Mitochondria are organelles responsible for the vast majority of energy production within the cell and are centrally involved with processes including signaling, calcium regulation, and programmed cell death. Mitochondria retain a small circular genome which codes for 13 genes, 22 tRNAs, and 2 rRNAs which are essential for respiration. A single cell contains hundreds to thousands of mtDNAs. Because of its central role in respiration, changes in mtDNA CN can have severe effects on metabolism and cellular function. In particular, drastic changes in mtDNA CN have been repeatedly observed in a variety of cancers. Measurement of mtDNA CN has traditionally been performed using real-time quantitative PCR; however, data generated from next-gen sequencing provides an opportunity to quickly and easily quantify mtDNA CN without the need for additional experimental techniques. Several methods have been described to measure mtDNA CN using sequencing data, but the accuracy of these methods has not been well verified. We use a simulation study to compare the accuracy of various methods for measuring mtDNA CN and determine an optimal approach for measuring mtDNA CN using next-gen sequencing data. Accurate estimation of mtDNA CN from sequencing data will provide a tool to incorporate mtDNA CN into association studies and measure changes in mtDNA CN during disease progression.

1509S

An accurate and integrative computational approach for cancer genome studies. *L.T. Fang[1], M. Mohiyuddin[1], J.C. Mu[2], P.T. Afshar[2], A. Kiani[3], N. Bani Asadi[1,3], L. Bullinger[4], A. Dolnik[4], C. Yau[5], W.H. Wong[6,7], H.Y.K. Lam[1].* 1) Department of Bioinformatics, Bina Technologies, Redwood City, CA; 2) Department of Electrical Engineering, Stanford University, Stanford, CA; 3) Department of Engineering, Bina Technologies, Redwood City, CA; 4) Department of Internal Medicine III, University Hospital of Ulm, Ulm, Germany; 5) Wellcome Trust Centre for Human Genetics, Oxford University, Oxford, UK; 6) Department of Statistics, Stanford University, Stanford, CA; 7) Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA.

Identifying somatic mutations is a key analysis and challenge in cancer research. Its complexity lies in the impure and heterogeneous nature of the samples, rendering each data set a unique problem. An algorithm in one mutation detector may work well for one data set but poorly for another. In this regard, we take an integrative approach to identify and rank the most clinically important mutations based on a combination of different algorithms, sequencing features, and prior knowledge.

We incorporate MuTect, SomaticSniper, VarScan2, and JointSNVMix2 in our cancer analysis pipeline for somatic SNP/Indel detection. For copy number aberration (CNA), we incorporate Control-FREEC, BIC-seq, and OncoSNP-SEQ. We have also developed a brand-new ranking method, OncoRank, which ranks the mutations in each tumor-normal study based on a number of features such as consensus calls, sequencing features (e.g. coverage, strand bias, and allele frequency) and knowledge base (e.g. dbSNP, COSMIC, and Cancer Gene Census). Here, we present our results for a breast cancer cell line (HCC1143) where the samples were pure, as well as two challenging leukemia (AML) data sets where heavy cross-contamination was expected.

For the HCC1143 study, a total of 189,142 somatic single-point mutation candidates were called, of which 90 candidates were COSMIC v54 entries and 38% (71,879) were dbSNP138 entries. Among those, 14,418 candidates were reported by all four detectors with 81 in COSMIC and 8.8% in dbSNP. Using our ranking method, we reported a similar size of call set with a score of at least 6/10. The call set had 13,146 candidates with 87 in COSMIC and 2% in dbSNP, indicating our method was more sensitive (with more COSMIC mutations) and more specific (with fewer known common calls) than the consensus of all four detectors. With our copy number analysis, we correctly detected the ploidy of the HCC1143 cell line as 3.8. Copy number gains were also detected by all the three detectors for breast cancer associated genes *GATA3*, *CCND1*, and *AKT1*. For the AML study, our pipeline successfully reported top-scored causative variants such as the *NPM1* insertion, *DNMT3A* mutation, and *KRAS* mutation in two leukemia patients. These results showed that our approach can increase both the sensitivity and specificity of somatic mutation detection, and is robust enough to handle vastly different data types from pure tumor cell line to challenging liquid cancers.

1510M

Accurate Randomized Dimension Reduction with Applications to Linear Mixed Model Corrections of eQTL Data. *G. Darnell[1], S. Georgiev[2], S. Mukherjee[3], B. Engelhardt[3,4].* 1) Computational Biology and Bioinformatics, Duke University, Durham, NC; 2) Genetics Department, Stanford University, Palo Alto, CA; 3) Department of Statistical Science, Duke University, Durham, NC; 4) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC.

Expression quantitative trait loci studies have power to detect many causal variants associated with gene expression. In many studies, particularly those of model organisms such as mice, complex population structure can overwhelm true signal and cause spurious associations. Linear mixed models have been shown to partition data in genome-wide association studies into fixed and random effects, accurately accounting for population structure and sample substructure (Zhou et al. 2012). Using sophisticated dimension reduction based on randomized singular value decomposition (SVD), we show it is possible to correct for confounding in many eQTL studies where computational burden made it previously infeasible. We show up to a 50% increase in speed when using our method for randomized SVD compared to optimized out of the box libraries for SVD, while maintaining accuracy above 99%. Our method applied to many eQTL datasets from the GTEx project yields results significantly faster than current methods such as GEMMA (Zhou et al. 2012).

1511T

Evaluation of a genotyping array design for tagging common variation across Africa. *T. Carstensen[1,2], GDAP Investigators.* 1) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) Cambridge University, Cambridge, United Kingdom.

Background: Most current genotype arrays have been ascertained largely on European populations, and are not designed to capture genetic variation across several populations in Africa. Designing genotype arrays to efficiently capture common genetic variation in Africa could have important implications for large-scale GWAS, including power to identify novel susceptibility loci in the region. Here, we explore the first such design to capture common genetic variation across Africa. Methods: For tagging, we used whole genome sequence (WGS) data generated through the African Genome Variation Project (320 samples sequenced at 4x coverage from Ugandans, Zulu in South Africa and Ethiopians) and phase 1 of 1000 Genomes (97 Luhya from Kenya and 88 Yoruba from Nigeria). To design the African SNP array we used a greedy pairwise tagging algorithm, written to be computationally efficient and scalable across the whole genome. Tagging was carried out across populations using a window size of 500kbp flanking each SNP, at an linkage disequilibrium (LD) threshold of 0.80 to capture variants with a MAF≥0.05. To examine efficiency of the genotype array, we carried out imputation using the tagged markers, and calculated the correlation between imputed genotypes and WGS genotypes. We calculated coverage across the genome as the proportion of common variants imputed with r2≥0.80 across all populations. Results: Using our tagging algorithm, we show that approximately 2.7 million SNPs are required to tag all common variation across the five African populations. With the most informative 1 million tagging SNPs identified using the tagging algorithm, we were able to tag more than 80% of common variation across 5 populations in Africa. Conclusions: We present the first exploration of a genotype array design to capture common variation across Africa. We show that while tagging common variation across Africa requires a reasonably large number of variants, even 1 million tagging variants can capture a substantial proportion of common genetic variation across African populations. Our findings suggest that designing a genotype array for large-scale GWAS in Africa is feasible, and can be made cost-effective so as to scale to large sample sizes. We envisage that using hybrid approaches for design that include cyclical tagging and imputation will further improve the efficiency of such an array.

1512S

StrandOmics: Accelerating clinical interpretation and reporting though integration of genomic, structural, functional and phenotypic information. *S. Agrawal, A. Ghosh, S. Krishna, S. Sankaran, A. Mannan, V. Veeramachaneni, R. Hariharan.* Strand Life Sciences, Bangalore, Karnataka, India.

Increasing demand for NGS tests in our clinical testing laboratory has driven the need for a clinical interpretation process with three key features. The first is *comprehensiveness*: handling of different types of panels, disease indications, variants, phenotypes, and family situations, a plethora of bioinformatics predictions and knowledge of genes and variants reported in literature. The second is *ergonomics*: the ability of an interpretation team to collaboratively assess literature and bioinformatics evidence, review it and then create accurate reports systematically on a large number of cases. The third is *efficiency*: minimising the time needed to go from reads to report. Thus StrandOmics, a clinical interpretation and reporting platform was created. We illustrate the clinical utility of this platform using some interesting cases comprising:

a) analysis of variants from whole-genome, whole-exome and targeted panel data

b) analysis of single individual as well as more complex multi-individual cases

c) molecular diagnosis for affected individuals and risk prediction for unaffected individuals

d) troubleshooting false negatives by Sanger sequencing

StrandOmics combines knowledge from our internal curated literature content with various publically available data sources and bioinformatics prediction tools to integrate genomic, phenotypic, structural and functional information. This integrated knowledge is then used to automatically prioritize a list of variants based on ACMG guidelines, the inheritance model and disease phenotypes. All the information needed for interpreting the clinical significance of a variant is displayed in a user-friendly manner, which greatly reduces the time for interpretation. Our team of clinical interpreters has thus far used StrandOmics to analyze over 500 cases with a variety of clinical indications and phenotypes. After its introduction, the average end-to-end time for interpretation, review and reporting was reduced by 67%. In addition, StrandOmics also supports the ability to share curated variant information with the medical community through an integrated interface for direct submissions to ClinVar. This feature not only makes our laboratory workflows comprehensive, ergonomic, and efficient, but also collaborative.

1513M

Performance survey of protein mutational prediction methods. *D.A. Baird[1], A. Bierzynska[2], I.N.M. Day[1], M.A. Saleem[2].* 1) Bristol Genetic Epidemiology Laboratories (BGEL), School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN; 2) Academic Renal Unit, School of Clinical Sciences, University of Bristol, Dorothy Hodgkin Building, Whitson Street,Bristol, BS1 3NY.

Point mutations are the most common form of variation within the human genome. Next Generation Sequencing (NGS) has accelerated research into the effects of these mutations in human disease. The major and fundamental challenge facing researchers in this area is determining the pathogenic mutations amongst a background of many neutral variants. Protein prediction tools are often used as a guide for this. Due to the diverse ways in which an amino acid residue change may impact protein structure, mutational prediction is challenging and a wide range of tools using a variety of computational approaches exist. There is a lack of literature jointly reviewing the performance of tools, with most comparisons done on an individual ad hoc basis during the development stage of the tool. This study extends and updates the surveys of Thusberg et al [1] and Shihab et al [2], by comparing the performance of 9 recent and commonly used web based prediction tools. Mutations annotated as neutral and pathogenic in SwissProt was selected as the benchmark for the performance comparison. The benchmark dataset was submitted to each prediction tool and performance statistics (accuracy, specificity, sensitivity and Matthews Correlation Coefficient (MCC)) and ROC curves calculated. The overall best performing tools were Condel 2 and FATHMM. 1.Thusberg, J., Olatubosun, A. & Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. Hum. Mutat. 32, 358-368 (2011). 2.Shihab, H. A. et al. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. Hum. Mutat. 34, 57-65 (2013).

1514T

A bioinformatics approach to prioritizing candidate explanatory variants in whole genome sequences from patients affected with rare diseases. *D. Bodian[1], B. Solomon[1], A. Khromykh[1], R. Iyer[1], R. Baveja[2], J. Vockley[1], J. Niederhuber[1].* 1) Inova Translational Medicine Institute, Falls Church, VA; 2) Fairfax Neonatal Associates PC, Inova Children's Hospital, Falls Church, VA.

Nextgen sequencing technologies are making significant advances in the study of rare diseases, contributing both to the discovery of disease-associated genes and to the identification of likely causative genetic variants in affected individuals. A significant challenge in this approach is identification of the clinically relevant mutation(s) among the millions of variants uncovered by whole genome sequencing. As part of a research study of rare congenital anomalies, we are evaluating the effectiveness of publicly available bioinformatics methods in prioritizing causative mutations. Initially we are testing a genotype to phenotype approach, in which variants are filtered based on inheritance patterns, population frequency, and predicted deleteriousness. The highest ranked variants are then evaluated for congruency with the clinical phenotype. Significant reductions in the number of variants requiring manual assessment come from adapting the tools for combined quality filtering and inheritance-based analysis of family trios, and from comparison to our internal database of whole genome sequences from an ethnically diverse cohort of >1,500 healthy family trios representing >80 countries of birth. Implementation of this approach is in progress; preliminary analyses revealed causative variants for 5 probands enrolled in our congenital anomalies study, all of which have been validated.

1515S

Comparisons on whole exome capturing homogeneity among different versions of capturing kits and populations. *M.G. Borges[1,2], C. Rocha[1,2], B.S. Carvalho[1,2], I. Lopes-Cendes[1,2].* 1) State University of Campinas, Campinas, Sao Paulo, Brazil; 2) Brazilian Institute of Neuroscience and Neurotechnology - BRAINN, Brazil.

Coding regions of the genome, which correspond to less than 2% of its entirety, are known as exome: the portion of the human genome believed to concentrate most of the disease-causing mutations. It has been shown that when performing whole genome sequencing of a single individual we expect to find approximately three million variants; however, if one focus only on coding regions this number drops to less than twenty thousand. In this context, exome sequencing is thought to be a cost-effective strategy for high performance molecular diagnosis applied to genomic medicine and large scale population studies. However, to best apply this approach to different populations it is important to determine whether ethnic differences can affect sensibility and specificity of the method. These effects could have a significant negative impact in exome capturing efficiency and homogeneity among different populations, leading to the necessity of protocol adjustments when analysing patients from a mixed ethnic background. In order to investigate the impact of ethnicity in exome homogeneity, we selected 120 individuals from the 1000 Genomes Consortium. These exomes were sequenced and aligned at three different time-points. For each of those temporal subsets of data there are 40 individuals: ten of each of the four considered populations (ACB - African Caribbean from Barbados; GBR - British in England and Scotland; YRI - Yoruba in Nigeria; JPT - Japanese from Tokyo). We obtained the mean homogeneity for each exon within the human exome as well as the homogeneity at specific genes considered to be clinically relevant by the American College of Medical Genetics and Genomics, as presented at "ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing". After using multidimensional scaling to reduce dimensionality and capture variability, we identified patterns of exome homogeneity that can be related to time progression and, therefore, protocol advancement. We did not observe evidence for association between ethnicity and mean homogeneity at both levels of data analysis (whole exome and clinically relevant genes). Additionally we perceived that for the clinically significant genes the homogeneity density tends to assume a more normalized distribution as a newer version of capture kits is used. This heterogeneity over time can represent an important issue in large sequencing projects, which are to be developed over the next years.

1516M

Prediction consequences of amino-acid substitutions in the IDS gene using in silico tools. *A.C. Brusius-Facchin[1], R. Giugliani[1,2,3], S. Leistner-Segal[1,2].* 1) Hospital de Clinicas de Porto Aelgre, Porto Alegre, Brazil; 2) Programa de Pós-Graduação em Medicina: Ciências Médicas,UFRGS; 3) Departamento de Genética,UFRGS.

Computational methods are used to predict the molecular consequences of amino-acid substitutions on the basis of evolutionary conservation or protein structure, being relevant tools for establishing the consequences of novel mutations in heterogeneous genetic diseases such as Hunter Syndrome (MPS II).We used four different prediction computer programs that are based on evidence of evolutionary conservation of amino-acids, identification of amino acid positions, sequence homology, protein folding, crystal structure and information from a database of hotspot mutations in a specific protein, in order to analyze 17 novel missense mutations found in the IDS gene. PANTHER scores of -3 infer a 50% probability of being deleterious, while -12 is associated with a 100% probability of being deleterious. The POLYPHEN scoring predicts three outcomes for mutations: 'benign' (most likely lacking any phenotypic effect), 'possibly damaging' (may affect protein structure or function) and 'probably damaging' (high degree of confidence that protein structure function will be affected). The SIFT scores range from zero to one, with zero predicted to be the most deleterious mutation and one the least deleterious. PMUT classifies as neutral or pathological. The results obtained after in silico analysis was: PANTHER, mutations p.D45V; p.D45G; p.S61Y; p.Q80R; p.C84Y; p.160H; p.L314H; p.D308H; p.D334Y; p.D334V and p.H342P showed scores between -3 to -10 and the mutations p.Q81Y; p.R95S; p.H138Y; p.N265K; p.E344K and p.V503D, showed scores below -3, SIFT: The mutations p.D45V, p.D45G, p.S61Y, p.Q80R, p.C84Y, p.H138Y, p.P160H, p.L314H, p.D308H, p.D334Y, p.D334V and p.H342P were predicted as damaging and the mutations p.Q81Y, p.N265K, p.V503D, p.E344K were predicted as tolerated, Polyphen: the mutations p.D45V, p.D45G, p.S61Y, p.Q80R, p.Q81Y, p.C84Y, p.R95S, p.H138Y, p.P160H, p.L314H, p.D308H, p.D334Y, p.D334V and p.H342P were predicted as probably damaging and the mutations p.N265K, p.V503D, p.E344K were predicted as possibly damaging, PMUT: analysis predicted that all mutations were neutral except for p.C84Y that was predicted as pathological. Three softwares used showed agreement to predict the molecular consequences and can be considered good tools to characterize the novel alterations found in the IDS gene. The mutations p.N265K, p.E344K and p.V503D in all softwares did not show severe consequences in the protein which should be correlated with a specific (mild) phenotype.

1517T

Investigating the relationship between allele frequency of benign variants used in training mutation impact predictors and their stringency at calling deleteriousness. *A. Carroll[1], A. Sidow[2, 3].* 1) DNAnexus, Mountain View, CA; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Department of Pathology, Stanford University, Stanford, CA.

The growing use of sequencing to identify variants of interest in research and clinical data has generated large databases of well characterized mutations in the human population (e.g. dbSNP, OMIM, and ClinVar). In addition, due to the large amount of rare variation in the human population, variants of unknown, but potentially critical, significance are often identified. When novel variants are discovered, researchers and clinicians often look to mutation impact predictors to gain insight. A wide array of mutation impact predictors have been developed (for example SIFT, Polyphen, GERP, MutationTaster, MutationAssessor, or CADD, amongst many others). During the development of these predictors, scientists use sets of previously characterized benign variants as either explicit or implicit training sets to build their predictor models. Similarly, in the assessment of predictors, accuracy is assessed on sets of known disease-causing mutations and mutations existing in the human population considered to be benign. We demonstrate that the choice of benign (or negative) variants used in training and assessment of these models has a substantial impact on predictor results and accuracy assessments. Specifically, choosing a negative set of variants that exist in the population at a high allele frequency results in predictors that are more aggressive at calling variants as damaging, and give more favorable assessments to predictors which aggressively call damaging variants. Conversely, choosing a negative set of variants that exist at low frequencies results in more stringent prediction of damaging mutations and rewards more stringent callers. We show the fraction of variants predicted as damaging by a number of popular predictors across the human exome. In addition, we use machine learning approaches to take the output of a predictor and tune it new training sets. With this, we have generated results for each predictor that represents training using various negative sets with different allele frequencies. This allows investigators to choose predictors tuned for varying levels of stringency. These datasets are available for download and a tool to annotate a VCF with this data is available for use on DNAnexus.

1518S

Developing a new approach to transcriptomic characterization of mesial temporal lobe epilepsy models through next-generation sequencing. *B.S. Carvalho, A.H.B. Matos, A.S. Vieira, I. Lopes-Cendes.* Department of Medical Genetics, State University of Campinas - Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, SP, Brazil.

Mesial temporal lobe epilepsy (MTLE) is the most prevalent type of epilepsy in humans. It is a chronic condition characterized by recurrent seizures and caused by a number of factors, including genetic predisposition. It is often refractory to treatments, drawing a significant amount of attention from the medical community. In this study, we used recently described MTLE models (pilocarpine-treated rats without status epilepticus), whose lesions are better related to those found in patients, to improve our understanding regarding epileptogenesis in MTLE. We use results from high-throughput sequencing to describe gene expression and alternative use of exons. The statistical methodology is based on count tables, rather than on RPKM/FPKM measurements, combined with negative-binomial models to account simultaneously for technical and biological variabilities. This provides accurate estimates for changes in expression and alternative exon usage when comparing affected to control individuals. Through the use of statistics that were developed specifically to compare expression levels of individuals from distinct biological statuses, we improve our findings by using computational methods to explore sets of genes whose frequencies do not match expectations, being identified as systematically over- or under-represented (OR/UR). We investigate pathways and gene ontologies, allowing for a better understanding of gene networks and enhancing the findings obtained through the simplified models used for the gene expression estimation. We identified ontologies that presented significant association to epilepsy at different levels: 1) molecular (N-acetylglucosamine-phosphate-deacetylase activity/ATP binding); 2) biological processes (negative regulation of signal transduction/macromolecule catabolic process); and 3) cellular components (ion channel complex/intracellular organelles). Our data analysis strategy focuses on the use and development of open-source and high-performance tools, improving accessibility and allowing for greater opportunities of improvement and reproducible research. Combining these factors with the expertise of our group members, we believe to be in a unique position that entitles us to A) properly compare these to results generated with other models; B) confirm these findings through further experiments; C) contrast these findings to those identified on patients; and D) translate these results to the clinic. (Supported by FAPESP).

1519M

Comparative analysis of computational pipelines for RNA sequencing in genetical genomics studies. *J. Chen[1,2], C. Ye[1], A.C. Villani[1], M.N. Lee[1,3,4], T. Raj[1,3,5], P.L. De Jager[1,3,5], C. Benoist[3,5], T. Bhangale[6], W. Ortmann[6], T. Behrens[6], N. Hacohen[1,3,4], A. Regev[1,7,8].* 1) Broad Institute, Cambridge, MA; 2) Division of Health Sciences and Technology, MIT, Cambridge, MA; 3) Harvard Medical School, Boston, MA; 4) Massachusetts General Hospital, Boston, MA; 5) Brigham and Women's Hospital, Boston, MA; 6) Genentech Inc., South San Francisco, CA; 7) Department of Biology, MIT, Cambridge, MA; 8) Howard Hughes Medical Institute.

RNA-sequencing (RNA-seq) has emerged as an important tool for profiling gene expression variation because it allows a comprehensive characterization of the transcriptome, including transcript levels, splicing, and allele specific events. Several large scale studies have now used RNA-seq to identify the genetic basis of gene expression variation in humans across different cell types and cell stimuli. Although a number of computational approaches have been developed to (i) align reads, (ii) assemble or annotate transcripts, and (iii) quantify transcript abundances, different computational pipelines aimed at defining quantitative traits from RNA-seq have not been systematically compared.

Here, we evaluate three pipelines for identifying the genetic basis of gene expression in humans. We: (1) align to the transcriptome and estimate fraction of transcripts using RSEM, (2) align to the genome, assemble and estimate fraction of transcripts using TopHat and Cufflinks, and (3) align to the genome, and directly count reads using HtSeq. We apply each pipeline to generate expression or isoform usage quantitative traits for three datasets: (i) the GEUVADIS lymphoblastoid cell line data across 462 individuals, (ii) whole blood across 922 individuals using single end RNA-seq, and (iii) ImmVar dendritic cell stimulation response in 576 samples. We evaluate the performance of each pipeline based on its ability to produce a reproducible set of genetic associations to expression (eQTL) or isoform ratio (irQTLs) within each dataset through cross validation or across datasets.

In the well annotated human genome, the simplest pipeline, which uses RSEM to estimate relative abundances from the annotated transcriptome, produced highly reproducible expression levels as compared to microarrays and Nanostring measurements on the same samples. Furthermore, within each pipeline, quantitative traits based on estimations of fraction of transcripts, such as transcripts per million and isoform ratio, produced highly reproducible eQTLs and irQTLs across all three datasets. Overall, our analysis highlights that a computational pipeline based on estimating fraction of transcripts from multiply mapped reads is a robust method for discovering variants that control expression and splicing from RNA-seq data.

1520T

Bioinformatic analysis of novel pathogenic missense mutation of ARSB gene in a colombian patient whit Maroteaux -Lamy. *G. Giraldo[1], P.A. Ayala[1], J. Acosta[2], R. García[2], J.C. Prieto[1], J. Gonzalez[3].* 1) Instituto de Genètica Humana, Pontificia Universidad Javeriana, Bogotà, Bogotà D.C, Colombia; 2) Instituto de Investigación en Nutrición, Genética y Metabolismo Universidad el Bosque, Bogotá D.C, Colombia; 3) Depto de bioquimica, Pontificia Universidad Javeriana, Bogotá D.C Colombia.

Introduction: Maroteaux -Lamy - MPS VI (OMIM# 253200) is a lysosomal storage disease characterized by systemic clinical manifestations and significant functional compromise. The reported worldwide incidence is 1:248.000 to 1:300.000 live births [1] and in Colombia 27 cases are known, 10 belong to indigenous groups [2]. To date, 133 mutations have been reported: 100 missense / nonsense, 9 in Splicing site, 17 small deletions, 3 small insertions in the Arylsulfatase B (ARSB) gene. The ARSB mutations cause loss of enzyme activity. To understand the structural basis for the disease, we approximated bioinformatics methods to analyze one ARSB missense mutations at both the sequence and structural level. Material and methods: In a previous study, 13 patients were analyzed, to date, sequencing for 4 exons in 7 patients have been performed, finding in one patient , a variant of exon 2, not previously reported , c.1618A > C (p.H111P). To evaluate the effect of mutation, different programs were run like Polyphen2, Mutation taster, Mustab, Panther and Provean. The amino acid sequence of human ARSB was taken from Uniprot entry P15848 and the crystallographic coordinates were obtained from the PDB database (ID: 1FSU) To prediction of the effect of point mutations on protein stability was performed using structural models generated by FoldX. The analysis of residue conservation was conducted with the ConSurf server and for illustration of conserved amino acids in the sequence we used MultiDisp. Results and discussion: We evaluated the effect of the change c.1618A>C (p.H111P) in the function on the protein, finding a highly deleterious and a potentially pathogenic variant. Histidine is a α-amino acid with an imidazole functional group and is considered a proteinogenic amino acid. This α-amino acid is involved in the catalytic reactions of enzymes. The proline change interrupts the habitual conformation of the side chain and makes the chain direction changes abruptly. The presence of proline disrupts the formation of any regular repeating structure Using bioinformatics programs was observed that this amino acid is conserved in all species; this variant causes a conformational change in the protein and affects the catalytic site seriously, responsible for the enzyme deficit.

1521S

Simultaneous detection of copy number variations (CNV) and point mutations with next generation sequencing (NGS) using Agilent HaloPlex custom designs. *C. Haag[1], K. Hauschulz[2], J. Strub[3], E. Schulze[1].* 1) Endocrine Practice and Medical Genetics Laboratory, Heidelberg, Germany; 2) Agilent Technologies Sales & Services GmbH & Co. KG, Germany; 3) JSI medical systems GmbH, Kippenheim, Germany.

Copy number variation (CNV) is a form of structural variation in the genome. Usually, CNV refers to the duplication or deletion of DNA segments larger than 1 kbp. Copy number variations have been recognized as pathogenic mutations for many years, for example in mental retardation disorders or Duchenne muscular dystrophy (DMD), where single or multiple exons are deleted or duplicated, respectively. Alterations in DNA copy number are also a common feature in different cancers and the detection of these changes shows promise for the diagnosis of a disease and also for therapeutic or prognostic purposes. Two different methods are traditionally used for mutation detection, one for CNVs and one for point mutations. For the detection of copy number changes in selected regions array Comparative Genomic Hybridisation (aCGH) or PCR-based methods such as Multiplex Ligation-dependent Probe Amplification (MLPA) are well established, whereas Sanger-based sequencing or next generation sequencing are the standard methods for the detection of point mutations and small deletions or insertions. We used the Agilent HaloPlex Target Enrichment system (on the Illumina MiSeq platform) for the combined detection of point mutations and copy number variations. Copy number variations are detected by comparing the coverage of regions in the sample of patients with control samples. The copy number variation analysis was performed with the CNV function of the Sequence Pilot software (JSI Medical Systems GmbH, Kippenheim, Germany). We used three different Haloplex custom designs including 21 genes for hereditary endocrine diseases like pheochromocytoma or multiple endocrine neoplasia for the validation of the Haloplex assay. Altogether, we analyzed 45101 bp covered by 2999 amplicons. The Haloplex assay was validated using known positive control samples comprising samples with point mutations, indels as well as exon duplications/deletions. All mutations were successfully identified with a high coverage (>1000 fold average exon coverage) and 99.2% of the targeted bases were covered with a least 50 reads. Regions with no coverage account for 0.5%. In summary, Haloplex custom designs can be used as a targeted NGS resequencing approach that facilitates the detection of point mutations, indels and duplications/deletions in parallel.

1522M

Gene-based burden analysis of imputed low frequency variants identifies associations with LDL in an African American cohort. *H. Hakonarson[1,2], A. Krämer[3], D.R. Richards[3], P.M.A. Sleiman[1,2].* 1) Center for Applied Genomics, Children's Hosp Philadelphia, Philadelphia, PA; 2) The Perelman School of Medicine, University of Pennsylvania Philadelphia, PA, USA; 3) QIAGEN Silicon Valley, CA, USA.

Genome wide association studies (GWAS) have been very successful in identifying loci associated with complex disease, however, the genotyping arrays were not designed to capture low frequency and rare variation. As a result, the role of variants in that category of allele frequency remains largely unexplored in complex disease. One possibility would be to generate sequencing data on a large number of samples, however, the cost of large-scale studies remain prohibitively expensive. Genotype imputation presents a viable alternative for sample sets with existing genotypes from dense whole-genome arrays. Data generated from large sequencing projects such as the 1000 Genomes project and the NHLBI exome sequencing project (ESP) can be used as reference datasets to impute missing variation into the sample genotype data. Recent improvements in imputation algorithms and the expansion of reference datasets have improved the accuracy of imputation for even low minor allele frequency variants. Imputed variants can then be assessed against binary phenotypes or quantitative laboratory values derived from patients' electronic medical records. The Center for Applied Genetics (CAG) maintains a biorepository of over 150,000 genotyped samples, 45,000 of which are pediatric samples randomly recruited from the Children's Hospital of Philadelphia (CHOP) with complete electronic medical records. As a proof of principle, we imputed missing variants into a subset of just under 100,000 samples that had been previously genotyped on the Illumina BeadChips from the CAG biorepository. Imputation of untyped markers (~39M) was carried out using IMPUTE2 against the 1KGP Phase I integrated variant set after prephasing with Shapeit. Imputed genotypes were converted to a binary encoding and uploaded to QIAGEN's Ingenuity Variant Analysis platform for annotation and gene-based analysis using the SKAT algorithm. As proof of principle we carried out an association against EMR-derived LDL values from 2,500 African American children. The results of the gene-based analysis showed highly significant association of the APOE (P 3.44×10-10) and PCSK9 (P 3.34×10-7) genes both of which have recently been associated with LDL in African Americans through genotyping on exome array chips. Imputation of low frequency and rare variation into genotyped data thus represents a viable alternative to sequencing and re-genotyping on exome array chips for the study of this class of variation in complex disease.

1523T

NIH Genetic Testing Registry (GTR): A data mine available through programmatic access. *B. Kattman, A. Malheiro, J. Lee, D. Maglott, V. Hem, M. Ovetsky, G. Song, C. Wallin, K. Katz, R. Villamarin-Salomon, B. Gu, S. Chitipiralla, W. Rubinstein.* NCBI/NLM/NIH, Bethesda, MD.

The NIH Genetic Testing Registry (GTR; www.ncbi.nlm.nih.gov/gtr/) houses detailed information on more than 17,000 genetic tests voluntarily submitted by laboratory test providers. Interest in programmatic access to GTR data has burgeoned as GTR has become the world's most comprehensive repository of publicly available data about genetic tests. The GTR website supports interactive data access for a high volume of daily users. Recently the entire GTR dataset was made available through XML files from the ftp site, as well as summary data via NCBI's application programming interface, E-utilities.

The GTR includes germline and somatic tests using molecular, cytogenetic and/or biochemical methodologies. Tests for drug responses, and complex panels utilizing next-generation sequencing or array technologies, are in scope. As of June 2014, GTR has submissions from more than 400 testing laboratories from 39 countries. In the past 7 months there has been a 66% increase in the number of registered tests. Registered tests evaluate 4,456 conditions, with molecular tests targeting 3,014 genes (91% of clinically relevant genes reported by ClinVar). Submission of complex tests is also increasing, with 651 tests that evaluate 5 or more genes. Next-generation sequencing (NGS) is a component of 9.2% of molecular tests. Data surrounding the evidentiary basis of tests are available for a growing number of records: Analytical validity (100%), Target population (36%), Clinical validity (14%), and Clinical utility (15%). Information is also available about proficiency testing, FDA approval/clearance, laboratory certification and much more.

GTR has long maintained an FTP site to support unrestricted access to standard terminologies and provide summary data (ftp://ftp.ncbi.nlm.nih.gov/pub/GTR/_README.html). In response to stakeholders, GTR added the comprehensive extraction of test data as XML. In addition, data are accessible by NCBI programmatic tools (www.ncbi.nlm.nih.gov/gtr/docs/maintenance_use/).

GTR has a mission to improve transparency surrounding genetic testing, and is being sought by a wide variety of stakeholders interested in surveying the genetic testing landscape. This presentation will illustrate how to access GTR data for knowledge discovery.

1524S

Impact of statin on gene expression in human lung tissues. *J. Lane[1], B.Y. Aminou[1], S. vanEeden[2], D. Sin[2], M. Obeidat[2], S. Tebbutt[2], W. Timens[3], D.S. Postma[3], M. Laviolette[1], P. Paré[2], Y. Bossé[1,4].* 1) Institut universitaire de cardiologie et de pneumologie de Québec, Quebec, Quebec, Canada; 2) University of British Columbia Center for Heart Lung Innovation and Institute for Heart and Lung Health, St. Paul's Hospital, Vancouver, Canada; 3) University of Groningen, University Medical Center Groningen, GRIAC research institute, Groningen, The Netherlands; 4) Department of Molecular Medicine, Laval University, Quebec, Canada.

Objectives: Statins inhibit 3-hydroxy-3-methylglutaryl-coenzyme A reductase which is involved in the synthesis of cholesterol and about a 1/3 of COPD patients are actively using statins. Statins may improve the respiratory health of patients with asthma, COPD and lung cancer by reducing the adverse impact of chronic inflammation and cardiovascular co-morbidities. We studied the effect of using statins on gene expression in human lung in order to identify new molecular pathways underpinning the effects of statins. **Methods:** Human non-tumor lung tissues were obtained from patients who underwent lung resection. Gene expression was measured on a custom Affymetrix array in a discovery cohort (n=408) and two replication sets (n=282 and 341). Genes differentially expressed between patients taking or not taking statins were tested by linear regression, adjusting for age, gender and smoking status. A propensity score or substitution variables also served as covariates as appropriate. The results of each cohort were combined in a meta-analysis and biological pathways were studied using Ingenuity Pathway Analysis, DAVID and Gene Set Enrichment Analysis. **Results:** The discovery set includes 267 statin users. Fifteen genes were found significantly up-regulated in the lungs of statin users (FDR < 0.05). Twelve of these genes were replicated in the first replication set, but none in the second (p-value < 0.05). The meta-analysis improved the significance of the 15 up-regulated genes. Biological pathways analyses provided 21 significant pathways (FDR < 0.05). The first pathway suggested that statins up-regulate expression of genes involved in cholesterol synthesis by activating the sterol regulatory element-binding protein. The top 10 pathways affected by statins included genes involved in isoprenoïds synthesis and genes regulated by the cytochrome P450 reductase. Genes in the insulin receptor substrate/protein kinase B pathway were also affected by statins. **Conclusion:** To our knowledge this is the first study to report the effect of statins on the transcriptome of the human lung. The results suggest that statins may improve respiratory health by modulating pulmonary expression of genes involved in biological pathways associated with asthma, COPD and lung cancer.

1525M

A network approach to investigate the respective roles of common and rare variants in Attention-Deficit/Hyperactivity Disorder. *L.A. Lima[1,2], A.C. Feio-dos-Santos[3], D.B. Mariani[2], S.N. Simões[2,4], R.P. Silva[1], R.F. Hashimoto[2,5], L.A. Rohde[6], H. Hakonarson[1], H. Brentani[2,3].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Programa Interunidades de Pós-Graduação em Bioinformática, Universidade de São Paulo, São Paulo, SP, Brazil; 3) Instituto de Psiquiatria, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brazil; 4) Instituto Federal do Espírito Santo, Serra, Espírito Santo, Brazil; 5) Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, SP, Brazil; 6) Departamento de Psiquiatria e Medicina Legal, Faculdade de Medicina, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

Attention-Deficit/Hyperactivity Disorder (ADHD) is the most common psychiatric disorder in infancy and adolescence, affecting 5.8% of children and adolescents in the world. Many studies have been applied trying to investigate the genetic susceptibility of ADHD. The present work aimed at examining the differences and similarities in rare and common variants in this disorder. We used previously published results of a GWAS meta-analysis and four copy number variation (CNV) studies of ADHD involving pediatrics/adolescent Caucasian samples. Using only genes expressed in brain (post-mortem pediatric samples from Brain Atlas), we constructed a protein-protein interaction (PPI) network as a background, using only physical interactions confirmed by at least two sources. Two sets of genes were defined based on the candidate hits from the GWAS meta-analysis and CNV analyses results. These sets of genes served as seeds to construct two independent networks: 1. related to common variants (from GWAS meta-analysis data) and 2. related to rare variants (CNV data). The two networks were constructed with the seed genes and their direct interactions on the background PPI network. The overlapping genes from common and rare variants networks confirmed some known functions involved with ADHD, such as Ubiquitin mediated proteolysis, Neurotrophin signaling pathway, Insulin signaling pathway, Focal adhesion and MAPK signaling pathway. The genes exclusively mapped in the rare network (CNV) were enriched with new candidate functions, as Spliceosome and Long-term potentiation. The network analysis approach can be easily applied to other diseases caused by both rare and common variants.

1526T

Optimizing an imputation panel for admixed Latin American populations. *W. Magalhães[1], N. Araujo[1], T. Leal[1], M. Machado[1], R. Moreira[1], M. Gouveia[1], F. Soares[1], M. Rodrigues[1], F. Kehdy[1], A. Horimoto[5], M. Barreto[2], B. Horta[3], M. Lima-Costa[4], A. Pereira[5], E. Tarazona-Santos[1], The Brazilian EPIGEN Consortium.* 1) Biologia Geral, Federal University of Minas Gerais, Belo Horizonte, Brazil; 2) Universidade Federal da Bahia, Brazil; 3) Universidade Federal de Pelotas, Brazil; 4) Fundaçao Oswaldo Cruz, Centro de Pesquiça René Rachou, MG, Brazil; 5) Instituto do Coracao, Universidade de Sao Paulo, Brazil.

As part of the largest Latin-American genomic initiative, we studied three Brazilian longitudinal population cohorts: Salvador-Bahia (n=1309), Bambui (n=1442) and Pelotas (n=3736) from Northeast, Southeast and Southern Brazil respectively. We genotyped the Omni2.5M-Illumina for the 6487 individuals, the Omni5.0M-Illumina for 265 individuals and sequenced 30 genomes (average coverage: 42X). Admixture is an important feature of the Brazilian population, which affects the genotype imputation process and must be considered as a key point, producing a changed linkage disequilibrium when compared to parental populations. In order to overcome the former point, using individuals with the same ancestry background, we combine our high-density array (Omni5.0M) and the whole genome dataset to propose a new reference imputation panel for Latin-American admixed populations. The software Shapeit for phasing followed by the software IMPUTE2 for imputation. We evaluated the accuracy of our panel with the public panel available by the phase I of the 1000 Genomes Project, wherein 1,092 individuals were genome and exome sequenced and genotyped to produce an integrated variant call set. Our preliminary results (CHR22) suggest that the effect of ancestry background and the size of the reference panel play important role on the accuracy of the imputed genotypes.

1527S

Functionally characterizing common variants associated with psychiatric disorders. *J. Moore, Z. Weng.* Program in Bioinformatics & Integrative Biology, UMass Medical School, Worcester, MA.

Schizophrenia, bipolar disorder, and major depressive disorder are psychiatric disorders that affect millions of people every year. The etiology of these disorders in unknown, though each have strong hereditary components. Genome wide association studies associated over 400 single nucleotide polymorphisms (SNPs) with these disorders with a majority of these variants lying in non-coding regions of the genome. In order to characterize the function of these non-coding SNPs, we used epigenomic data from the Encyclopedia of DNA Elements (ENCODE) Consortium and the Roadmap Epigenomics Project to define regulatory regions overlapping each of the variants. In agreement with previously reported results, we found a majority of variants and their linkage disequilibrium (LD) partners overlapped active epigenomic marks. While a majority of the SNPs intersected tissue specific marks, a subset of these variants overlapped marks from hundreds of cell lines. Additionally, we analyzed transcription factor binding motifs, which overlapped the associated SNPs and predicted the effect size of allele change. Most SNPs did not dramatically change the probability of observing the motif, but there were almost 500 loci with allele specific motifs. These motifs corresponded to transcription factors involved in neural pathways as well as inflammatory and immune response. Additionally, our functional predictions were validated by performing allele specific regulation analysis. We remapped reads from ENCODE and Roadmap datasets to observe whether the change in allele affected binding of transcription factors, DNase I hypersensitivity, and gene expression. We then predicted gene targets for these SNPs by correlating DNase I signal across multiple tissue types. These target genes were enriched in similar pathways as the motifs. Several of these pathways were connected through the regulation of cytokines such as transforming growth factor beta and interleukin 2, both of which were previously reported as dysregulated in psychiatric disorders.

1528M

A network-based approach to dissect the cilia/centrosome complex interactome. *M. Morleo[1], R. Amato[1], L. Giaquinto[1], D. di Bernardo[1,2], B. Franco[1,3].* 1) TIGEM, Telethon Institute of Genetics and Medicine, Naples, Italy; 2) Department of Computer and Systems Engineering, University of Naples "Federico II", Naples, Italy; 3) Department of Medical Translational Sciences, Federico II University, Naples, Italy.

Background Cilia are microtubule-based organelles protruding from almost all mammalian cells that, when dysfunctional, result in genetic disorders called "ciliopathies". High-throughput studies have revealed that cilia are composed of thousands of proteins. However, despite many efforts, much remains to be determined regarding the biological functions of this increasingly important complex organelle. Results We derived an online tool from a systematic network-based approach to dissect the cilia/centrosome complex interactome (CCCI). The tool is able to integrate all current data available into a model which provides an "interaction" perspective on ciliary function. We generated a network of interactions between human proteins organized into functionally relevant "communities", i.e. a group containing genes that are both highly inter-connected and strongly co-expressed. We then combined sequence and co-expression data in order to identify the transcription factors responsible for regulating genes within their respective communities. Our analyses revealed communities significantly specialized for delegating specific biological functions such as mRNA processing, protein translation, folding and degradation processes that had never been associated with ciliary proteins until now. Conclusions CCCI will allow us to clarify the roles of previously unknown ciliary functions, elucidate the molecular mechanisms underlying ciliary-associated phenotypes, and apply our knowledge of the functional roles of relatively uncharacterized molecular entities and disease phenotypes to new clinical applications.

1529T

Identification and Clinical Assessment of Deletion Structural Variants in Whole Genome Sequences of Acutely Ill Neonates. *A.C. Noll[1,2], L.D. Smith[2], N.A. Miller[2], C.J. Saunders[1,2], I. Thiffault[2], A. Newton[1], K. Detherage[1], J. Hoang[2], L.K. Willig[2], L.D. Cooley[2], S.D. Fiedler[2], E.G. Farrow[2], S.F. Kingsmore[1,2].* 1) University of Kansas Medical Center, Kansas City, KS; 2) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO.

Effective management of acutely ill newborns with likely genetic diseases requires rapid and comprehensive identification of causative haplotypes. While whole genome sequencing (WGS) can identify pathogenic nucleotide variants in newborns in less than 50 hours, deletion structural variants (DSVs) >50 nucleotides are also an important component of the mutation burden for genetic diseases and are missed by many variant detection pipelines. Concomitant identification of DSVs in rapid WGS data for diagnosis of acutely ill newborns would be highly valuable. Here we describe the development of DSV detection methods that combine consensus calls from two WGS DSV detection tools (Breakdancer and GenomeStrip) with a novel filtering and clinical strategy. WGS simulation data demonstrated Breakdancer and GenomeStrip consensus calls had 83% sensitivity, 99% positive predictive value, and high precision. Upon inspection of read depth, and alignment distance of paired reads in the integrated genome viewer (IGV), DSV consensus calls overlapping with SNP array calls were found to be 95% true positive. Consensus DSV calling with parameterized filtering was implemented in an 8-hour computational pipeline called SKALD (Screening Konsensus and Annotation of Large Deletions). IGV evaluation of SKALD results in a tetrad demonstrated more than 80% were true positives, however the sensitivity was low. To assess the clinical utility of SKALD, WGS from 12 familial trios with an acutely ill newborn proband in which causative nucleotide variants had not been identified were analyzed. A heterozygous deletion of exons 1-3 of MMP21 (NC_000010.11 g.127,460,714_127,461,028del) was found in trans with a heterozygous frameshift deletion (p.Met122SerfsX55) in two siblings with transposition of the great arteries and abdominal heterotaxy (TGAH). These variants are expected to be deleterious to MMP21, resulting in a truncated or degraded protein. Mmp21 -/- mice exhibit TGAH, supporting MMP21 as a novel disease gene. In a newborn female with dysmorphic features, ventricular septal defect, and persistent pulmonary hypertension, SKALD identified the breakpoints of a heterozygous, de novo 1p36.32p36.13 deletion (NC_000001.11 g.4,848,728_18,503,068del). In sum, SKALD has the potential to increase the diagnostic yield of WGS in acutely ill neonates and to discover novel disease genes.

1530S

Protein functional domain annotation in single gene association in Parkinson Disease. *K. Nuytemans[1,2], V. Inchausti[1,2], L. Maldonado[1,2], A. Mehta[1,2], E.R. Martin[1,2], G.W. Beecham[1,2], L. Wang[1,2], W.K. Scott[1,2], J.M. Vance[1,2].* 1) Human Genomics, University of Miami, Miami, FL; 2) Miami Udall Center, University of Miami, Miami, FL.

Although whole exome sequencing (WES) allows identification of all variants in the coding regions of the genome, additional analyses are needed to assess the potential impact of any of those variants on the protein and thus potentially the disease mechanism. So far, many studies have used amino acid change prediction programs based on conservation or 3D structure (e.g. SIFT, POLYPHEN). Alternatively, filtering on presence in the protein's functional domains will augment the potential effect of variants to be included in the analysis. Most protein databases currently available, however, do not allow for annotation on the genomic level. As a proof-of-principle we used the Pfam database available as a UCSC track to filter variants for analysis. We performed WES in 410 PD patients and 229 controls. Functional annotation of identified variants for effect in protein and presence in protein functional domains (UCSC Pfam track) was performed through ANNOVAR. Rare variants (MAF<5%) were identified using the 1000Genomes and the NHLBI Exome Sequencing Project datasets. Association with risk for PD was assessed for sets of variants in genes using the optimal Sequence Kernel Association test (SKAT-O). Genes with p-value <0.01 were included in the enrichment analyses in KEGG pathways using WEB-based GEne SeT AnaLysis Toolkit (webGestalt). In total, 37,965 protein sequence altering variants (nonsense or missense) were annotated to a protein functional domain in 9736 genes. The top 10 genes ($p<0.002$) using rare or all variants were similar and included genes previously reported in PD (*SUMF2*) or involved in processes linked to PD (endosomal trafficking genes; *AP5M1*, *VPS4B*, axon guidance; *PPFIBP2*, cell-cell interaction; *PCDHB12*). No KEGG pathways were associated after correcting for multiple multiple testing. Top (nominally significant) pathways included "cell adhesion molecules" and "regulation of actin cytoskeleton", both previously reported in PD pathway analyses. These proof-of-principle analyses confirm the validity of using presence in a protein functional domain as a variant inclusion requirement. The analyses presented here are limited by the data available on genomic level for the Pfam (predicted) protein domain database. Therefore, we will set out to evaluate this approach of protein domain annotation for genomic variants including different algorithms (e.g. TIGR, PANTHER,..) to widen the search for potentially damaging variants.

1531M

Large pedigrees in human sequencing studies: toward a more resolved and accurate picture of genetic disease. *J.A. O'Rawe[1,2], Y. Wu[1,2], H. Fang[1], L.T.J Barrón[1,4], K. Wang[3,5], G.J. Lyon[1,2,3].* 1) Cold Spring Harbor Laboratory , Cold Spring Harbor, NY; 2) Stony Brook University, Stony Brook, NY; 3) Utah Foundation for Biomedical Research, Salt Lake City, UT; 4) Universidad Nacional Autonoma de Mexico, Cuernavaca, Morelos, MX; 5) Zilkha Neurogenetic Institute, Department of Psychiatry and Preventive Medicine, University of Southern California, Los Angeles, CA.

Introduction: While studying larger multi-generational families can increase the accuracy of "whole" genomes by reducing the effect of undetected sequence variation, using a variety of sequencing and analytical pipelines can generate a more comprehensive set of sequence variations. We describe a whole genome sequencing (WGS) study of one large multi-generational family containing two affected male children, aged 10 and 12, with severe intellectual disability, autism-like behavior and very distinctive facial features. Methods: WGS was performed on ten family members using the Illumina HiSeq2000 platform, with four (the two affected boys and their parents) being additionally sequenced using the Complete Genomics (CG) WGS and analysis platform. Illumina reads were mapped to the hg19 reference genome using BWA v.0.6.2-r126, and variant detection was performed using the GATK v. 2.4-9. Additional variant detection procedures included using Novoalign and the FreeBayes caller, Scalpel for insertion or deletion (INDEL) detection, RepeatSeq for INDEL detection in short tandem repeat regions, and the ERDS method for detecting larger copy number variations. Disease variant prioritization was performed using filtering techniques (ANNOVAR, Golden Helix SVS and GEMINI), and a statistical method for identifying disease variants (VAAST). Results: CG WGS covered >85% of the genome and >90% of the exome, both with 20 or more reads. Illumina WGS covered >90% of the genome with 30 reads or more and with >80% of the bases having a quality score of >30. We find a 2 to 5-fold difference in the number of variants detected as being relevant for various disease models when using different sets of sequencing data and analysis pipelines. We derive greater accuracy when more pipelines are used in conjunction with data encompassing a larger portion of the family, with the number of putative de-novo mutations being reduced by 80%, due to false negative calls in the parents. Conclusions: These data are particularly salient in terms of contrasting 'trio' or 'quad' studies with larger family studies, as both 'trio' and 'quad' study designs are limited in their power to resolve genetic variations contributing to disease. We demonstrate increased power to resolve putative disease related mutations using larger family based studies, and explore the feasibility and performance of using a wide variety of informatics methods to elicit a more complete "Whole Genome".

1532T

Novel bioinformatics driven imaging-genetics approach exploring the aetiology of Alzheimer's disease. *S. Patel[1,2,5], MT. Park[3], J. Pipitone[3], MM. Chakravarty[4,5,6], J. Knight[1,2,5], The Alzheimer's Disease Neuroimaging Initiative.* 1) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, Canada; 3) Kimel Family Translational Imaging-Genetics Research Laboratory, the Centre for Addiction and Mental Health, Toronto, Canada; 4) Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada; 5) Department of Psychiatry, University of Toronto, Toronto, Canada; 6) Douglas Mental Health University Institute, Verdun, QC, Canada.

Introduction: Alzheimer's disease (AD) is a devastating illness, affecting over 35 million people worldwide and expected to increase to 115 million by 2050. Our goal is to identify biomarkers for AD-progression based on structural magnetic resonance imaging (MRI) and genetic data using a systematic biological and bioinformatically-driven approach to shed light on the genetic burden associated with AD. **Methods**: T1 weighted MR images were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Hippocampus segmentation was carried out using the MAGeT Brain algorithm on 162 AD subjects, 317 with mild cognitive impairment and 183 healthy controls [Pipitone et al., 2014]. Hippocampal volumes were used for association testing with both genotypes from the Human 610-Quad BeadChip and imputed genotypes (n=5,706,974) and the following covariates were controlled for: gender, age, first principle component from multidimensional scaling, baseline diagnoses, APOE status and intracranial volume. Stratified false discovery rate (sFDR) [Sun et al., 2006] was used to prioritize SNPs in genes selected using Gene Ontology (GO) networks. A list of genes associated with AD was obtained from a previous study by Lambert et al., 2013. Common GO terms within the gene list were identified and Cytoscape was used to visualize the relationships of the GO terms in a network format. The GO network was then pruned to enrich for GO terms containing genes from our list. We then extracted all SNPs from genes associated to all the GO terms in the nervous system development GO network which formed our stratum for sFDR. **Results** : A total of 249,001 out of 5,706,558 SNPs were selected from 1146 genes from the nervous system development stratum to be up-weighted in sFDR. No significant SNPs were found to be associated with hippocampal volume. **Conclusion**: SNPs in the nervous system development stratum list were not significantly associated with hippocampal volume. SNPs in this stratum may not play a specific role in hippocampal volume, for example SNPs in gene region, growth arrest-specific protein 7 is involved in neuronal development and is mainly expressed in mature cerebellar Purkinje cells [Ju et al., 1998]. Further modifications to prune the nervous system development GO network by using the 'Extension' column of the GO database to extract genes annotated to GO terms with an extension of neuro-anatomical cells in the region of the hippocampus.

1533S

Non-coding RNAs and transcription expression underlying neuropathic pain following sciatic nerve injury. *H.B. Raju[1,3], E. Capobianco[3], N.F. Tsinoremas[1,2,3].* 1) Human Genetics, University of Miami, Miami, FL; 2) Department of Medicine, University of Miami, Miami, FL; 3) Center for Computational Science, University of Miami, Miami, FL.

Background: Neuropathic pain (NP) causes damage to the nervous system. It is vital to detect the targets underlying NP, essential for improving the health of millions of people affected. Cellular functions are likely associated with changes in gene expression causal to NP developing over time Objective: Identify, classify non-coding RNAs and detect transcriptional expression for the two different types of neuronal tissues (dorsal root ganglia (DRG) and proximal sciatic nerve (SN)) following sciatic nerve injury. Method: Microarray probe sequences for the differentially expressed genes at 4 0d, 1d, 4d, 7d and 14d post-sciatic nerve injury in rats were obtained for DRG and SN from the gene expression omnibus (GEO) generated by the Agilent-014879 Whole Rat Genome Microarray 4x44K G4131F. Using these datasets and the expression changes at different time points post-nerve injury, we identified and categorized different ncRNA for DRG and SN at 4 different time points nerve injury. In the primary analysis, the extracted probe sequences were aligned against mouse reference to identify the homologous sequences between the two rodent species. In t the secondary analysis, the raw microarray probe sequences were analysed using GeneSpring 12.6 GX to identify differentially expressed genes across the different time points and potentially, understand the gene regulatory effect related to nerve injury. Results: From the primary analysis, we detected increase in the expressions of ncRNAs such as antisense, lincRNA and pseudogenes in DRG tissues from 1d-14d post-injury. In SN tissues, we detected about 190 lincRNAs at 1d which increased to 310 lincRNAs at 7d post-injury. Comparatively in SN, more ncRNA categories (N=15) were identified than DRG (N=22). Discussion: To the best of our knowledge, this is the first report of the detection of different categories of ncRNAs and transcription expression underlying NP following sciatic nerve injury. Interestingly, the increase in the expression levels of long ncRNAs at extended day's post-injury raises many questions related to tissue repair or other mechanisms related to NP.

1534M

Unrevealing the genomic architecture of chromosomal breakpoint region using multiparametric computational approach. *R.M. Rawal.* Dept. of Cancer Biology, The Gujarat Cancer & Research Institute, Ahmedabad, Gujarat, India.

Recurrent non-random chromosomal translocations are hallmark characteristics of leukemogenesis however molecular mechanisms underlying these rearrangements are less explored. The fundamental question is, why and how chromosomes break and reunite so precisely in the genome. Meticulous understanding of mechanism leading to chromosomal rearrangement can be achieved by characterizing breakpoints. To address this hypothesis, a novel multiparametric computational approach for characterization of six most frequent leukemic translocations within and around breakpoint region was performed. To best of our knowledge, this study is unique in finding the presence of Segmental Duplications (SDs) flanking breakpoints of all major leukemic translocation. Breakpoint islands were also analyzed for other complex genomic architecture and physical properties e.g. SIDD, repetitive elements, recombination signal sequence, base composition, Topo-II binding site & restriction endonuclease cleavage site etc. Our study distinctly emphasizes on the probable role of SDs and various genomic features in the occurrence of breakpoints. Further, it also highlights what those potential features may be which play a crucial role in causing double-strand breaks leading to translocation.

1535T

Assessing the hidden genome architecture of structural variants with Globus Genomics Galaxy pipelines. *A. Rodriguez[1], S.G. Potkin[3], R. Madduri[1,2], F. Macciardi[3,5,6,7], S. Gaudi[8], G. Guffanti[4].* 1) Computation Institute, University of Chicago, Chicago, IL; 2) MCS, Argonne National Laboratory, Argonne, IL; 3) Department of Psychiatry & Human Behavior, University of California, Irvine, School of Medicine, Irvine, CA; 4) Division of Child and Adolescent Psychiatry, Department of Psychiatry, Columbia University/NYSPI, New York, NY; 5) Center for Autism Research and Treatment (CART), University of California, Irvine, California; 6) Center for Epigenetics and Metabolism, University of California, Irvine, California; 7) Department of Pharmacological and Biomolecular Sciences, University of Milan, Milan, Italy; 8) Department of Infectious, Parasitic and Immune-Mediated Diseases Istituto Superiore di Sanità Rome, Italy.

Next-generation sequencing (NGS) opened a new phase of discovery of structural variants. The list of programs and algorithms to assess insertions, deletions, duplications, inversions and mobile elements is in continuous expansion. To get a comprehensive picture of the structural variants genomic architecture, researchers need to use different programs with advanced programming-based query interface, which is often demanding due to the large number of required tools and considerable computational infrastructure and needs. We used cloud-based Globus Genomics platform for development of comprehensive pipeline for assessing the "hidden" genomic architecture of structural variants that includes all steps from processing and quality control of data generated by NGS to functional annotation and data visualization. Globus Genomics enabled us to share our analysis pipelines, collaboratively build the pipeline, share the results of the analysis with our collaborators and ultimately make our published results accessible to the research community at large thus promoting transparency and reproducibility. The research platform integrated Globus for data transfer and management, Galaxy for tool and workflow management and Amazon Web Services (AWS) for compute resources. In addition, we leveraged the Swift parallel language into the execution of tools for seamless parallelization of datasets at a chromosomal level. We optimized each tool to use best available Amazon Web Services resource based on memory and computation needs. Optimizations include identification of tools that can be parallelized to minimize execution time and costs which allowed us to reduce the execution time by a factor of 5. At the time of this writing, we implemented in our pipeline the most up to date, publicly available tools: ForestSV used for detection of deletions and duplications; INDELS and mobile elements insertions are detected using Retroseq and Tangram; reference-based mobile elements (LINE, ALUs and HERVs) are detected with RepeatMasker. Finally, we assembled the structural variants identified by each specific algorithm. The structural variants can be further visualized using circus plots for genome-wide scale summary graphs. In conclusion, we developed a comprehensive pipeline for structural variants analysis featuring a live online supplement (interface with UCSC) providing access to exact analyses and workflows.

1536S

Integrative Analysis of Cancer Genetic Data for Drug Discovery. *S. Saisanit, F. Milletti, Z. Albertyn, J. Cai.* Pharma Research and Early Development Informatics, Roche Innovation Center New York, New York,, NY.

Our objective is to perform integrative analysis of genetic variations such as mutations, gene fusions, copy number, and transcript abundance in tumor samples from patients and cultured tumor cell lines. We envision that these analyses will help identify cancer cell lines best represent a given primary tumor type. We will also discover candidates for biomarkers and therapeutic targets. We process saw RNA sequencing data for identification of gene fusion events. We obtain other processed genetic data from TCGA Firehose and from ICGC Data Portal. We devise mapping strategies specific to each sequencing data type; differential mutation and differential gene and transcript expression (from RNA sequencing data); identification of genetic variations shared by patient samples and cell lines representing the corresponding primary tumor; compound screening panels on cancer cell lines for selected tumor types. Where applicable, we utilize in silico analysis pipelines to handle mapping and variant calling (RNASeq and WXS), gene fusion detection, splice junction detection, transcript-level expression, and isoform usage (RNAseq). Our effort leads to several candidate biomarkers and targets. Experimental validations will be needed to confirm the utility of these candidates. We can also use our integrated data for selecting cell lines to use in compound screenings.

1537M

Identification of differentially methylated genes potentially associated with neurological diseases. *W. Souza, B. Carvalho, D. Dogini, I. Lopes-Cendes.* Department of Medical Genetics, UNICAMP, Campinas, Brazil.

Epigenetic marks, including DNA methylation and histone modifications could play a role in neurological disorders with complex inheritance, such as epilepsy and stroke. However, this issue has not been sufficiently investigated. In this context, our aim was to perform an epigenome-wide association study (EWAS) using a public dataset (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37579), to assess methylation profile across distinct tissues of patients with neurological disorders and compared this to that obtained from a control group. We used principal component analysis (PCA) for data quality control and empirical Bayes methods to adjust for systematic differences allowing better comparisons between tissues and disease-related conditions. Our results show a latent variable that stratified control samples into two subgroups and could represent a potential source of bias for downstream analyses. After further investigation we observed that the two clusters were tightly related to the different ethnicities of the individuals in the control group. After correcting for this variable we applied recent methods for methylation profiling using bumphunter Bioconductor package, which allows one to search for differences in methylation at the regional level rather than at the single-CpG level. We identified differentially methylated regions (DMR) when comparing DNA methylation profiles from different tissue groups. Finally, we compared the list of differentially methylated regions to a curated list of genes known to be associated with neurological disorders. We found that regions of a set of genes related with epilepsy and stroke (including *RELN*, *ARHGEF9* and *BDNF*) were differentially methylated. In conclusion, we identified and corrected a source of bias in the dataset, performed an analysis for differential methylation and identified DMRs that included genes potentially associated with epilepsy.

1538T

Genome-Wide Association Study Of Cerebrospinal Fluid Prostatic Acid Phosphatase Levels. *L.A. Staley[1], P.G. Ridge[1], M.H. Bailey[1], C. Cruchaga[2], A. Goate[3], J.S.K. Kauwe[1].* 1) Biology, Brigham Young University, Provo, UT. USA; 2) Biology and Biomedical Sciences, Washington University in St. Louis, St. Luis, MO. USA; 3) Psychiatry, Washington University School of Medicine in St. Louis, St. Luis, MO. USA.

Prostatic Acid Phosphatase(PAP) is an enzyme that is produced in the prostate in males, and it's physiological function is thought to deal with the liquefaction process of semen. Understanding the genetics of this enzyme is very important because this enzyme may be found in highly increased levels in men who have prostate cancer and moderately increased levels for bone diseases, such as Paget's disease or hyperparathyroidism and blood cell diseases, such as sickle disease, or multiple myeloma or lysosomal storage diseases, such as Gaucher's disease. For the initial association analysis in each series we used PLINK to perform linear regression and evaluated the association between the additive model for 5.8M SNPs and each phenotype. Age, gender, and the principle components from Eigensoft were included as covariates. Meta analysis was performed using default settings in METAL. Genomic inflation factor scores (GIF) were estimated using the R package GenABEL. The initial threshold for significance was P<5×10-8. SNPs that met this threshold were further filtered using the following criteria. First, we rejected markers where the direction of the effect was different in the Knight ADRC and ADNI datasets. Second, we removed all markers where the minor allele frequency was less than 5% (unless they were directly genotyped or had a clear functional annotation). Finally, we also rejected all associations with phenotypes where the genomic inflation factor was greater than 1.03 (GIF was calculated without SNPs where MAF is <0.05). We found 233 genome-wide significant markers(P<5×10^-8). We will analyze each of these markers individually and study the biological functions of each, to discover which is casual of the increased levels of this enzyme for men who have cancer or other diseases.

1539S

European Psoriasis Differences are Defined by Variation in the Epidermal Differentiation Complex. *C.E. Tanes, L. Jackson, A. Tozeren.* Biomedical Sciences, Engineering and Health System, Drexel University, Philadelphia, PA., USA.

Psoriasis is an autoimmune skin condition affecting 2-3% of Americans. Interestingly, European Americans are almost twice as likely to have psoriatic plaques as African Americans (3.6% and 1.9% respectively). Previous studies have shown that psoriasis has a strong genetic signature in human populations, leading us to examine the role that genetic variation might be playing in these prevalence differences. We harness the numerous curated genes identified through single gene and genome wide association studies to broaden the investigation of candidate polymorphisms in ethnic populations. Psoriasis genes curated through the National Center for Biotechnology Information (NCBI) Gene were mapped onto the human genome and regions with an overabundance of psoriasis genes were identified. These genomic hotspots were projected onto gene ontology categories and cellular pathways to draw a bioinformatics portrayal of psoriasis. HapMap (11 populations) and Human Genome Diversity Panel (51 populations) polymorphism data were analyzed for the psoriasis hotspot regions. Our analyses identify six psoriasis hotspots located on chromosomes 1, 2, 5, 6, 17 and 20. Functional annotation of these hotspots showed significant ($p<0.05$) biological processes consistent with keratinization, B cell proliferation and antigen presentation, core disruptions consistent with previous psoriasis studies. Population based polymorphic analyses identified a European specific pattern in genes making up the epidermal differentiation complex (1q21), which bear further investigation as the potential cause of population differences in psoriasis prevalence.

1540M

Exploiting whole exome-seq data for variant discovery from highly divergent regions in the human genome. *S.L. Tian, H.H. Yan, S. Slager.* Mayo Clinic, Rochester, MN.

Whole Exome Sequencing (WES) has been widely used to characterize genetic variations within human genes. However, mutation rate varies substantially across the human genome, which is a key determinant for variant calling. For example, human leukocyte antigen (HLA) region (6p21.3) contains the most polymorphic genes in the human genome and is associated with over 100 diseases. Accurate identification of variants from such highly divergent regions (HDRs) often relies on Sanger sequencing. While frameworks have been established for genome-wide variant calling for WES data, their feasibilities in HDRs have yet to be assessed. Using simulated and real data generated from our CLL (chronic lymphocytic leukemia) project, we have conducted a comprehensive assessment of numerous pipelines for variant discovery. We propose a strategy that achieves both high sensitivity and specificity over a wide range of divergence. Five short-read aligners were selected, including bwa, novoalign, gsnap, NextGenMap and stampy. From simulated data we found that four of the mappers were highly comparable in mapping accuracy (>99%) for reads with <= 1% divergence; novoalign was 3-4% lower instead, a result largely caused by its extensive use of 'soft clipping' to mask unaligned portion in a read. At an increased divergence of 5%, gsnap, NextGenMap and stampy still achieved ~99% mapping accuracy, versus up to 20% unmapped rate by bwa. At 10% and 15% divergence in particular, stampy represents the most effective mapper, followed by NextGenMap and gsnap. We next investigate the overall performance of five selected variant callers: GATK UnifiedGenotyper, GATK HaplotypeCaller, freebayes, platyus and SAMtools/mpileup. GATK UnifiedGenotyper showed the highest sensitivity for SNP calling across nearly all the divergent levels, from simulated reads mapped by gsnap (at >=1% divergence rates) or by bwa and stampy (at <1% divergence rates). More importantly, we observed similar trends in performance for these mappers and callers when applied to the CLL data, which were supported by comparing discovered variants to the calls made from our in-house Illumina iSelect genotype array data and to the known variants in dbSNP. Our findings highlight several key factors and combinations recommended to use in order to have a successful WES-based variant discovery in HDRs.

1541T

An in silico Post-GWAS Analysis of C-Reactive Protein Loci: a Pipeline of Sequential Bioinformatics-Based Approaches. *A. Vaez[1], R. Jansen[2,3], B. Prins[1], J. Hottenga[3,4,5], E. de Geus[4,5], D. Boomsma[3,4,5], B. Penninx[2,3,5], I. Nolte[1], H. Snieder[1], B. Alizadeh[1].* 1) Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; 2) Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands; 3) Neuroscience Campus Amsterdam, VU University and VU University Medical Center, Amsterdam, The Netherlands; 4) Department of Biological Psychology, VU University, Amsterdam, The Netherlands; 5) EMGO+ institute, VU University and VU University Medical Center, Amsterdam, The Netherlands.

Genome-wide association studies (GWASs) have successfully identified a number of Single Nucleotide Polymorphisms (SNPs) associated with serum levels of C-reactive protein (CRP). An important limitation of GWASs is that the identified variants merely flag the nearby genomic region and do not necessarily provide a direct link to the biological mechanisms or pathways underlying their corresponding phenotype. To this end we assembled an integrated pipeline of sequential bioinformatics-based approaches for post-GWAS analysis of human traits or diseases. In the first phase of 'in silico' sequencing it explores the nearby genomic region to identify all linked variants, with a focus on non-synonymous SNPs as potentially functional variants. In the second phase of eQTL analysis, it attempts to identify all nearby genes whose expression levels are associated with the corresponding GWAS SNPs. These two phases generate a number of relevant genes that serve as input to the next phase of functional network analysis. We applied this pipeline to the 18 SNPs that had previously been associated with CRP at a genome-wide significant level[1]. Our in silico sequencing analysis using 1000 Genomes Project data identified 3,801 linked variants, including 25 non-synonymous SNPs. Our eQTL analysis, based on one of the largest single datasets of genome-wide expression probes (n>5000) assessed in participants from the Netherlands twin Register (NTR) and the Netherlands study of Depression and Anxiety (NESDA)[2] identified 23 significantly associated expression probes belonging to 15 genes (FDR<0.01). The final phase of functional network analysis, which was based on 40 relevant genes identified in the previous two phases, revealed 93 significantly enriched biological processes (FDR<0.01). Our post-GWAS analysis of CRP GWAS SNPs confirmed the previously known overlap between CRP and lipids biology. Additionally, it suggested an important role for interferons in the metabolism of CRP.

References

1.Dehghan, A. et al. Meta-Analysis of Genome-Wide Association Studies in >80 000 Subjects Identifies Multiple Loci for C-Reactive Protein Levels. Circulation 123, 731-738 (2011).

2.Jansen, R. et al. Sex differences in the human peripheral blood transcriptome. BMC Genomics 15, 33 (2014).

1542S

The Variant Characterization of 211 Whole Genome Sequences: The Cache County Study on Memory Health and Aging. *M.E. Wadsworth[1], B.D. Pickett[1], J.B. Miller[1], J.D. Duce[1], R.G. Munger[2], C.D. Corcoran[2], J.T. Tschanz[2], M.C. Norton[2], J.S.K. Kauwe[1], K.V. Voelkerding[3,4], P.G. Ridge[1].* 1) Department of Biology, Brigham Young University, Provo, UT., USA; 2) The Cache County Study on Memory and Aging, Utah State University, Logan, UT., USA; 3) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT., USA; 4) Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT., USA.

The 1000 Genomes Project is the current standard for population based variant profile analysis and accurately accounts for 95% of common variants in the human genome. Consequently, the 1000 Genomes dataset has been used to develop novel bioinformatics methods, design genetics studies, and evaluate the accuracy of new variant datasets. While the 1000 Genomes Project is an unparalleled resource for the study of human genetics, rare variants (<5%) were poorly characterized in the 1000 Genomes because of the low coverage of the majority of genomes in the dataset. This leaves an important gap in our understanding of population-level genetic variation since many important variants are rare. In this research we used 211 whole genomes sequenced to an average depth of 37 from The Cache County Study on Memory and Aging (Cache). Each of the genomes is from elderly (>65 years old), healthy individuals of European descent. Many of the CEU genomes in the International HapMap study were derived from individuals in the Cache study and are comparable to the European genomes from the 1000 Genomes Project. Due to the high coverage of these genomes it is possible to accurately characterize rare variants. We created a variant profile for each of the 211 individuals and the Cache dataset as a whole. In total, we observed ~16 million unique single nucleotide variants (SNV) in the dataset. Each individual had ~3.9 million SNVs and ~38,000 InDels. We further classified variants by type, and specifically analyzed variants located in the 57 genes the America College of Medical Genetics and Genomics defined as clinically relevant. Our analyses revealed that each individual had an average of 35 nonsynonymous SNVs, 13,180 SNVs, and 8,136 InDels located in 57 genes (exons and introns). Our research provides important insights into the type of genomic variation that can be expected in a healthy population of Caucasian individuals.

1543M

Genome-wide haplotype-based association study in Chinese Han population identified novel susceptibility locus for systemic lupus erythematosus. *Y. Wang[1], Y. Zhang[1], J. Yang[1,2], Y. Lau[1], W. Yang[1,2].* 1) Department of Paediatrics and Adolescent Medicine, LKS Faculty of Medicine and Queen Mary Hospital, The University of Hong Kong, Hong Kong, China; 2) Centre for Genomic Sciences, LKS Faculty of Medicine and Queen Mary Hospital, The University of Hong Kong, Hong Kong, China.

Systemic lupus erythematosus (SLE) is a heterogeneous and chronic autoimmune disease, resulting in recurrent inflammatory response and multiple-organ damage. Family and twin studies suggested that genetic factors are a crucial component contributing to the disease. The first wave of Genome-wide association studies have identified more than 40 susceptibility loci robustly associated with SLE. However, all of them together only explained a small fraction of disease heritability. In this study, we carried out a genome-wide haplotype-based association study, using the previously published GWA data from two Chinese Han cohorts, Hong Kong (Southern China) and Anhui (Northern China), respectively. Our analyses identified a novel region (meta haplotype-based p-value 6E-07) in chromosome 2q22.1 (upstream of CXCR4). The haplotype-based P value is smaller than any single-SNP P values within the region, suggesting haplotypes are more informative than single SNPs in this case. This strategy was demonstrated to be helpful to partially explain the missing inheritability of the complex diseases including SLE. Here we have provided a detailed account of the methodology (haplotype-based association analysis) and the association signals from this novel locus.

1544T

Next-Generation Sequencer Analysis: The Accurate Somatic SNV Detecting Workflow. *M. Yamaguchi.* Amelieff Corporation, Tokyo, Japan.

Next-generation sequencing technique has dramatically improved the efficiency and the speed of large scale sample analysis. As a result, picking out variations that have a significant impact on the disease of interest from a massive number of detected variations has become a difficult challenge. To overcome this problem, we established an analysis workflow to detect realistically verifiable number of somatic SNVs. In this workflow somatic SNVs are detected by SomaticSniper, the existing software designed to detect somatic SNVs. After detection, known variations, variations of low quality, and variations with low impacts are filtered out. We analyzed publically available NGS data of tumor-normal pair samples to verify the accuracy of our workflow.

1545S

Identification of recurrent drive gene fusions in melanoma using RNA-Seq data. *T. Zhang[1], M. Xu[1], P. Johansson[2], A. Pritchard[2], N.K. Hayward[2], K.M. Brown[1].* 1) National Cancer Institute, Bethesda, MD; 2) QIMR Berghofer Medical Research Institute, Brisbane, Australia.

Molecular targets for effective therapeutic intervention in melanoma have recently been identified within the RAS-RAF-MEK-ERK and, to a less extent, PI3K-AKT pathways. Despite the promise of approaches targeting these pathways, therapeutic resistance remains a significant problem, and the need to identify new molecular targets remains a critical need. Gene fusion events resulting from inversions, interstitial deletions, or translocations occur commonly in cancers, leading to the generation of novel fusion proteins with unique oncogenic properties. In order to identify fusions driving melanoma development and/or progression, we performed paired-end RNA-Seq of 72 melanoma cell lines using the Illumina HiSeq 2000. After filtering out low quality data, TopHat was used to align RNA-Seq reads to the human genome (hg19), and TopHat-Fusion was subsequently applied to identify potential fusions. In order to assess the potential of these fusions to act as driver events, we used a naïve Bayesian classifier Oncofuse to highlight fusions with common characteristics of known 'driver' events. In total, we observed 295 potential fusion events, 112 of which involved neighbor genes located within 1 Mbp on the same chromosome and 101 involving more distant same-chromosome partners. Only 82 of potential fusions were inter-chromosomal. In summary, we detected gene fusions involving PRKAR1A, RBMS3, CBX3, PTPRO, RERG, PTEN and BRAF recurrently occurring in at least 3 samples with the same partner genes. Notably, we identified multiple fusions involving known melanoma oncogenes or tumor suppressors, including BRAF, PTEN, GRIN2A and NF1. The Oncofuse pipeline identified the significant potential driver fusion genes such as BRAF, ATF2, PRKAR1A, MECP2 and IGF1R (FDR<0.01). In addition, we also identified and summarized fusion events in the TCGA Skin Cutaneous Melanoma dataset and found the recurrent/drive fusion genes in melanoma cell lines were also identified in TCGA samples. Particularly, PTEN, MAPKAPK5, RERG, PTPRO and BRAF fusions in melanoma cell lines were also observed in multiply TCGA samples with the same break point. There were a total of 9 BRAF fusions in our cell lines and TCGA samples, and 7 of them were BRAF wild type with the protein kinase domain retained. We are presently functionally characterizing several of these gene fusions, including BRAF, PTEN, ATF2, and PRKAR1A to establish their role in driving melanoma and as potential novel therapeutic targets.

1546M

Application of gene expression deconvolution to the translation of gene expression signatures from pre-clinical models to the clinic. *C. Campbell, A. Reddy, H. Huet, K. Venkatesan, W. Winckler, M. Morrissey, H. Bitter.* Novartis Institutes for BioMedical Research, Cambridge, MA.

Global gene expression profiling by microarray or RNA-seq has been used successfully to discover prognostic and predictive gene signature biomarkers in a number of cancer types and therapies. Translation from pre-clinical cell line models can be difficult because in primary tumors, the average expression is measured in a mixed population of cells, including not only the tumor cells themselves but non-cancerous cells from the surrounding stromal tissue, infiltrating immune cells, and others. Methods using cell-type frequencies in mixed cell populations have been developed to deconvolute cell-type specific gene expression to obtain average cell type expression levels across samples. We have performed *in silico* mixture experiments between cell line expression data from different lineages (pancreatic and myeloid (N=32)). Using these simulated data, we have developed a method to obtain deconvoluted expression values on the individual sample level that show high concordance to the true expression values prior to *in silico* mixing. We are using simulations to analyze the effect of non-uniformity and imprecision in the tumor purity estimates that would be inherent to primary tumor datasets. In addition, we have applied deconvolution to a gene expression signature that was predictive of sensitivity to death receptor 5 activation in cell lines and mouse xenograft models of pancreatic cancer. We examined TCGA pancreatic adenocarcinoma samples (N=39) for which RNA-seq as well as measurements of tumor purity were available. The sensitivity scores in the TCGA primary tumors (mean=0.21) were lower than in pancreatic cancer cell lines (mean=0.36) (p=0.003). When we deconvoluted these expression data (N=18,030 genes), most genes showed a similar level of expression in the tumor and non-tumor cells (96 % within an order of magnitude). We found that the average score in tumor cells (0.23) was higher than in the non-tumor cells (0.15). In addition, tumors with low scores tended to show high expression of genes associated with pancreatic acinar cells, which comprise the bulk of the pancreas, suggesting that contamination of these tumors with normal cells was leading to lower scores. The results point to the utility of this approach for accurately measuring gene expression signatures in the specific cell-type of interest. Such an approach may boost the clinical utility of signatures developed in cancer cell lines or xenograft models.

1547T

Whole Exome Sequencing in Two Siblings with Developmental Regression and Hypermetabolism. *S. Jougheh Doust[1], J. Clark[1], B. Robinson[2], J. Cameron[1], R. Cohn[1].* 1) Clinical and Metabolic Genetics, The Hospital For Sick Children, Toronto, Canada; 2) Mitochondrial Laboratory, DPLM, The Hospital for Sick Children, University of Toronto.

Hypermetabolism is physiological state of increased rate of metabolic activity. The patients have an abnormally high intake of calories followed by continuous weight loss. It is a common feature of metabolic and mitochondrial disorders. We report two sisters with developmental regression, hyperactivity and hypermetabolism confirmed with indirect calorimetry. Other symptoms include: seizures, aggressive behavior and oral intake of about 4000 kilo calories per day. The growth curves were at 25th and 50th percentile for their weights. The patients had coarse facial features with synophrys. The parents are consanguineous (half first-cousins) and Caucasian. They have one healthy son. The rest of the family history is non-contributory. Extensive investigations including brain imaging, electroretinography, genetic and metabolic work up were inconclusive. Mitochondrial assays showed an increase in mitochondrial lactate production, normal pyruvate production and a normal lactate to pyruvate ratio. Pyruvate dehydrogenase(PDH) enzyme activity was decreased. Sequencing of *PDH1* and *PDK3* (pyruvate dehydrogenase kinase 3) genes were negative. Whole exome sequencing performed on genomic DNA of one of the siblings at TCAG. In searching for candidate gene(s) responsible for these sisters' phenotype we looked for homozygous mutations in either mitochondrial genes or the genes that are expressed in the brain. No mitochondrial candidate genes identified. However, a homozygous frameshift mutation in *MAPK8IP2*(mitogen-activated protein kinase 8 interacting protein 2) gene, aka, *IB2* (Islet Brain 2), is reported in our patient. *IB2* is located on long arm of chromosome 22 and is one of the deleted genes in 22q13 microdeletion syndrome. The function of *IB2* is unknown in human. Recent studies showed that IB2 protein is expressed in brain and disruption of *IB2* gene can cause autism-like phenotype in knock-out mice (Gisa et al 2010). Sanger sequencing did not confirm the presence of the mutation in both siblings. Therefore we excluded *IB2* gene mutation as a cause for our patients' presentation. The next step is to look for other candidate gene(s) in our patient's exome library and explore more about the new candidate gene(s).

1548S

Genome-wide association study of serum metabolites using non-targeted metabolomics to identify new metabolic loci. *L. Lind[1], S. Gustafsson[2], S. Salihovic[1], A. Ganna[3], T. Fall[2], A.C. Syvänen[2], T. Axelsson[2], C.M. Lindgren[4], A.P. Morris[5], C.D. Broeckling[6], J. Prenni[6], E. Ingelsson[2].* 1) Dept of Medical Sciences, Uppsala University, Uppsala, Uppsala, Sweden; 2) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 3) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 4) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 5) Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 6) Proteomics and Metabolomics Facility, Colorado State University, Fort Collins, Colorado, USA.

Background: One way to understand human metabolism in more detail is to identify genes involved in metabolic pathways. The most recent efforts have linked around 150 genetic regions to different metabolites via genome-wide association studies (GWAS) of the metabolome. The present study aimed to extend this knowledge using three large independent cohorts in which multiple metabolic features were measured using non-targeted metabolomic profiling. Methods: In the ULSAM (n=1,028), PIVUS (n=970) and TwinGene studies (n=1,670), up to 10,000 metabolic features were measured in serum/plasma using the same technical platform across studies (UPLC and MS/MS). Genotyping was performed using the Illumina Human OmniExpress or Omni2.5M array and subsequently imputed using 1000 Genomes CEU data. A GWAS was performed for each metabolic feature using linear regression adjusting for age, sex and population-based principal components. We used TwinGene as the discovery dataset and attempted replication of genome- and metabolome-wide significant findings in the ULSAM and PIVUS cohorts. Results: We could confirm 42 of the previously reported associations between genetic regions and metabolites. Top findings included associations with the PYROXD2 locus (p-value=4.0×10-323) and the FADS gene cluster (P=4.6×10-111). In addition, we identified 8 novel associations between genetic variants and circulating metabolites (p-value range 1.1×10-11-9.6×10-93). Likely genes in those 8 regions were PNLIPRP1, LOXL4, FDX1, BCMO1, CYP2D6, UGT2B17, SLC22A1, and CYP3A43. Conclusion: In a GWAS of non-targeted metabolomics data, we could identify eight new genetic loci being related to metabolic traits, and replicate a large number of previously reported loci. Our findings will contribute towards an improved understanding of human metabolism.

1549M

A β-cell specific protein subnetwork significantly enriched for association with GLP-1 stimulated insulin secretion: A DIRECT study. *V. Gudmundsdottir[1], H.K. Pedersen[1], L.M. 't Hart[2], K. Banasik[3,4], D. Boomsma[5], E. de Geus[5], M. Eekhoff[6], M. Diamant[6], M. McCarthy[3,7], E. Pearson[8], C. Workman[1], R. Gupta[1], S. Brunak[1].* 1) Technical University of Denmark, Lyngby, Denmark; 2) Leiden University Medical Center, Leiden, The Netherlands; 3) University of Oxford, Oxford, UK; 4) University of Copenhagen, Copenhagen, Denmark; 5) VU University Amsterdam, The Netherlands; 6) VU University Medical Center, Amsterdam, The Netherlands; 7) Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford, UK; 8) University of Dundee, Dundee, UK.

The heritability of insulin secretion response (ISR) during a modified hyperglycemic clamp with glucose, glucagon-like peptide 1 (GLP-1) and arginine stimulation has been shown to be considerable (h2=0.52 to 0.80) and it is of great interest to identify genetic variants influencing these responses. Evaluating ISR with a modified hyperglycemic clamp is not feasible in larger cohorts of sizes preferred for genome-wide association studies (GWAS). When sample sizes are small, GWAS may be complemented with systems biology approaches to aid the prioritization of genetic variants. Within the DIRECT consortium a GWAS was performed on GLP-1 stimulated ISR and data integration used to add biological context to the results and facilitate variant prioritization. GLP-1 stimulated ISR was measured with a modified hyperglycemic clamp in 130 twins and sibs from the Netherlands twin register. The cohort was genotyped using the Illumina HumanCore Exome BeadChip and association analysis was performed using the QTassoc software and adjusted for age, sex, familial relationships and insulin sensitivity index. Gene-based P-values were mapped onto a β-cell specific protein-protein interaction (PPI) network, which was created by pruning high confidence PPIs from InWeb 3.0 using published β-cell RNAseq data. Connected components in the network enriched for high scoring genes were identified with the Cytoscape plugin jActiveModules and their significance evaluated by comparing to 10,000 degree-preserved randomly sampled subnetworks from the β-cell PPI. None of the variants tested in the GWAS reached a genome-wide significance of $P \leq 5.0E-8$. However, the top scoring subnetwork (25 genes) had a significantly higher combined z-score than expected by random ($P \leq 1.0E-4$). It was most strongly enriched for the Gene Ontology terms “cell junction” (P = 3.6E-4), “plasma membrane part” (P = 2.2E-4) and “cell projection” (P = 4.9E-3) and contained a number of genes known to affect β-cell mass and function (FOXO1), insulin secretion (WFS1, RYR2) or be implicated in type 2 diabetes (MAGI2, CTNNA2 and PTPRD). We have identified a β-cell PPI network enriched for genes with nominal associations with GLP-1 stimulated ISR, demonstrating how data integration can highlight biological mechanisms underlying a phenotype where GWAS results on their own may be insufficient. Furthermore, the network can be used to prioritize genetic variants to take forward for replication in independent cohorts.

1550T

Analysis of Whole Exome Datasets to Test the Hypothesis of Digenic Inheritance in Stargardt Disease. *K. Lee[1], D.S. Marchuk[1], M.J. Friez[2], C. Bizon[3], D. Young[3], D. Gillis[3], P. Owen[3], K. Wilhelmsen[1], K.E. Weck[1,4], S. Garg[5], J.P. Evans[1], J.S. Berg[1].* 1) Dept Genetics, Univ North Carolina, Chapel Hill, NC; 2) Greenwood Genetic Center, Greenwood, SC; 3) Renaissance Computing Institute, Chapel Hill, NC; 4) Dept Pathology and Laboratory Medicine, Univ North Carolina, Chapel Hill, NC; 5) Dept Ophthalmology, Univ North Carolina, Chapel Hill, NC.

Whole exome sequencing (WES) is a valuable tool in determining the molecular etiology for retinal disorders given their genetic heterogeneity. We performed WES in 45 patients with various retinal disorders. Two patients with simplex Stargardt disease (SD; [MIM 248200]) harbored one pathogenic variant in ABCA4 and one pathogenic variant in a second retinal disease gene (CNGB3 or GUCA1A) suggesting the possibility of digenic inheritance. Digenic inheritance has been reported in other retinal disorders, but not in SD. Approximately one-third of simplex cases have only one identifiable mutation in ABCA4; digenic inheritance may explain a subset of these cases and would be supported by finding an excess of mutations in other retinal disease genes in such patients. We evaluated the frequency of double carrier status for ABCA4 and another retinal disease gene in the general population by searching for carriers of ABCA4 pathogenic variants in WES datasets not enriched for retinal disorders (1,189 individuals). ABCA4 variants previously reported as pathogenic, or rare (<1% minor allele frequency) and possibly damaging (missense, nonsense, frameshifting indels, canonical splice site), were selected for further analysis. Pathogenicity was assessed by allele frequency, conservation data, in silico modeling and cosegregation data. For the 40 (3%) ABCA4 carriers identified, we analyzed 448 variants within 213 genes associated with retinal disorders. No individual had two pathogenic ABCA4 variants. One of 8 variants of possible pathogenicity was identified in 9 ABCA4 carriers within PDE6A, EYS, ALMS1, USH2A, BBS9, BBS10, ABCC6 or NR2E3; two individuals harbored the same variant previously reported as a pathogenic mutation in the NR2E3 gene. However, no carrier had a known pathogenic variant in other genes associated with SD, including PROM1, PRPH2, ELOVL4 and CNGB3, or GUCA1A. The finding of 2 heterozygous variants associated with SD among 45 patients with retinal disorders is significantly enriched compared to population controls (Fisher’s exact test, P=0.0013). Our preliminary analysis suggests that double carrier status for ABCA4 and another SD disease gene is rare in the general population, supporting the possibility that double heterozygous mutations in two SD genes could be causative for SD. WES datasets are an effective method to evaluate heterogeneous disorders for the presence of double carrier status as a means to test the hypothesis of digenic inheritance.

1551S

Efficiency of exome sequencing for the molecular diagnosis of Pseudoxanthoma Elasticum. *M.j. Hosen[1], F. Van Nieuwerburgh[2], D. Deforce[2], L. Martin[3], G. Lefteriotis[4], A. De Paepe[1], P.J. Coucke[1], O.M. Vanakker[1].* 1) Centre for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 2) Department of Pharmaceutics, Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium; 3) Department of Dermatology, Angers University Hospital, Angers, France; 4) Department of Vascular Physiology and Sports Medicine, Angers University, Angers, France.

The molecular etiology of pseudoxanthoma elasticum (PXE), an autosomal recessive connective tissue disorder, has become increasingly complex as not only mutations in ABCC6 but also ENPP1 and GGCX can cause resembling phenotypes. Identification of modifier genes, such as VEGFA, has further contributed to the molecular heterogeneity of PXE. In such heterogeneous diseases, Next Generation Sequencing allows to perform mutation screening of several genes in a single reaction. We explored whole exome sequencing (WES) as an efficient diagnostic tool to identify the causal mutations in ABCC6, GGCX, ENPP1 and VKORC1, in 16 PXE patients. WES identified the causal mutation in 27 of 32 alleles (ABCC6 or GGCX, with no causal mutations in ENPP1 or VKORC1). Exomes with insufficient reads (≤20 depth) and patients with no or single mutations were further evaluated by Sanger sequencing (SS) and MLPA, but no additional mutations were found. The potential of WES is to explore multiple genes at a time, the ease to update target genes and the opportunity to search for novel genes when targeted analysis is negative. Together with low cost, rapid and less laborious workflow, we conclude that WES complemented with SS can provide a tiered approach to molecular diagnostics of PXE.

1552M

Unraveling Genetic Architectures Spanning Mendelian and Complex Phenotypes with Data Driven Electronic Medical Record Validation. *B.S. Glicksberg[1,2], L. Li[1], R.Z. Castellanos[1], J. Hakenberg[1], W. Cheng[1], S. Khader[1], M. Ma[1], L. Shi[1], H. Shah[1], J.T. Dudley[1,3], R. Chen[1].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Health Evidence and Policy, Icahn School of Medicine at Mount Sinai, New York, NY.

While Mendelian and complex phenotypes are traditionally organized into distinct categories, emerging studies suggest there may be pervasive shared genetic connections within and across these categories. To further explore this intriguing relationship, we have compiled and analyzed a comprehensive, integrated knowledgebase comprised of known published mutation-gene-phenotype relationship representing over 2,500 Mendelian phenotypes with around 4,000 unique variant-associated genes collected from the Online Mendelian Inheritance in Man (OMIM) and Human Gene Mutation Database (HGMD) online sources. Additionally, we compiled genetic disease associations for >700 complex phenotypes and >3,500 variant-associated genes from the open Genome-Wide Association Studies (GWAS) Catalog as well as a proprietary database ActiVar which provided additional GWAS findings from manual curation of ~13,000 publications. We performed a statistical analysis based on the Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine similarity method to test for enrichment between Mendelian and complex diseases based on shared genetic associations. This analysis identified 1,116 significantly enriched Mendelian-complex disease pairs at q-value < 0.05 and 609 enriched pairs at q-value < 0.01. We further evaluate these novel Mendelian-complex disease pair enrichments by analyzing disease co-morbidities across Electronic Medical Records (EMR) for 3.5 million unique patients comprising more than 2 billion clinical encounters in the Mount Sinai Electronic Medical Record Warehouse (MSDW). Of the genetically enriched pairs (q<0.01), we found 23 that were also significantly enriched in the MSDW using a Bonferroni corrected Fisher's Exact test (p<0.0001). There were 25 Mendelian phenotypes hubs and 44 complex disease hubs that had 10 or more significantly enriched connections. These results from our analyses have been incorporated into a comprehensive phenotype network for visualization and exploration. Furthermore, borrowing information from these phenomic cross connections, new insights can be extracted regarding inheritance patterns of complex diseases as well as candidate modifier genes in Mendelian phenotypes. Additionally, these connections can inform clinical aspects of these phenotypes utilizing data from the MSDW validation: phenotypes in a validated pair can be used as a diagnostic marker of risk, symptom alteration, and prediction.

1553T

Integration of GWAS signals, measures of polymorphic structure and linkage disequilibrium to discover clinically relevant biomarkers and improve identification of causal variants. *M.W. Lutz[1], R. Saul[2], O. Chiba-Falek[1], D.K. Burns[3], A.M. Saunders[1], A.D. Roses[1,3].* 1) Department of Neurology, Duke University School of Medicine, Durham, NC; 2) Polymorphic DNA Technologies, Alameda, CA; 3) Zinfandel Pharmaceuticals, Inc., Durham, NC.

Variants in non-coding RNA, epigenetic regulators, introns, promoters, and distal regulatory elements are associated with a variety of complex phenotypes including longevity and neurodegenerative diseases. We have also shown that small insertion/deletion polymorphisms within noncoding regions of the gene *TOMM40* are associated with tissue-specific modulation of *APOE* and *TOMM40* gene expression, a mechanism presumably underlying their contribution to genetic susceptibility to late-onset Alzheimer's disease (LOAD). These variants were identified through phylogenetic analysis and consideration of linkage disequilibrium (LD) for a 10 kb genomic region on chr 19 that consistently showed a strong genome-wide association (GWA) signal for LOAD. Here, we present a genome-scale implementation of an approach that integrates GWAS signals with measures evaluating the polymorphic potential of genetic variants (small insertion/deletion polymorphisms, structural variants). The approach utilizes bioinformatics methods, i.e., preferential LD (Zhu Q., 2012) and localization of a minimal region containing a causal variant (Bochdanovits Z., 2013), to identify potential causal regulatory variants. We employed this approach for examining the association of 20 contiguous genes on chr 19 (spanning 1.4 Mb) with longevity and LOAD using the AD Neuroimaging Initiative (ADNI) database and a dataset from a study on longevity (Sebastiani P., 2012). The two methods identified 88-Kb and 300-Kb regions likely to contain causal variants. In addition to the strong GWAS signal for this region, genomic structural features suggest possible inter-relatedness of genes and variants in this region. 209 variants were identified and scored with respect to their likelihood of being highly significant. An unusual strand alignment in this region, unlikely to occur by chance, involves 15 consecutive genes (PVR through CLASP) all located on the + strand. Our findings have a biological significance; co-regulation of APOE expression by variants located in the *TOMM40* gene has been reported and four neighboring genes in the cluster (*APOE*, *APOC1*, *APOC4* and *APOC2*) belong to a gene family and exhibit similar biological function. Application of the general approach has the potential to elucidate the biochemical basis for disease-gene associations, identify genetic biomarkers for clinical application, and enrich the analysis of biochemical pathways.

1554S

The Utah Genome Project is successfully discovering and diagnosing genetic disease using VAAST, pVAAST and Phevor. *M.V. Singleton[1], B. Kennedy[1], Z. Kronenberg[1], M.G. Reese[3], M. Yandell[1,2].* 1) Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112, USA; 2) Utah Science, Technology, and Research Center for Genetic Discovery, University of Utah, Salt Lake City, UT 84112, USA; 3) Omicia Inc., 1625 Clay Street, Oakland, CA 94612, USA.

The Utah Genome Project (UGP) is a large-scale, intramural genome-sequencing project, the aim of which is to discover new disease-causing genes and diagnose inherited diseases. What sets the UGP apart from similar projects is its variant interpretation and diagnostic pipeline that includes innovative bioinformatics tools, including VAAST, pVAAST and Phevor. VAAST and pVAAST are probabilistic disease-gene finders capable of prioritizing every variant and gene in the human genome. Single affected and case control cohorts are analyzed using VAAST while pVAAST adds linkage information for additional power in large pedigrees. Successful molecular diagnosis using an exome or genome sequence hinges on accurate association of damaging variants to the patient's phenotype. Unfortunately, many clinical scenarios (e.g. single affected or small nuclear families) have little power to confidently identify damaging alleles using sequence data alone. Today's variant interpretation tools are simply underpowered for accurate diagnosis in these situations, limiting successful diagnoses. Phevor alleviates this lack of power by dynamically incorporating the patient's phenotype into the disease-gene search. Phevor works by expanding known phenotype associations using knowledge resident in biomedical ontologies, like the Human Phenotype and Gene Ontologies, and combines those results with the VAAST/pVAAST outputs. Phevor accurately reprioritize candidates identified by VAAST/pVAAST in light of knowledge found in the ontologies. Phevor is not limited to a fixed set of genes or phenotypes. This enables Phevor to improve diagnostic accuracy for established diseases and previously un-described or atypical phenotypes. The UGP pipeline, including Phevor, has been successfully applied to several clinical cases. The UGP provides successful diagnosis of genetic disease for those with known and novel disease-alleles, novel disease-genes, atypical phenotype presentation and family specific mutations discovered in large pedigrees. Presented are several clinical cases where the UGP provided a molecular diagnosis where the common technique (filtering and querying disease databases) failed. Free from the limitations imposed by disease databases, the UGP analysis pipeline promises to improve diagnostic rates for inherited disease.

1555M

GWAS analysis of epigenetic age acceleration. *A. Lu[1], R. Ophoff[1,3,4], K. Eijk[4], J. Bell[5], I. Erte[5], P. Tsai[5], T. Spector[5], K. Hao[6], S. Dracheva[7], S. Horvath[1,2].* 1) Human Genetics, UCLA, Los Angeles, CA; 2) Biostatistics, UCLA, Los Angeles, CA; 3) Center for Neurobehavioral Genetics, Department of Psychiatry, UCLA, Los Angeles, CA; 4) Rudolf Magnus Institute of Neuroscience, Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Department of Twin Research and Genetic Epidemiology, King's College London, United Kingdom; 6) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine; 7) Department of Psychiatry, Mount Sinai School of Medicine.

Recent articles describe methods for estimating the age of a tissue sample based on DNA methylation levels. By contrasting estimated age, referred to as DNAm age, with chronological age, one can arrive at a measure of age acceleration. Age acceleration is highly heritable (estimated to be around 40%). Age acceleration is a quantitative measure that can be used in a QTL analysis to identify SNPs. Here we report the findings of a large meta-analysis involving over a dozen data sets comprised of both DNA methylation and SNP data. Imputation was performed for each study using 1000 Genome reference panels newly released in December 2013. This yields ~7 M reliable markers with common variants, minor allele frequency (MAF) >0.05 in at least one study, available for association analysis. A genome-wide QTL analysis was performed in each dataset. We used a meta-analysis to aggregate the results from the different data sets identifying 34 SNPs (1 called and 33 imputed) in 8 genes. The imputed markers passed stringent quality controls (genotype information measure > 0.74 and mean MAF = 0.16). When we restricted the analysis to participants who are younger than 70, we found 5 more SNPs in two additional genes. Overall, our results demonstrate that epigenetic age acceleration is under strong genetic control.

1556T

Differentially Expressed Genes in Asthma Differ by Tissue-type. *T.B. Mersha[1], E. Geh[2], D. Ghosh[2], J.A. Bernstein[2], G.K.K. Hershey[1].* 1) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Department of Internal Medicine, University of Cincinnati, OH.

Several studies have identified differentially expressed genes (DEGs) between healthy versus asthmatic patients using microarray technology. This approach, however, have led to considerable variation in the lists of genes generated by various groups due to variation from random noise to biological samples. There are two objectives of this study a) conduct a comprehensive cell-type specific expression profiling analysis in order to identify the target genes, pathways and networks that are enriched in each cell or tissue-type, b) develop cell and tissue-type specific predictions models. Based on cell/tissue origin, a total of 400 samples and 22,590 DEGs were divided into bronchial epithelia (BEC), nasal epithelia (NEC), bronchial biopsy (BB), nasal biopsy (NB), alveolar macrophages (AM), lymphocytes (CD) and lung fibroblasts (FIB). Multivariate analysis of heatmaps of DEGs from similar cells shows a clear separation between asthmatics and controls. Moreover, DEGs from various tissue-specific datasets share more commonality at the pathway level than at the gene level as reflected by Jaccard similarity index of 0.65 between BEC and NEC. These results also highlight the importance of a cell-type based approach when analyzing multiple gene expression data. Our study identified novel asthma-related genes and pathways as future therapeutic targets and prediction models.

1557S

Smoking-related microRNAs and mRNAs in human peripheral blood mononuclear cells. *M.W. Su[1,3], K.Y. Tung[1], S.L. Yu[2], C.H. Tsai[1], Y.F. Wu[1], Y.J. Lai[3], C.H. Liu[3], Y.L. Lee[1,3].* 1) Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan; 2) Department of Clinical Laboratory Sciences and Medical Biotechnology, National Taiwan University College of Medicine, Taipei 100, Taiwan; 3) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

Teenager smoking is of great importance in public health which may increase the incidence of asthma and allergy. Tobacco smoke would be sensed by monocytes as an injurious stimulus and induced the expression change of genes associated with immune, and inflammatory. However, the regulatory mechanisms and the benefits of smoking reduction in alteration of immunotoxicological system remained poorly understood. In 2011, we conducted the Teenager Smoking Reduction Trial (TSRT), which was a community-based trial aiming to evaluate the effectiveness of smoking reduction in teenagers, to examine the causal relationship between smoking reduction and alterations in immunotoxicological system. A total of 100 teenagers with active smoking habits were recruited and 12 teenagers who reported a substantial reduction in smoke quantity and an decrease in cotinine/creatinine ratio were included in genomic analyses. We examined the whole genome mRNA expression in peripheral blood mononuclear cells (PBMC). A total of 292 genes with a false discovery rate adjusted p-value < 0.01 were deemed to be differentially expressed in TSRT. The Gene Ontology biological process annotations based on Gene Set Enrichment Analysis was conducted on the 292 differentially expressed genes. The smoking reduction induced expression changes mainly involved in the signal transduction. We found *BCL3*, *TNF*, *RHOB*, *BCL6*, and *NLRP3* were important genes in regulation of multiple biological processes after smoking reduction. We increased the gene expression fold changes cut-offs >1.5 to depict the regulatory pathways by Ingenuity Pathway Analysis. The primary regulatory pathways characterized to be differentially regulated in PBMC after smoking reduction were "Glucocorticoid Receptor Signaling", "Differential Regulation of Cytokine Production in Intestinal Epithelial Cells by IL-17A and IL-17F", "IL-6 Signaling", "TNFR2 Signaling", and "Communication between Innate and Adaptive Immune Cells". Integrated analyses of microRNA and mRNA expression further identified eleven differentially expressed microRNAs negatively correlated with their predicted target genes ($p<0.05$). Among the eleven microRNAs, CD83 molecule was regulated by mir-4498, which was critical for the communication between innate and adaptive immune cells and played an important role in regulation of immunotoxicological system of allergy. Real-time PCR confirmation and functional validation of miR-4498 and CD83 molecule were undergoing.

1558M

The mitochondrial mutational landscape of human cancer and its impact on tissue- and tumor-specific gene expression. *S. Grandhi, N. Dhami, C. Bosworth, T. Laframboise.* Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH.

Unlike that of the nuclear genome, the regulatory landscape of the mitochondrial genome remains largely uncharacterized, particularly in the context of cancer. In contrast to normal cells, which rely primarily on mitochondrial oxidative phosphorylation for energy production, cancer cells instead rely heavily on aerobic glycolysis, even in the presence of oxygen. Although there is an accumulation of somatic mitochondrial DNA (mtDNA) mutations in cancer, the impact of these mutations on the role of mitochondria in reprogramming cancer cell bioenergetics is unclear. Even in the context of normal cells, mitochondria are highly sensitive to their cellular environment: heart, muscle, kidney, and liver cells have higher mitochondrial gene expression levels than lung, adrenal, and thyroid cells, likely in response to tissue-specific energy demands.

Here we investigate the role of the mitochondrial genome in tumor cell bioenergetics while addressing tissue specificity. To achieve this, we mine publicly-available next-generation sequencing (NGS) data across 15 cancer types from the Cancer Genome Atlas (TCGA), the Cancer Cell Line Encyclopedia (CCLE), and the Encyclopedia of DNA Elements (ENCODE). Using these datasets, we survey mitochondrial DNA mutational patterns, mRNA expression levels, and mtDNA transcription factor occupancy.

Our analysis of TCGA RNA-sequencing data from approximately 1,800 cancer patients suggests that mitochondrial-encoded gene expression is significantly dysregulated in cancer cells and also clusters distinctly based on the tumor's tissue of origin. Additionally, we observe that patterns of differential gene expression in tumor as compared to matched normal samples vary across tumor types. To interrogate potential mechanisms underlying these gene expression patterns, we integrate them with mtDNA copy number, mtDNA mutations, and mtDNA-binding transcription factors, while addressing the overall somatic mutation frequency in mtDNA across cancer types.

This pan-cancer query of the mitochondrial regulatory landscape distinguishes tumor- and tissue-specific mitochondrial aberrations on the DNA and RNA levels by integrating multiple NGS platforms from disparate sources and analyzing them in tandem.

1559T

Somatic Mutation Detection by Whole Exome Sequencing in Patients with Adult-onset Still's Disease. *Z. Deng, C. Jeon, A. Almeida de Jesus, M. Ombrello, R. Goldbach-Mansky.* Intramural Research Program, NIAMS, National Institute of Health, Bethesda, MD.

Adult-onset Still's disease (AOSD) is a severe, inflammatory disease that is characterized by high, spiking fever, arthritis, a characteristic skin rash and increased acute phase reactants. The molecular mechanism underlying AOSD is unknown, but given its apparent lack of heritability, we hypothesize that somatic mutations or mosaicism may contribute to its pathogenesis. To investigate whether somatic mutations play a role in AOSD, we have sequenced matched tissues (whole blood, fibroblasts and monocytes) from 5 patients. Using fibroblasts as the control tissue to determine germline variation, we test the hypothesis that monocytes or another cell population in whole blood might harbor somatic mutations that contribute to the disease. Whole exome sequencing was performed by a sequencing service provider (Otogenetics) using Agilent SureSelect V5 (51Mb) capture and Illumina HiSeq instruments. The average coverage of each sample is about 60X with greater than 95% of the exome having at least 10X coverage. Four different computational methods have been utilized for detection of somatic mutations: BWA-PICARD-GATK, MuTect, VarScan2 and Diamund. While the first three tools require BAM files generated by aligning sequence reads to the reference genome, Diamund directly compares the reads and thus doesn't require the reference genome for variant detection. The number of potential somatic mutations called by each method ranges from fewer than 100 to over 1,000 per matched pair. Among the methods used, Diamund reports the most candidate somatic mutations, while muTect identifies the least. The results generally don't overlap between the different methods and are comprised of mostly low quality variants, suggesting that most of the somatic calls are likely to be false positives. This observation is consistent with the 0.5% to 5% discordant rates observed between technical replicates and the assumed low mutation rates in the affected tissues. The few mutations called by more than one method are being evaluated and could provide important clues on the molecular etiology of AOSD.

1560S

Weighted gene co-expression network analysis suggests white matter might play a role in epilepsy and episodic motor disorders. *L. Silveira-Moriyama[1, 2,3,4], M. Ryten[4,5], P. Forabosco[6], C. Bettencourt[4], A.J. Lees[1,2], H. Houlden[4], J. Hardy[1,4], T. Warner[1,2,4], UK Brain Expression Consortium.* 1) Reta Lila Weston Institute, UCL Institute of Neurology, London, London, United Kingdom; 2) Queen Square Brain Bank for Neurological Studies, London, UK; 3) Neurology Department, University of Campinas, Campinas, Brazil; 4) Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK; 5) Department of Medical and Molecular Genetics, King's College London, London, UK; 6) Istituto di Ricerca Genetica e Biomedica CNR, Cagliari, Italy.

PURPOSE: There is a significant phenotypic overlap of genes causing epilepsy and episodic motor phenomena (such as episodic dyskinesia, episodic ataxia and episodic hemiplegia). We used genome-wide gene expression (GWGE) data to study co-expression of these genes. METHODS: GWGE data generated from 788 samples of 101 adult control human brains (10 distinct brain regions each) was collected as part of the UK Human Brain Expression Consortium (UKBEC). Weighted gene co-expression network analysis was used to group all expressed genes into modules in an unsupervised manner. Target genes were: "episodic" genes (ATP1A2, SCN1A, SLC1A3, ATP1A3, KCNMA1, CACNB4, CACNA1A, KCNA1, GCH1, SLC2A1, PNKD and PRRT2) and "epilepsy" genes (PRRT2, ARX, CDKL5, CHRNA2, CHRNB2, GABRG2, GABRA1, KCNQ2, KCNQ3, LGI1, PCDH19, SCN1A, SCN1B, SCN2A, STXBP1, NDE1, TBC1D24, GOSR2, AFG3L2, GRN, KCNT1,CHRNA4, PLCB1, and SLC2A1). The overrepresentation of each group in modules was assessed using chi-squared tests with Yates correction. Gene ontology (GO) and KEGG pathway enrichment analysis (with Bonferroni correction) was performed. RESULTS: Episodic genes highlighted a single white matter gene co-expression module (blue module), which was enriched for such transcripts (8/12 genes, 13/17 transcripts, $p<0.0001$) and for epilepsy transcripts (18/23 genes, 20/26 transcripts, $p<0.0001$). This module was enriched for synaptic transmission (p-value = 4.02 x 10-44) and more specifically the glutamate signalling pathway (p-value = 4.89 x 10-6) and others. It was also enriched for a number of KEGG pathways including the long term potentiation (p-value = 7.47 x 10-8) and axon guidance (p-value = 1.17 x 10-7). Of the 60 top genes in this module (module membership 0.9 or higher) 12 are known to cause human disease. Of these, eight cause epilepsy (SCN8A, ARHGEF9, GABRB3, GRIN2B, SCN2A, SYN1, GLRB and GABRG2). CONCLUSION: Our data reinforces the link between epilepsy and episodic disorders, and the notion that PRRT2 and PNKD - relatively unknown genes - relate to synaptic transmission. It suggests that white matter might play a role in these disorders. The main caveat of the finding is that globally genes in the blue module have lower expression in the white matter than in other brain regions. Further studies are warranted to understand these findings at depth including replication of the clustering in an independent data set and in-situ studies of gene expression in the various white matter cell types.

1561M

Novel intergenic large non-coding RNAs (lincRNA) in Human Retina and RPE/Choroid. *L. Tian[1], A. Bowman[2], Y. Liu[1], C. Curcio[3], H. Hakonarson[1], M. Li[4], D. Stambolian[2].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pa. 19104. USA; 2) Department of Ophthalmology, University of Pennsylvania, Philadelphia, Pa. 19104. USA; 3) 4. Department of Ophthalmology, University of Alabama School of Medicine, Birmingham, Al. 35294. USA; 4) 3. Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pa. 19104. USA.

Large intergenic non-coding RNAs (lincRNAs) are emerging as key regulators of diverse cellular processes. Recent advances in RNA sequencing (RNA-Seq) and computational methods allow for a comprehensive analysis of lincRNAs. There have been no comprehensive studies involving the presence and role of lincRNA in the eye tissues and eye related pathologies till date. To build a comprehensive catalogue of novel lincRNAs expressed in eye, we performed RNA-Seq on eight human post-mortem eyes and generated a comprehensive set of data for both region-based (macular versus peripheral retina) and tissue-based (retina versus RPE/choroid) studies. We have processed RNA-seq data and identified ~2000 potential novel lincRNAs per sample which are not present in the catalog of the Human Body Map lincRNAs. Briefly, we first assemble the transcripts using Cufflinks and Scripture. To differentiate true lincRNA transcripts from assembly artifacts, we use the software Sebnif, which implements a series of filtering steps such as filters of transcript length, expression level and coding potential. Each lincRNA has been annotated with genomic features around its promoter and gene body, such as 36 of histone H3 (H3K4me3, and H3K36me3), Expressed Sequence Tag (EST) and Cap Analysis of Gene Expression (CAGE) tags. We also looked for colocalization and coexpression of the novel lincRNAs with known AMD GWAS hits to identify possible links between lincRNA and AMD genes.

1562T

Gene variant modification in keratinocyte cell samples irradiated by UV using RNA-seq. *V. Mijatovic[1], C. Patuzzo[1], G. Buson[2], A. Dal Molin[1], A. Mori[1], E. Caviola[3], M. Delledonne[2], P. Pertile[3], P.F. Pignatti[1], G. Malerba[1].* 1) Dep. of Life and Reproductions Sciences, University of Verona, Verona, Verona, Italy; 2) Department of Biotechnology, University of Verona, Verona, Italy; 3) Cutech Srl, Padova, Italy.

Currently the levels of gene expression can be estimated at the genomic level using the high throughput sequencing platforms. Therefore it is theoretically possible to identify genetic variants present in most expressed genes. Over-exposure to ultraviolet (UV) radiation is important environmental risk factor related to skin cancer development. We studied keratinocyte cell samples irradiated by 2 different dosages of UV rays (250 e 500 mJ/cm2) to investigate if UV irradiation might modify the sequence of some gene transcripts, reflecting changes on DNA sequence, and if such modifications might be detected in RNA-seq experiments. We started focusing our attention on heterozygous single-nucleotide polymorphisms (SNPs) and alternative-allele homozygous SNPs. This study was performed on two human keratinocyte cell samples. Both samples were divided into three plates. The first plate did not undergo UV treatment, while the other two plates were treated with 250 and 500 mJ/cm2, respectively. Afterwards they were sequenced with Illumina Hi-seq 1000, producing from 23 to 45 million reads (depending on the sample), and aligned to the reference genome GRCh37.74 with GSNAP software. The variant calls were performed with Samtools software. The analysis were performed on variants called by at least 10 reads. The preliminary results on heterozygous SNPs detected 3 loci with different variation calls at the following positions: chr12:52841320 (KRT6B gene, after the treatment with 250 mJ/cm2), chr12:52745931 (KRT6B gene, after the treatment with 500 mJ/cm2) and chr3:195694460 (SDHAP1 pseudogene, after the treatment with 250 mJ/cm2). The first two loci were detected in the first sample whereas the variant at chr3:195694460 was detected in the second sample. The nucleotide at chr12:52841320 is located in the splicing site of the KRT6B gene and it resulted to be conserved in vertebrates while nucleotide at chr12:52745931 is in a non coding, not conserved KRT6B region. The nucleotide at chr3:195694460 resulted to be not conserved across vertebrates. In conclusion we used very stringent conditions to call genotypes from mRNA transcripts and preliminary results suggest the possibility to employ successfully RNA-seq for variant call detection. We plan to study in more detail the effects of UV radiations on keratinocyte cell samples extending the analysis on the entire spectrum of detectable variations that might be then validated by independent technologies.

1563S

TRRUST: A reference database of human transcriptional regulatory network. *H. Han, H. Shim, D. Shin, J. Shim, Y. Ko, J. Shin, H. Kim, E. Kim, T. Lee, H. Kim, K. Kim, C. Kim, S. Shin, I. Lee.* Biotechnology, Yonsei University, Seoul, South Korea.

Transcriptional regulatory networks (TRNs) that represent regulatory relationship from transcriptional factors (TFs) to target genes are important resources to study genetic regulatory programs for complex phenotypes in higher organisms including human. While many human TRNs inferred from various large-scale experiments are publicly available, only few reference databases for human TRNs have been published. Here we present an extensive database of human reference transcriptional regulatory network, TRRUST (Transcriptional Regulatory Relationships Unraveled by Sentence-based Text-mining), which contains 5,460 regulatory relationships between 646 TFs and 1,823 target genes identified by manual curation of 18,446 relevant sentences extracted from more than 15,000 abstracts. To collect the relevant sentences, which might contain information about human transcriptional regulation, we scan more than 20 million Pubmed abstracts for a set of key words implying human transcriptional regulation. TRRUST is the largest reference database for TF-target links based on literature curation to date to our best knowledge. TRRUST provide disease & biological process ontology terms that are enriched with target gene sets of TFs. Also we constructed a network of TF cooperativity, which might be useful to study co-regulatory complex. TRRUST will provide a reference database of TF-target relationships for validation of large-scale mapping of TF-target links and learning new TF-target links from biological data mining. TRRUST data is freely available at http://www.inetbio.org/trrust.

1564M

MokaSeq: Initial validation of the sequence analysis module of an NGS software platform for clinical diagnostics. *J.W. Ahn[1], N. Chandler[2], N. Parkin[2], M.A. Simpson[3], S.C. Yau[2], C. Mackie Ogilvie[1], K.J.P. Ryan[2].* 1) Genetics, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom; 2) Genetics, Viapath, London SE1 9RT, United Kingdom; 3) Division of Genetics and Molecular Medicine, Kings College London, London SE1 9RT, United Kingdom.

Guy's and St Thomas' NHS Foundation Trust and Viapath represent a centre of excellence in clinical genetics, servicing the South East Thames region population of ~6 million. We currently offer NGS-based testing for a number of heterogeneous genetic disorders. For each of these tests, DNA libraries are prepared targeting specific genes using custom Agilent SureSelect kits and sequenced on the Illumina MiSeq and HiSeq systems. As part of the newly formed bioinformatics core within Viapath, we are developing a bespoke NGS software platform (Moka) to perform high quality analysis within a robust and scalable framework that can be adapted in line with diagnostic demands. The Moka platform will provide an end-to-end system from patient test request to diagnostic report, and will include components for LIMS, sequence analysis, a variant database, variant interpretation, and cross-referencing/auditing our resulting data. Moka's sequence analysis module (MokaSeq) uses the Novoalign aligner alongside Picard Tools to carry out BAM file processing, and Samtools for variant calling of SNVs and small indels. Variants not relating to the regions of interest specified by our library panels are filtered out with Vcftools and the remaining variants are then annotated using Annovar, and a compendium of variant attribute repositories. CNV detection is performed using the R package ExomeDepth. Here we report findings from our initial validation of the sequence analysis module; this validation investigated the efficacy of detecting SNVs and CNVs previously identified by NGS and confirmed by Sanger sequencing/MLPA. To date, Moka has shown robust detection of all Sanger confirmed SNVs (n= 114) at a coverage of 30x or greater. Similarly all previously confirmed CNVs have also been detected, albeit from high coverage data (~6000x). Downsampling BAM files from CNV-containing samples by a factor of 5, 10, 50 and 100 to simulate coverage depths of 1200x, 600x, 120x and 60x respectively did not affect the capacity to detect any of the CNVs. However, reduction in coverage to 120x and 60x did result in the detection of a number of spurious false positives.

1565T

Combining callers across different sequence contexts improves somatic SNV detection. *K. Arora[1], E. Grabowska[1], R. Darnell[1,2,3], T. Bloom[1], V. Vacic[1].* 1) New York Genome Center, New York, NY; 2) Howard Hughes Medical Institute, New York, NY; 3) The Rockefeller University, New York, NY.

Somatic single nucleotide variants (SNVs) are the most frequent class of driver mutations in majority of common cancer types and consequently the most studied group of somatic alterations. Although SNVs are comparatively easy to detect using sequencing technologies, initial comparisons showed a poor overlap between callsets and necessitated development of ad hoc filters, extensive manual curation and validation of called variants, and motivated ongoing development of better methods for somatic variant calls. Despite considerable improvements over the initial algorithms, calling SNVs from whole genome tumor-normal pairs remains an open problem, aggravated by the issues of tumor purity and clonality, and difficulties in distinguishing between true low frequency mutations and sequencing errors. We conducted a systematic analysis of eight popular SNV callers using synthetic datasets as well as whole genome, exome and validation data from the TCGA glioblastoma, acute myeloid leukemia and thyroid carcinoma projects, with the goal of finding the best combination of callers to be included in our pipelines. Based on data from the Platinum Genomes resource that included a trio sequenced to 200x depth and exomes sequenced in house, we combined random lanes of a single sample into mock tumor and matched normal pairs and analyzed false positive calls. In our experiments different callers returned between 900 and 60,000 false calls per genome pair, indicating a lower bound of specificity in SNV calling. In order to estimate sensitivity, we followed the previously described "virtual tumor" approach where a synthetic tumor is generated by swapping reads between two germline samples at sites known to have discordant genotypes. Having a controlled ground truth dataset, built around a systematic grid of variant allele frequencies and read depths, allowed us examine type I and II errors across the genome as well as within local sequence and technical covariates contexts. Most somatic calls have poor calling rates on the extremes of the variant allele frequency spectrum; the same is true of the variant scores range. The analysis of local contexts of SNV calls points to differences in calling errors across different methods. This suggests a conditional combination method for calling SNVs, which in our preliminary experiments improved both sensitivity and specificity of calling SNVs, an important first step towards understanding the mechanisms leading to oncogenesis.

1566S

Functional interpretation of variant and non-variant positions in whole-exome sequencing data. *M. Delledonne[1,2], M. Garonzi[1], F. Griggio[1], C. Centomo[1], L. Xumerle[1], L. Venturini[1], A. Minio[1], A. Dal Molin[1], G. Martinelli[3], I. Iacobucci[3], B. Salisbury[4], M. Harvey[4], M. McManus[4], E. Zago[2], A. Ferrarini[1].* 1) Department of Biotechnology, University of Verona, Verona, Italy; 2) Personal Genomics s.r.l., Verona, Italy; 3) Department of Hematology and Oncological Sciences 'L. and A. Seragnoli', Institute of Hematology L. and A. Seragnoli', University of Bologna, Bologna, Italy; 4) Knome Inc., Cambridge, MA, USA.

Hematological diseases are malignancies with an high etiologic heterogeneity. These diseases may also encompass considerable heterogeneity in risk and in response to therapeutic treatments among subsets of patients. Within the project "Next Generation Sequencing platform for targeted Personalized Therapy of Leukemia" (NGS-PTL; FP7 european grant) we are involved in the setup of a bioinformatic pipeline for the identification of sequence variants as leukemia disease markers and therapy response markers from the analysis of Next Generation Sequencing (NGS) data. Methods for the analysis of whole-genome and -exome data are currently based on detection of variant sequences by comparison of sequencing reads with the reference human genome which are then stored in the Variant Call Format (VCF). However, the reference sequence contains both common and rare disease risk variants, including rare susceptibility variants for acute lymphoblastic leukemia and the Factor V Leiden allele associated with hereditary thrombophilia (Chen et al. 2011). In fact, out of 16,400 variant positions associated with disease, more than 4,000 variants are represented in the reference genome by the minor allele (Dewey et al. 2011). This poses a seriuos limitation to a comprehensive evaluation and detection of markers related for example to response to drugs and treatments. To overcome this problem we are developing, in collaboration with Knome Inc., an integrated hardware-software platform for the annotation and interpration of variation in genomes/exomes. The platform uses the emerging genome Variant Call Format (gVCF) which allows to store the information of both variant and non-variant positions in the genome/exome. Analysis of patients affected by disease with the pipeline based on knome system (knoSYS 100) identified about 500 variants represented in the genome by a minor allele (MAF < 5%) per patient. These variants will be correlated with clinical data as they may represent important candidates as molecular markers of predisposing factors of disease and of response to drugs. Chen R, Butte AJ (2011) The reference human genome demonstrates high risk of type 1 diabetes and other disorders. Pac Symp Biocomput. pp 231- 242. Dewey, F. E., Chen, R., Cordero, S. P., Ormond, K. E., Caleshu, C., Karczewski, K. J., ... Ashley, E. a. (2011). Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. PLoS Genetics, 7(9), e1002280.

1567M

Reducing INDEL errors in whole-genome and exome sequencing. *H. Fang[1,2], G. Narzisi[3], J.A. O'Rawe[1,2], Y. Wu[1,2], M.C. Schatz[3], G.J. Lyon[1,2].* 1) Stanley Institute for Cognitive Genomics, One Bungtown Road, Cold Spring Harbor Laboratory, NY, USA; 2) Stony Brook University, 100 Nicolls Rd, Stony Brook, NY, USA; 3) Simons Center for Quantitative Biology, One Bungtown Road, Cold Spring Harbor Laboratory, NY, USA, 11724.

Background: INDELs, especially those disrupting protein-coding regions of the genome, have been associated with human diseases. However, there are still many errors with INDEL variant calling, with these errors being driven by library preparation, sequencing biases, and algorithm artifacts. We have recently developed and reported a new INDEL-calling algorithm, Scalpel, with substantially improved accuracy. Results: We have analyzed simulation, whole genome sequencing (WGS), whole exome sequencing (WES), and PCR-free sequencing data from the same samples to investigate the properties of false-positive and false-negative INDEL errors. Simulation data show that Scalpel can achieve sensitivity of 90% at 30X mean coverage, while maintaining a reasonable false discovery rate (0.06%) for large INDELs. We developed a classification scheme for INDELs utilizing extensive high-depth sequencing validation data, and we show that low-quality INDELs have ~2.7-fold higher error rates than high-quality INDELs (32% vs. 12%). The mean concordance of INDEL detection between our WGS and WES data is ~52%, while WGS data uniquely identifies ~10.8-fold more high-quality INDELs, and WES data uniquely identifies ~1.9-fold more low-quality INDELs. INDELs called within both datasets are designated as "high confidence INDELs". Low-quality homopolymer A/T (Poly-A/T) INDELs are enriched in the WES-specific call set (82.6%), relative to the high confidence (44.2%) and WGS-specific call sets (45.1%). Concordance of INDEL detection between with-PCR and PCR-free data is ~71%, while PCR-free data uniquely yields ~4.4-fold more high-quality INDELs and with-PCR data uniquely yields ~6.3-fold more low-quality INDELs. We demonstrate that these types of INDEL errors are significantly reduced with a PCR-free library protocol, implying that these errors are introduced with PCR amplification. We have also calculated that ~95% of INDELs can be detected with 60X mean coverage WGS data using 2*100 bp reads from the Illumina HiSeq 2000 platform. Accurate detection of heterozygous INDELs naturally requires ~1.2-fold higher coverage relative to that required for homozygous INDELs. Conclusions: Homopolymer A/T INDELs are a major source of low quality INDELs and multiple signature events, and these are highly enriched in the WES data. We recommend WGS for human genomes at 60X mean coverage with PCR-free protocols, which can substantially improve the quality of personal genomes.

1568T

Single-nucleotide mosaicisms in whole-genome sequences of clinically unremarkable individuals. *Y. Huang[1], X. Xu[2], Y. Ye[3], Q. Wu[3], L. Yan[1], B. Zhao[3,4], X. Yang[3], Y. He[3], S. Wang[3], B. Gu[3], H. Zhao[1], M. Wang[1], H. Gao[1], G. Gao[1], Z. Zhang[2], X. Yang[2], X. Wu[2], Y. Zhang[2], L. Wei[1,3].* 1) Center for Bioinformatics, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, People's Republic of China; 2) Peking University First Hospital, Peking University, Beijing 100034, People's Republic of China; 3) National Institute of Biological Sciences, Beijing 102206, People's Republic of China; 4) Graduate School of Peking Union Medical College, Beijing 100730, People's Republic of China.

Postzygotic single-nucleotide mosaicisms (SNMs) have been studied in cancer and a few other overgrowth human disorders at whole-genome scale and found to play critical roles. However, in clinically unremarkable individuals, SNMs have never been identified at whole-genome scale largely due to technical difficulties and lack of matched control tissue samples, and thus the genome-wide characteristics of SNMs remain unknown. We developed a new Bayesian mosaic genotyper and a series of error filters, using which we were able to identify 17 SNM sites from whole-genome sequencing of peripheral-blood DNAs from three clinically unremarkable adults. The SNMs were thoroughly validated using pyrosequencing, Sanger sequencing of individual cloned fragments, and multiplex ligation-dependent probe amplification. In addition to peripheral-blood samples, the identified SNMs were also detected in the majority of other samples non-invasively obtained from the same donors, implying the whole-body presence of SNMs.

1569S

A Next-Gen Sequencing Software Workflow for Gene Panel Validation Control. *M. Keyser[1], J. Carville[1], T. Schwei[1], T. Durfee PhD[1], A. Pollack-Berti PhD[1], D. Nash[1], J. Stieren[1], S. Baldwin[1], R. Nelson PhD[1], K. Dullea[1], J. Schroeder[1], P. Pinkas PhD[1], G. Plunkett III PhD[1,2], F. Blattner PhD[1,2,3].* 1) DNASTAR, Inc., Madison, WI; 2) University of Wisconsin, Department of Genetics, Madison, Wisconsin, USA; 3) Scarab Genomics LLC, Madison, Wisconsin, USA.

DNASTAR offers an integrated suite of software for assembling and analyzing sequence data from all major next-generation sequencing platforms. DNASTAR software supports a variety of key workflows on a desktop computer. A new Gene Panel Validation Control workflow supports several types of data sets, including Ion Torrent AmpliSeq™ Comprehensive Cancer Panel, Illumina TruSight Cancer Panel, as well as custom gene panels and evaluates the efficacy of gene panel targeting and the accuracy of variant calling. The accuracy of gene panel targeting is determined by multiple factors, including specificity of primers and probes used for gene panel design, efficiency of the sequencing technology, accuracy of assembly, accuracy of SNP calling, and the SNP filters applied. DNASTAR's SeqMan NGen and ArrayStar programs provide an accurate alignment algorithm and variant caller. They then utilize a validated SNP set in the form of a VCF file and a BED or Manifest file that specify targeted regions, to calculate SNP-calling sensitivity, specificity, and accuracy. By utilizing this workflow, users can ultimately validate their entire process to verify that their known variants are being identified.

1570M

Whole Genome Sequencing of 30 Admixed Brazilians. *M. Machado[1], R. Moreira[1], A. Pereira[5], M. Barreto[2], B. Horta[3], M.F Lima-Costa[4], A. Horimoto[5], N. Esteban[5], M. Scliar[1], G. Soares-Souza[1], R. Zamudio[1], M. Santolalla[1], C. Zolini[1], F. Kehdy[1], M.R. Rodrigues[1], W. Magalhães[1], E. Tarazona-Santos[1], The Brazilian EPIGEN Consortium.* 1) Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; 2) Universidade Federal da Bahia, Brazil; 3) Universidade Federal de Pelotas, Brazil; 4) Fundação Oswaldo Cruz, Centro de Pesquiça René Rachou, MG, Brazil; 5) Instituto do Coração, Universidade de São Paulo, Brazil.

As part of the Epigen-Brazil Initiative on population genomics and genetic epidemiology, we resequenced the complete genome of 30 Brazilian admixed individuals randomly selected from each of three Brazilian population-based cohorts with different histories of admixture: Salvador from North East (n= 10), Bambuí from the South East (n=10) and Pelotas from Southern Brazil (n=10). Salvador has predominant African ancestry (51%), 43% of European and 6% of Native American ancestry. Bambuí and Pelotas are predominantly European (>76%), with 16% of African and 7-8% of Native American ancestry. Sequences were obtained by paired-end reads strategy using Illumina technology. On average, each genome was sequenced 42.7X, with 128 Gbases that successfully passed filter and aligned to the reference HumanNCBI37_UCSC genome, 82% of bases that showed a quality QScore>=30, 96% of Non-N reference bases with a coverage >=10X. Agreement with Illumina HumanOmni2.5 genotyping was 99.26 % and with HumanOmni5 was 99.53%. We identified 15,033,510 SNVs and 1,479,746 of these are new (not present in dbSNP nor in 1000Genomes databases on May 2014). Importantly, 989 of these have a non-reference allele frequency >0.50 in our sample of 60 haploid genomes. Most of SNVs were classified as intergenic (58.03%) or intronic (34.88%) variants, the remaining variants were classified in other functional categories. Considering only exonic SNPs, similar proportions of non-synonymous and synonymous SNPs were identified: 49.91% and 47.88% respectively, a result that matches other studies. By comparing the allele frequency spectra obtained by WGS with those from the Illumina HumanOmni2.5 on the same cohorts, we observed that for this array, rare SNPs are under-represented in the population with higher African ancestry (Salvador). Thus even if this array was inspired in a multiethnic database (1000Genomes) and enriched for rare alleles, a population-based bias persists. Finally, we observed 9,143 deleterious variants (CONDEL score > 0.51) and we are now analyzing how admixture and the demographic history of these populations affected the additive and recessive genetic load in the Brazilian sample. Funding: Brazilian Ministry of Health/FINEP.

1571T

IntSplice: A tool to predict aberrant splicing of an SNV at intronic positions -50 to -3. *K. Ohno, A. Shibata, T. Okuno, M.A. Rahman, Y. Azuma, A. Masuda.* Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan.

Precise spatiotemporal regulation of splicing is mediated by splicing cis-elements on pre-mRNA. Single nucleotide variations (SNVs) affecting intronic cis-elements potentially compromise splicing, but no efficient tool has been available. First, we analyzed the effect size of each intronic nucleotide on alternative splicing events annotated in the ENSEMBL database. We found that nucleotides at intron -13 to -5 (Int-13:Int-5), Int-3, exon +1 (Ex+1), and Ex+2 were critical splicing determinants. We further extracted a total of 111 parameters that possibly dictated the strength of splicing signals. Second, we calculated percent-splice-in (PSI) at all 3' splice sites of 14 RNA-seq data of normal human tissues and generated support vector regression (SVR) models with the 111 parameters. However, the correlation efficiencies between the calculated and predicted PSI's were less than 0.3. Third, we generated support vector machine (SVM) models to directly differentiate pathogenic SNVs in the human gene mutation database (HGMD) and normal SNVs in dbSNP using the 111 parameters as well as PSI's predicted by the SVR models above. HGMD had 1,162 SNVs (mutations) at Int-50:Int-3. We randomly chose 1,162 SNVs among 63,605 SNVs in dbSNP. A total of 2,324 SNVs were randomly divided into five groups, and four were employed to generate a SVM model and the remaining one was used to validate the efficiency. We repeated the modeling and validation five times. The generated SVM models yielded sensitivities of 0.800 ± 0.041 (mean and SD) and specificities of 0.849 ± 0.021. We compared efficiencies of our models with PSSM and MaxEnt, and found that sensitivity as well as the sum of sensitivity and specificity of our models were better than those of PSSM and MaxEnt. We serially introduced "A" nucleotide at Int-11:Int-3 in RAPSN minigene, and introduced them in cultured cells. RT-PCR analysis revealed that three out of the nine mutants caused aberrant splicing. We found that these mutants lost binding to U2AF65. We tested these nine mutants with our IntSplice model and found that splicing consequences were properly predicted in eight mutants. We created a web service program, IntSplice (http://www.med.nagoya-u.ac.jp/neurogenetics/IntSplice) to predict pathogenic SNVs at Int-50:Int-3. We hope that IntSplice will disclose yet unidentified splicing mutations and unveil pathomechanisms of human diseases.

1572S

Next generation sequencing approach to molecular diagnosis of auto-inflammatory diseases: from gene panel design to variant call. *M. Rusmini[1], F. Caroli[1], A. Grossi[1], L. Tattini[2], A. Magi[3], R. Ravazzolo[1, 4], A. Martini[4,5], M. Gattorno[5], I. Ceccherini[1].* 1) U.O.C. Genetica Medica;Istituto G Gaslini, Genova, Genova, Italy; 2) Dipartimento di Neuroscienze, Psicologia, Area del Farmaco e Salute del Bambino, Ospedale Pediatrico Meyer, Firenze, Italy; 3) Dipartimento medicina sperimentale e clinica, Università degli Studi di Firenze, Firenze, Italy; 4) Dipartimento di Neuroscienze, Oftalmologia, Genetica e Materno Infantile (DINOGMI), Università di Genova, Genova, Italy; 5) Pediatria II, Reumatologia, Istituto G Gaslini, Genova, Italy.

Auto-Inflammatory Disorders (AIDs) are a heterogeneous group of monogenic diseases caused by primary dysfunction of the innate immune system. Previous diagnosis by Sanger sequencing performed on a restricted number of genes has resulted time and cost consuming in addition to fail in detecting mutations in around 86% of patients recruited to our Unit. Clinical misdiagnosis, mutations in untested gene regions, genetic heterogeneity are possible explanations. We thought the Next Generation Sequencing approach could be the solution. By using the Ion AmpliseqTM Designer (LifeTech) online tool, 203 amplicons were included in the screening panel, for a total of 121 exons and 22 Kb of target DNA. The design has been performed on 11 genes mainly involved on AIDs, as reported in the database for these disorders, the Infevers database. Ion PGMTM has been used for runs and the mean coverage has turned out to be 347X, with 92.5% of amplicons at > 20X and 79.5% at > 35X. The analysis from FastQ to VCFs was carried out using 3 different workflows: i) Ion Torrent Alignment and Ion Reporter™ 4.0, specific for data generated by PGM ii) in-house pipeline based on open source tools like BWA and GATK, iii) CLC Bio software. We first focused on a set of 50 DNA samples already genotyped for the respective causative genes with the intention of comparing the three pipelines of analysis and assessing their sensitivity and specificity. Variants detected in genes which were not analyzed before have been validated by Sanger sequencing and differences in the variant calls from the three pipelines investigated. Through our analysis we found that the in-house pipeline has returned the most reliable results, missing only two of the 65 expected variants. However, this approach detected 34% of false positives, compared to the Ion Reporter™ and CLC Workbench which called only 2.5% of false positive variants. Moreover, only 2 of the 77 unexpected allelic variants detected with all the three pipelines and sequenced by Sanger have not been validated. Genetic diagnosis of patients affected by one of the Auto-Inflammatory Diseases will become more accurate by using these NGS approaches. Moreover, the complete analysis of all the genes present in the panel will allow us to correlate genotypes with the vast range of phenotypes, in addition to the assessment of allele frequencies of the variants detected and possibly involved in the pathologies.

1573M

From NGS back to Sanger Sequencing: Connecting and Synchronizing NGS and CE Variant Files with the Primer Designer Tool. *E. Schreiber, S. Berosik, M. Wenz, S. Chang, K. Gordeeva, S. Sharp.* Thermo Fisher Scientific, South San Francisco, CA.

Whole exome or panel sequencing projects performed by next generation sequencing (NGS) technologies typically reveal a large number of variants which may require verification by an orthogonal method. To that end, automated fluorescent Sanger sequencing is the method of choice since it is accurate, affordable and easy to perform. To facilitate the re-sequencing of any exon in the human genome we have recently made available to the scientific community a free to use tool called Primer Designer™ . The tool provides the designs for over 350.000 PCR primer pairs that cover 99% of all exons in the human genome. Amplicons generated with these primers can be readily sequenced using the BigDye® Direct, BigDye® Terminator v1.1 and v3.1 sequencing chemistries on the 3500xl Genetic Analyzer capillary electrophoresis platform. For variant identification the sequencing trace files are analyzed with Applied Biosystems Variant Reporter® Software which requires the import of a text file with a reference sequence for sequence alignment and comparison. Here we show the utility and workflow of a new on-line tool called VR Toolkit™ within the Primer Designer portal that generates a reference file from Primer Designer - derived PCR amplicons that contains the coordinates of the chromosomal location. Use of this annotated reference file in Variant Reporter® Software allows the generation of an output file that can be compared and matched to a variant call file (vcf) from an NGS instrument. The VR Toolkit enables the connection and synchronization between NGS data and traditional Sanger sequencing data analyzed with Variant Reporter® Software and should be of benefit for all researchers seeking to validate NGS data by Sanger sequencing.

1574T

A practical method to detect SNVs and indels from whole genome and exome sequencing data and an importance of in-house data for variant filtering. *D. Shigemizu, A. Fujimoto, S. Akiyama, T. Abe, K. Nakano, KA. Boroevich, Y. Yamamoto, M. Furuta, M. Kubo, H. Nakagawa, T. Tsunoda.* IMS, RIKEN, Yokohama, Kanagawa, Japan.

The recent development of massively parallel sequencing technology has allowed the creation of comprehensive catalogs of genetic variation. However, due to the relatively high sequencing error rate for short read sequence data, sophisticated analysis methods are required to obtain high-quality variant calls. Here, we developed a probabilistic multinomial method for the detection of single nucleotide variants (SNVs) as well as short insertions and deletions (indels) in whole genome sequencing (WGS) and whole exome sequencing (WES) data for single sample calling. Evaluation with DNA genotyping arrays revealed a concordance rate of 99.98% for WGS calls and 99.99% for WES calls. Sanger sequencing of the discordant calls determined the false positive and false negative rates for the WGS (0.0068% and 0.17%) and WES (0.0036% and 0.0084%) datasets. Short indels were also identified with high accuracy (WGS: 94.7%, WES: 97.3%). Currently we have applied this variant calling method to approximately 1,400 Japanese individuals and developed our own in-house database. Our own in-house database efficiently played a key roll in variant filtering for exome sequence analysis. We believe our variant method and own in-house database can contribute to the greater understanding of human diseases.

1575S

Group-based Variant Calling for a Large Cohort of Human Whole Genomes Leveraging Next-Generation Supercomputing. *K.A. Standish[1,2], T.M. Carland[2,3], G.K. Lockwood[4], W. Pfeiffer[4], M. Tatineni[4], C.C. Huang[5], S. Lamberth[5], Y. Cherkas[5], C. Brodmerkel[5], E. Jaeger[5], L. Smith[5], S. Szalma[5], G. Rajagopal[5], M. Curran[5], N.J. Schork[2].* 1) Biomedical Sciences Graduate Program, University of California, San Diego; 2) Department of Human Biology, J. Craig Venter Institute, La Jolla, CA; 3) The Scripps Research Institute, La Jolla, CA; 4) San Diego Supercomputer Center, University of California, San Diego; 5) Janssen R&D, LLC.

Recent advances in DNA sequencing technologies have increased the efficiency of sequencing, leading to an expanding deluge of high-quality data. However, DNA sequence reads have limited biological utility without relevant downstream processing and analysis, including read-quality assessment, alignment to a reference genome, variant identification, and individual genotyping. While the tools for performing these steps have improved, processing a whole genome from reads to variants remains an expensive and time-consuming aspect of sequencing studies. Furthermore, the accuracy of genotyping depends on accurate assessments of sequencing errors, which can be enhanced by processing large numbers of sequencing reads simultaneously. Here, we describe an efficient approach for obtaining high-quality variant calls from 438 whole genomes sequenced on an Illumina Hi-Seq 2500 platform. We exploited a group calling approach to minimize specific genotype assignment errors that arise from ignorance of sequencing error rates and inconsistent coverage. To accommodate the extraordinary computational and storage requirements associated with the scale and approach of this study, we used SDSC's Gordon supercomputer. It has 1,024 compute nodes, each with 16 cores, 64GB of DRAM, and 300GB of solid-state flash memory. To minimize cost, we optimized each step involved in genotyping to run as efficiently as possible on Gordon, parallelizing steps as appropriate during alignment and subsequent processing steps. To circumvent an I/O issue that arose while sorting the aligned reads, we used two "BigFlash" nodes with 4TB of SSD flash memory for temporary storage. Variants were then called with GATK's HaplotypeCaller in groups of 20-24 genomes. We found evidence that the ancestral makeup of a group influenced the accuracy of variant calls so groups were comprised of individuals with similar ancestral background. This approach provided increased power for variant detection while avoiding the prohibitively long computing time associated with large group variant calling. Our results suggest that this approach yields high-quality variant calls and genotype assignments in an efficient manner and highlights the need for sophisticated computational strategies in analyzing large numbers of human genomes.

1576M

Repeat-Aware Hidden Markov Models for the Comprehensive Joint Calling of SNPs, Indels, and Short Tandem Repeats. *A. Tan, H.M. Kang.* University of Michigan, Ann Arbor, MI.

Accurate detection of genetic variants in Next Generation Sequencing data has always been complicated by the presence of low complexity regions of the genome. While many indels located in regions with high sequence diversity are easy to assay, the challenge is in the large fraction (~50%) of indels that are located in low complexity regions of the genome, mostly in the form of short tandem repeats (STRs), often with inexact repeat units. In the vicinity of indels in low complexity regions, it is also difficult to classify nearby base mismatches as a SNP or as part of an inexact repeat within an STR.

To be an effective short variant caller, the calling algorithm must account for the inherent heterogeneous nature of indels. We thus developed a variant calling algorithm based on a set of novel repeat-aware Hidden Markov Models (raHMMs) that locally aligns a pair of sequences allowing for mismatches or indels, with an arbitrary number of repeat units.

For variant discovery, our approach first identifies the repeat unit in a candidate variant allele; using this candidate repeat unit, we then perform pairwise alignments using a pair of raHMMs to determine the most appropriate flanking sequences, even in the presence of inexact STRs. This procedure allows us to systematically define variants by their flanks and earmarks regions that a SNP should not be called indiscriminately.

For variant genotyping, we apply another raHMM that explicitly models both flanking sequences and the repeat unit. This model allows us to determine if a read contains the defining flanks of an indel, counts the number of repeats observed and keeps track of alternative alleles that are not explicitly modeled.

This calling algorithm not only allows us to better classify the variant types in repeat rich regions but also genotypes STRs with inexact repeat units. We demonstrate that our method comprehensively detects and genotypes SNPs, Indels and STRs using the deeply sequenced trios in the 1000 Genomes Project. This is implemented in vt and is available from http://genome.sph.umich.edu/wiki/Vt.

1577T

Describing complex rearrangements using HGVS sequence variation nomenclature, suggested extensions. *P.E. Taschner, J.T. den Dunnen.* Dept Human Genetics, Leiden Univ Medical Ctr, Leiden, Netherlands.

Breakpoints involved in translocation and chromothripsis are traditionally described using ISCN nomenclature based on chromosomal banding patterns (1). The sequence variation nomenclature guidelines of Human Genome Variation Society (HGVS, http://www.hgvs.org/mutnomen) traditionally focused on simple variants not requiring specific rules for detailed description of genetic rearrangements. This changed with the introduction of new technologies allowing rapid discovery of breakpoint sequences from complex structural rearrangements including translocations. The description of such complex variants challenges the existing guidelines. Previously, we suggested extensions for simple translocations (2). Here, we suggest extending the HGVS nomenclature guidelines to facilitate unambiguous description of more complex structural rearrangements including chromothripsis. A main requirement for the description is that precise chromosomal breakpoint sequences can be derived easily. The suggested format should provide sufficient flexibility and consistency limiting alternative interpretations and ambiguous descriptions. The new rules can be combined with those proposed previously for complex changes, which included: i) nesting to support description of changes within inversions and duplications, ii) composite changes to support concatenation of inserted sequences (3). We have applied the rules in practice by describing complex cases involving many breakpoints. The specifications should allow easy implementation in sequence variant nomenclature checkers (e.g. Mutalyzer, https://Mutalyzer.nl). We are planning to extend the functionality of Mutalyzer to incorporate the latest version of the HGVS sequence variation nomenclature guidelines as part of the development of curational tools for gene variant databases (Locus-Specific DataBases, LSDBs).

1) ISCN (2013). 2013. An International System for Human Cytogenetics Nomenclature. Shaffer LG, McGowan-Jordan J, Schmid M (eds). Basel: Karger.

2) http://www.hgvs.org/mutnomen/SVtrans_HGVS2013_PT.pdf

3) Taschner PE, den Dunnen JT. Hum Mutat. 32:507-511 (2011).

1578S

Comparing variant filters from transcriptome and exome sequencing data. *N. Thomson, C. Boysen, A. Arens, A. Heine, U. Appelt, A. Joecker, S. Monsted, B. Knudsen, R. Forsberg.* CLC division of QIAGEN, Fort Myers, FL.

Targeted gene panels, whole exome sequencing (WES), whole genome sequencing (WGS), and transcriptome sequencing (RNA-Seq), are revolutionizing cancer research. Many analyses, such as the comparison of tumor and normal cells, include the comparison of variant calls in the resulting VCF files. Here, we illustrate the benefits of leveraging read mappings in the comparison of WES and RNA-seq data using the CLC Cancer Research Workbench. WES and RNA-seq reads (paired-end, Illumina GA II, 100bp) from two patients (MM065 and MM089) are analyzed with built-in workflows in CLC Cancer Research Workbench. The sequences are available from SRA (bioproject 182345)1. When filtering the results for tumor-specific variants inducing an amino acid change 42 and 43 variants from WES and 827 and 1,497 variants from RNA-seq are identified in MM065 and MM089, respectively. A function in the Cancer Research Workbench, “Compare Shared Variants within a Group of Samples” identifies Variants from both WES and RNA-seq. Only two and five shared WES/RNA-seq variants, respectively, are identified for the two patients. “Identify Known Mutations from Sample Mappings”, searches RNA-seq read mappings for variants previously called from the exome data. This approach identifies 3 exome variants in the RNA-seq data from patient MM065 and 11 of the exome variants in the RNA-seq data of patient MM089. In contrast to comparing VCF files, the “Identify Known Mutations from Sample Mappings” tool provides a more reliable approach to recover previously detected variants in the WES data directly from the RNA-seq read mappings. Using this approach, 3 out of 42 variants for MM065 and 11 out of 43 variants for MM089 are confirmed by RNA-seq. 1Harbour JW et al., “Frequent mutation of BAP1 in metastasizing uveal melanomas”, Science, 2010 Nov 4;330(6009):1410-3.

1579M

Combining sets of indels with improved specificity and sensitivity using BAYSIC. *D. Weaver[1], B. Cantarel[1, 2], G. Benstead-Hume[1], V. Benstead-Hume[1], A. Mackey[3], J. Reese[1].* 1) Genformatic, Austin, TX; 2) Baylor, Scott & White Baylor Institute of Immunology, Dallas, TX; 3) University of Virginia. Charlottesville, VA.

Insertions and deletions ("Indels") comprise a source of human genome variation as significant as SNVs, and frequently cause clinically relevant phenotypes (Mulaney, et al., 2010). Moreover, the additional complexities of indel variation make their detection and accurate representation a more technically challenging problem than SNV detection. Many methods exist of the detection of indels, and as with SNVs, the concordance among sets of indels detected by different methods is poor. We previously described BAYSIC, a method of integrating variant calls using Bayesian statistics and latent class analysis to combine sets of SNVs with improved specificity and sensitivity. We extend this strategy here to combine sets of indels with increased sensitivity and specificity when compared to other currently popular methods for indel detection, e.g., GATK. BAYSIC provides the user with a posterior probability threshold that can be specified by the user to tune BAYSIC's performance according to the user's tolerance for false negative and false positive indel calls. Using genome data from the 1000 Genomes project, we evaluated the performance of BAYSIC and demonstrate BAYSIC's enhanced accuracy when compared to other popular indel detection methods.

1580T

CNV Detection Assessment. *J. White, X. Gai.* Mass. Eye & Ear Infirmary, Boston, MA.

Clinical assessment of patients at Mass. Eye and Ear includes diagnostic testing for pathogenic genetic variants. We assess both single nucleotide changes that result in missense or frameshift alterations, and structural/copy number changes that result from deletions or duplications. Our diagnostic rate based on SNP analysis alone is ~ 50%. In order to increase this rate we have used microarray chip based genotype analysis on the Illumina iScan platform. However, we wish to reduce costs by evaluating our exome sequence data to detect CNVs. To do so we compared CNV detection applications based on the following criteria: a) paired-end verses read-depth analysis, b) ease of installation and use, c) inclusion of baseline or matched-control samples, d) use of exome verses targeted exome sequence data, and e) detection of positive control samples. We have collected a number of data sets that have been found to include CNVs which have been confirmed to be associated with specific traits. We used these data sets as positive controls for detection of CNVs. We have so far compared XHMM, BreakDancer, and WaveCNV, with many more comparisons planned. We were unable to use WaveCNV due to its difficult installation and lack of documentation. We have evaluated whole exome sequence data in two ways: 1) comparison of family members against a group of "baseline" samples, and 2) comparison of family members against each other without baseline samples. •XHMM identified a deletion in samples known to have that particular deletion. However, XHMM was unable to identify the duplication in samples that had a known small duplications. •We evaluated the same family samples with BreakDancer. Neither the known deletion, nor the known duplication were identified by BreakDancer. BreakDancer generates a large amount of SV results that we filtered to search for known CNVs. •Inclusion of baseline samples did not improve CNV detection.

1581S

Cloud-based variation analysis using SRA sequencing data directly. *C. Xiao, E. Yaschenko, S. Sherry.* National Center for Biotechnology Information, National Library of Medicine, National Institute of Health, 45 Center Drive, Bethesda, MD 20892.

Variation analysis plays an important role in elucidating the causes of various human diseases. The drastically reduced costs of genome sequencing driven by next generation sequence technologies now make it possible to analyze genetic variations with hundreds or thousands of samples simultaneously, but currently with the cost of ever increasing local storage requirements. The tera- and peta-byte scale footprint for sequence data imposes significant technical challenges for data management and analysis, including the tasks of collection, storage, transfer, sharing, and privacy protection. Currently, each analysis group facing these analysis tasks must download all the relevant sequence data into a local file system before variation analysis is initiated. This heavy-weight transaction not only slows down the pace of the analysis, but also creates financial burdens for researchers due to the cost of hardware and time required to transfer the data over typical academic internet connections. To overcome such limitations, we here introduce a cloud-based analysis framework that facilitates variation analysis using direct access to the NCBI Sequence Read Archive. We demonstrate that it is cost and time effective to make variant calls using both public and control-accessed SRA sequence data without first transferring the raw or aligned sequence data into a cloud storage environment. Performance of this pipeline will be assessed against a comparable cloud analysis pipeline that first transfers all data into local storage, and a customized machine image (swift1.0 AMI) with preconfigured tools and resources essential for variant analysis, created for instantiation in an EC2 instance or instance cluster on Amazon cloud, will be introduced.

1582M

Low false-positive rate chromosomal structural variation detection procedure with statistical comparisons between case and control using paired-end reads. *K. Yamagata[1], A. Yamanishi[2], C. Kokubu[2], J. Takeda[2], J. Sese[1].* 1) Ochanomizu University, Bunkyo-ku, Tokyo, Japan; 2) Osaka University, Suita-shi, Osaka, Japan.

Recent high-throughput sequencing technology has uncovered the high frequency of both somatic and germinal chromosomal structural changes. To detect the changes, various software tools such as BreakDancer and GASV have been developed. However, most of those tools tend to yield a high rate of false-positives, which hampers the full understanding of the chromosomal changes in the cells. We here introduce a structural variation (SV) detection method with an extremely low false-positive rate by using the statistical comparison between case and control sequences. To detect SVs between two individuals, most existing tools need to perform the detection on each individual, and subsequently compare them with each other. However, this strategy may suffer from the problem of how to assess the differences in sequencing depth, quality and base biases between the samples. Our method first generates lists of candidate breakpoints using multiple sources of read information such as paired-end distances and the number of reads on each direction, and then makes statistical comparison of the candidate regions between samples. To evaluate our method's performance, we carried out SV analysis of mouse embryonic stem cell (ESC) lines by comparing gamma-ray irradiated (case) and unirradiated (control) cell populations. The frequency of irradiation-induced SVs was enhanced by increased genomic instability of the engineered ESCs in which the expression of the Blm helicase gene was transiently suppressed by doxycycline administration (Yamanishi A. et al, Genome Research 2013). We performed whole-genome sequencing (WGS) of both case and control samples using Illumina HiSeq-2000 at 30x coverage, and mapped the paired-end reads to the mouse reference genome using Bowtie2. Using the data, Breakdancer detected 3,177 deletions, 197 inversions and 125 translocations; GASV detected 12,563 deletions, 66 inversions and 138 translocations, most of which turned out to be false-positives by referring to the mapping results. In contrast, our method detected 21 deletions, 22 inversions, and 9 translocations. Among them, we examined 15 rearrangement breakpoints by direct PCR sequencing and confirmed them all as true-positives. This result indicates that our method largely contributes to the reduction of false-positive detection of chromosomal SVs in WGS data. In our presentation, we also show the result of SV analysis when our method is applied to human sequences in 1000 genomes database.

1583T

Multiplexing strategies for HLA genotyping using DNA barcoding methods for SMRT® sequencing. *S. Ranade[1], K. Eng[1], J. Harting[1], B. Bowman[1], E. Rozemuller[2], N. Westerink[2], M. Penning[2].* 1) Pacific Biosciences, Menlo Park, CA; 2) GenDx, Utrecht, The Netherlands.

Fully phased allele-level sequencing of highly polymorphic HLA genes is greatly facilitated by SMRT sequencing technology due to its long read lengths and ability to sequence an amplicon representing a gene allele in isolation. In the present work, we evaluate DNA barcoding strategies for efficient pooling of multiple samples and loci for simultaneous sequencing of these genes. Multiplex sequencing using symmetric and asymmetric barcode-tailed HLA amplification primers were tested. Full-length HLA class I genes HLA-A, -B, and -C were evaluated, using 8 different 16-bp barcode sequences in symmetric and asymmetric pairing. Eight barcodes generated 28 unique asymmetric pairings and allowed for simultaneous sequencing of 28 genomic reference DNA samples; symmetric barcodes could be used for unique tagging of 8 genomic reference DNA samples. Following amplification, the symmetric and asymmetric barcode-tagged libraries were pooled into two separate libraries for sequencing. The data was analyzed both separately and together to evaluate the ability of the barcode tags to uniquely de-convolute the samples. Consensus sequences were generated using LAA protocol in SMRT analysis 2.2. Allele sequences were typed using GenDx NGSengine HLA-typing software. We demonstrate the use of a DNA barcode strategy for multiplex full-length HLA gene sequencing to provide allele-level genotyping along with SNP phasing information.

1584S

Detection of Copy Number Variations in Cancer Genomes from High Throughput Sequencing Data. *G. Klambauer, S. Hochreiter.* Insitute of Bioinformatics, Johannes Kepler University Linz, Linz, Upper Austria, Austria.

"Copy Number estimation by a Mixture Of PoissonS" (cn.MOPS), is a well established and widely used method for detection of germline copy number variations (CNVs) in high-throughput sequencing data. cn.MOPS showed excellent performance at the detection of CNVs in HapMap samples, as well as in genomes of bacteria, fungi and plants. Since cn.MOPS constructs a model across samples for each genomic position, it is not affected by read count variations along chromosomes, and, therefore, geared to targeted sequencing. In a comparative study, cn.MOPS was the best performing method at the detection of CNVs in targeted sequencing data. However, the detection of somatic CNVs in cancer genomes is still challenging due to admixture of normal and tumor tissue, nondiploidy and very large copy number variations that affect normalization. Therefore, preprocessing, normalization, and the core algorithm of cn.MOPS have been optimized for CNV detection in cancer genomes. We demonstrate the improved performance of the enhanced cn.MOPS algorithm for cancer genomes on whole genome sequencing data from the International Cancer Genome Consortium (ICGC). cn.MOPS has been optimized for computation time and parallelized, which makes the method perfectly suited to analyze data sets of hundreds of cancer samples within a few hours.

1585M

Efficient variant pipeline for diagnosis of inherited cardiomyopathies associated genes using Ion Torrent PGM™ platform. *L. Cerdeira[1], T.G M. Oliveira[2], A. Pereira[2], M. Mitne-Neto[1].* 1) Research and Development, Fleury Group, São Paulo, SP, Brazil; 2) University of São Paulo - Heart Institute, São Paulo, SP, Brazil.

Hypertrophic cardiomyopathy (HC) is a primary cardiac disease characterized by hypertrophy of the left ventricle (LV) without dilation , usually asymmetrical and predominantly septal , in the absence of any other cardiac or systemic disease that can cause myocardial hypertrophy . Typically, CH is caused by mutations in genes encoding sarcomeric elements. Currently 19 genes have been discovered and linked to the CH spectrum, besides the filaments of the sarcomere, additional subgroups can be classified as related CH, as Z disc genes and calcium transport. Diagnosis is mainly clinical and usually only identified after the symptoms beginning. For that reason molecular genetic tests came up as a differential tool for the discovery of the mutations causing the phenotype. This study developed a bioinformatics pipeline for accurate molecular diagnosis of CH using Ion PGM data. The pipeline was developed using CLCBio Genomic Workbench 6.5 workflow and had as a first step a mapping assessment, with the 5 nucleotides at the 3'end trimmed and a Phred ≥ 20 used for quality control. Alignment against the human genome HG19 version was done using standard thresholds, followed by identification of variants by coverage and quality positioning. The identification of known variants was validated against the databases: dbSNP and clinicalvar and for further evaluation a prediction of splice site effect and amino acid change. The result were submitted to SIFT and Polyphen programs to obtain the values for protein damaging. To validate the pipeline we selected 15 DNA samples from previously analyzed patients, which had clinical and molecular diagnoses of HC from the Heart Institute (InCor - University of São Paulo, Brazil). The previous molecular diagnosis was performed by Sanger sequencing for the three most HC-associated genes: MYH7, MYBPC3 and TNNT2. All variants found were properly annotated for the three genes and were further used in the evaluation of NGS accuracy. The NGS pipeline presented here could identify > 97% of the Sanger sequencing identified mutations, showing its robustness and viability for HC and for other diseases with Mendelian heritability standard.

1586T

Assessing novel centromeric repeat sequence variation within individuals by long read sequencing. *K.H. Miga[1], J. Chin[2], A. Bashir[3,4].* 1) Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA 95064, USA; 2) Pacific Biosciences, Inc, 1380 Willow Rd, Menlo Park, CA 94025, USA; 3) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, NY 10029, USA; 4) Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, NY 10029, USA.

Centromeres and other heterochromatic regions are commonly enriched with long arrays of near-identical tandem repeats, known as satellite DNAs, that offer a limited number of variant sites to differentiate individual repeat copies across millions of bases. This substantial sequence homogeneity challenges available assembly strategies, and as a result, centromeric regions are vastly underrepresented in genomic studies. Further, as these sites are known to be variable among individuals in the population, it is necessary to not only characterize the sequence organization of these regions in a single genome, but to develop high-throughput methods to study this new source of human sequence variation among individual genomes. To advance characterization in these regions we have designed alphaCENTAURI (centromeric automated repeat identification for alpha satellite DNA) that takes advantage of Pacific Biosciences' long reads from whole-genome sequencing. Long reads allow direct determination of satellite higher-order repeat structure as opposed to using indirect inference methods, like assembly, with reads shorter than the underlying lengths of the high order repeat unit. Here we demonstrate a comprehensive assessment of higher-order repeat patterns for two human cell lines: NA12878 (diploid) and the hydatidiform mole (CHM1, haploid-like) genomes. First, we show the reliability of the method by validating consistency with existing centromere repeat references. Additionally, we are able to identify changes in repeat unit directionality that exist within arrays and between individuals, representing polymorphisms in the population or errors within existent assemblies. The analysis also represents a robust and straightforward methodology for characterization of higher-order repeat variants within the array that differ between individuals. Based on this analysis, resolution of higher-order repeats could be readily performed at low depth and reasonable cost across a population, or in genomes without high-quality references. This study demonstrates the methods to generate a sequence survey for regions enriched in satellite DNA that are typically omitted from genomic studies. We believe it establishes a foundation to extend and improve genomic characterization of any higher-order repeat structure using long reads.

1587S

Anchored Assembly: An algorithm for large structural variant detection using NGS data. *J. Bruestle, B. Drees.* Spiral Genetics, Seattle, WA.

Statement of purpose Characterizing large indels, inversions, and multi-nucleotide variants is important for understanding cancer, bacterial pathogens, and neurological disorders. Standard pipelines often miss these variants. Spiral Genetics has developed Anchored Assembly, a novel method using direct, *de novo* read overlap assembly to accurately detect variants from next-generation sequence reads. Anchored Assembly's range of detection and low false discovery rates may be useful for characterizing structural differences between tumor and normal samples.

Methods used Anchored Assembly was evaluated against Pindel and BWA + GATK using simulated read data. Datasets were generated by populating chromosome 22 of the human genome reference sequence with a set of SNPs, insertions, deletions, inversions, and tandem repeats.

Summary of results On human chromosome 22 data, Anchored Assembly detected over 90% of indels and structural variants up to 50 kbp and SNPs with false discovery rates well below 1%. In comparison, Pindel and BWA + GATK had overall false discovery rates of 10% and 9%, respectively. We detect, on average, over 90% of indels and structural variants up to 30 kbp in non-repetitive regions. The ability to detect deletions and structural variants is undiminished by variant size, and the ability to accurately detect and assemble insertions continues well into the 30 kbp range.

1588M

Short inversion detection by splitting and re-aligning poorly mapped next-generation sequencing reads. *R. Chen, W. Yang, Y. Lau.* Paediatrics & Adolescent medicine, The university of Hong Kong, Hong Kong, Hong Kong.

Rapid development of sequencing technology has enabled routinely discovery of deletions and insertions. Unlike these two kinds of structural variation, characterization of inversions is left behind. Summary of public databases and researches shows that little short inversions with length smaller than 500bp have been detected for now. Because in contrast to small insertions and deletions, which are considered by gap alignment and recorded in primary mapping files, inversions short enough to interrupt alignment probably cause poorly mapped reads, which are mostly left out of consideration for structural variation detection by existing methods. And as a result, signals of short inversions are lost. Here, we introduce SRinversion, which tries to reuse those poorly mapped reads by splitting and re-aligning them for inversion detection and in this way, improve the resolution of inversion detection to less than 10bp. Comparison with previous structural variation detection methods, i.e. Pindel, BreakDancer, and DELLY, using simulated data indicates that SRinversion performs much better when inversion size is smaller than 100bp. Testing on chromosome 21 of a high-coverage parent-child trio (NA12878, NA12891, and NA12892) from 1000 genomes project shows that SRinversion is the only method that is able to detect inversions smaller than 100bp. As for inversions with length from 100bp to 1000bp, besides SRinversion, Pindel is the only method that have output, which only include 7 regions, indicating relatively low power in detecting small inversions. PCR will be performed on randomly selected inversions detected by SRinversion on each individual of the trio for further validation. The method was also tested on a single African individual (NA18507) with both high-coverage and low-coverage whole genome data to check the effect of coverage, which will helps broader the spectrum of inversion at the same time.

1589T

Reproducible and repurposable toolkit of structural variant callers applied to 3,751 whole genomes and 10,940 whole exomes. *S. Ma[1], S. Ambreth[2], A. Carroll[1], A. Sabo[2], P. Mishra[2], W. Salerno[2], A. English[2], N. Veeraraghavan[2], E. Boerwinkle[2,3].* 1) DNAnexus, Mountain View, CA; 2) Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX; 3) University of Texas Health Science Center, Human Genetics Center, Houston, TX.

Despite the significant influence human structural variation and copy number variation has on mendelian disease and cancer genomics, accurate calling of these variants remains a difficult problem. Great efforts have been made in understanding the concordance of SNP and small indel calls, allowing for groups to harden these pipelines for large scale research and clinical applications. A similar systematic approach to structural variant calls has until now remained lacking.

Baylor College of Medicine's Human Genome Sequencing Center (HGSC) has collaborated with DNAnexus to package together SVPipe which runs BreakDancer, CNVnator, CREST, DELLY, Genome STRiP, Pindel, and Tiresias for whole genome analysis, converting each output to a common format. A pair of CNV callers, CoNIFER and XHMM were also set up for exome analysis. The SVPipe multi-algorithmic approach was run on 3,751 low coverage whole genomes and the CNV callers were run on 10,940 exomes from the CHARGE project. These call sets provide a deeper understanding into the overlap between the tools and the amount of structural variation present in a large human population.

Thanks to HGSC's work, all of these callers are publicly available on the DNAnexus platform. Users can run any caller without the need to install or configure software and each has been optimized to improve the speed and cost of its execution. These tools are provided with a number of benchmarks sets to validate the results of pipelines which use them.

1590S

Orthogonal Resequencing Support of Structural Variation in a Personal Genome. *W.J. Salerno[1], A.C. English[1], C. Gonzaga-Jauregui[2], O.A. Hampton[1], S. Ambreth[1], D. Ritter[1], S. White[1], C. Davis[1], P. Mishra[1], Y. Liu[1], C.R. Beck[2], M. Dahdouli[1], N. Veerarraghavan[1], A. Hawes[1], D.A. Wheeler[1], J.G. Reid[1], D.M. Muzny[1], J. Rogers[1], K. Worley[1], A. Sabo[1], W.J. Salerno[1], J.R. Lupski[1], E. Boerwinkle[1], R.A. Gibbs[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX., USA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX., USA.

Personal genome sequencing is becoming standard in research and clinical settings, and discovery of genetic disease etiologies increasingly requires genomic characterizations beyond single nucleotide variants and small insertions and deletions. The human genome is subject to large (>100 bp) deletions, insertions, inversions, translocations, and complex combinations of these events. While the number of such structural variant (SV) events is far fewer than the number of single nucleotide variants, SVs collectively affect a comparable or even larger fraction of the genome, directly impacting gene function and regulation. Any effort to robustly characterize structural variation must address the broad variation in SV size, type, and complexity. We have developed and implemented a novel approach for integrating whole-genome array comparative genomic hybridization (aCGH) and short-read next-generation sequencing (NGS) data with long-read (Pacific BioSciences RSII), long-insert (Illumina Nextera), and whole-genome architecture (BioNano Irys) data. These methods are illustrated via analysis of the genome of a phenotypically well-characterized individual with autosomal recessive Charcot-Marie-Tooth neuropathy. The integration of these orthogonal methods was performed with Parliament, a consensus SV-calling infrastructure that merges and evaluates the results from existing and novel SV-detection software and multiple input data sources. In this individual, Parliament identified 33,8862 loci (composing 2.65% of the genome) that are inconsistent with the hg19 reference assembly. Of these, 7,460 are supported as putative SVs by local hybrid assembly and 3,440 are supported by long-read force calling or multi-source heuristics. These 10,900 putative SVs have an aggregate length of approximately 20 Mbp (0.6% of the genome), including 4,427 deletions, 4,442 insertions, and multiple complex events such as a single 40 Kbp chimeric event on chromosome 11 (p15.5) that individual NGS methods classify as independent inversion, deletion, and insertion events. These data represent the most comprehensive characterization of SVs in a personal genome and provide a platform for SV validation efforts. We also update on de novo whole genome assembly, comparison of SVs in hg19 and hg38, and Parliament workflows for trio, cancer, and whole-exome samples. The complete Parliament infrastructure and a WGS NGS-specific workflow are available to the community on the DNAnexus platform.

1591M

Copy Number Variation Analysis using Single Cell Sequencing. *X. WANG.* Program in Public Health, Departments of Preventive Medicine, Biomedical Informatics, and Applied Mathematics & Statistics, Stony Brook University, Stony Brook, NY 11794.

Recently developed single cell sequencing (SCS) technology can amplify and sequence the genome of a single cell. It is a promising new tool to study tumor heterogeneity and evolution in cancer studies. In this study, we focus on methods for the copy number variation (CNV) detection using the low-coverage single cell sequencing data. Traditional methods for CNV analysis (based on CGH array or bulk-cell sequencing) are impractical for single cell, especially for the low-coverage sequencing. We proposed to use a two-step adaptive scheme for a more accurate CNV detection of the noisy data. In the initial step (after GC content correction and sample normalization steps), we apply a fast CNV calling algorithm based on the modified circular binary segmentation (CBS). To ensure better coverage and reliability, the data from cells with similar ploidy level are then pooled to create a bulk-sample data—with high purity. The preliminary ploidy estimates are validated by the lesser allele frequency (LAF) information from mapped reads. We propose two novel methods for more precise breakpoint detection in the second step: one method is based on mixture of Poissons and which can be used to model depths of coverage across samples at each position; the other is based on scan statistics which overcome the limitation of fixed window size in the first step. We tested our methods on simulated data and a prostate cancer dataset. All the scripts and pipelines are included in SinSat: a single-cell sequencing analysis toolkit.

1592T

A Convergent Clinical Exome Pipeline Specialised for Targeted Gene Analysis. *J. Plazzer[1], A. Oshlack[2], C. Gaff[3], N. Thorne[3], G. Taylor[4], H. Dashnow[5], A. Lonie[5], M. Bahlo[6], T. Bakker[3], D. Bauer[7], K. Siemering[8], P. James[1], S. Sadedin[2], Melbourne Genomics Health Alliance.* 1) The Royal Melbourne Hospital, Parkville, Victoria, Australia; 2) Murdoch Childrens Research Institute, Parkville, Australia; 3) Melbourne Genomics Health Alliance, The University of Melbourne; 4) Centre for Translational Pathology, The University of Melbourne, Parkville, Australia; 5) Victorian Life Sciences Computation Initiative, Carlton, Australia; 6) Walter and Eliza Hall Institute of Medical Research, Parkville, Australia; 7) Division of Computational Informatics, CSIRO, Sydney; 8) Australian Genome Research Facility, Parkville, Australia.

Efforts to move high throughput sequencing into the clinic must confront many challenges including meeting clinical standards for cost, reproducibility, quality, ethical and privacy considerations. The Melbourne Genomics Health Alliance was formed from a diverse group of institutions with the aim of sharing the burden of these challenges through a common sequencing and bioinformatics platform. Critical to this concept is that regardless of disease, all patients undergo full exome sequencing, and targeting to a specific disease is performed in silico. While this has many benefits, it also adds significant complexity to the analysis pipeline as it must meet the differing diagnostic needs of diverse disease cohorts. In this work we present the Melbourne Genomics Health Alliance bioinformatics analysis pipeline which has been specifically designed as a clinical grade exome sequencing pipeline with built in support for multiple sub-target regions. The pipeline is based on the Bpipe platform and includes support for customized sets of targeted regions, prioritized genes, regions blacklisted against incidental findings, automatic exclusion of sequencing artefacts and population variants, PDF provenance and quality reports. It allows genes to be prioritized at both a disease cohort and individual level and combines this with output from annotation tools to produce a clinically interpretable report that is available as an Excel spreadsheet or optionally can be imported into any LOVD (Leiden Open Variant Database) instance for curation. Being based on Bpipe, it offers powerful features for running many samples in parallel either on a cluster or dedicated computing resources as well as full traceability through auditable provenance of every file.

1593S

An ensemble variant calling approach to 10,000 low coverage whole genomes. *Z. Huang[1], N. Rustagi[1], L.A. Cupples[2], D. Muzny[1], R. Gibbs[1], E. Boerwinkle[1,3], F. Yu[1], CHARGE Consortium.* 1) Human Genome Sequencing Center and Molecular and Human Genetics Department, Baylor College of Medicine, Houston, TX 77030, USA; 2) Boston University School of Public Health, Boston, MA 02118, USA and the National Heart, Lung and Blood Institute (NHLBI) Framingham Heart Study, Framingham, MA, 01702, USA; 3) Human Genetic Center, University of Texas Health Science Center, Houston, TX 77030, USA.

Accurate calling of common and rare variants is critical for population genetic analyses and novel gene discovery, as is consideration of single nucleotide variants, indels and short tandem repeats (STRs). There is growing consensus of the practical utility of large sample sizes having both deep whole exome sequencing (WES) and low coverage whole genome sequence (WGS), made efficiently from the same library preparation. The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium is aiming to identify novel genes (and other noncoding genomic motifs) influencing health and disease, and is accruing deep WES and low coverage WGS (6-7x) in ~10,000 deeply phenotyped individuals. In order to call and integrate variants across a large sample of low coverage genomes, multiple variant calling pipelines were used to avoid caller specific bias, and consensus filtering was adopted to select high quality variants. We utilized GATK UnifiedGenotyper (UG) and HaplotypeCaller (HC), GotCloud and SNPTools for SNP calling; GATK-UG, GATK-HC, Pindel for INDEL calling; and LobSTR for STR calling. We compared two consensus strategies: the 3-out-of-4 and a machine learning based consensus filtering. We built an automated pipeline including individual and consensus calling and quality assessment. This pipeline is computationally intensive, requiring a fully parallelizable design in a cloud environment (e.g. DNAnexus) to achieve high efficiency speed. We benchmarked the method using chromosome 20 data extracted from 3,751 whole genomes. With the 3-out-of-4 consensus strategy, we identified 1.37 million SNPs with a transition-transversion ratio of 2.83. Among them, 40.2% are novel compared to 1000Genomes phase3, and 91.2% of novel SNPs have MAF<0.1%. Compared to CHARGE WES from the same individuals, the WGS call set achieved a sensitivity of 61.1%, specificity of 99.8%, and 95.4% of the missed SNPs are singletons and doubletons. The SNP rediscovery rate in exome region is 38.3% for singletons and 62.2% for doubletons, which is higher than in 1000G phase1. We also identified 8,141 unique STR loci on chromosome 20, with a rediscovery rate of 97.8% compared to 1000G phase1 results. The entire process, including genotype/haplotype inference and WES integration, requires 3000 CPU-month for the Chromosome20 pilot. Deploying the pipeline in distributed computing environments (i.e. the Cloud) holds promise to make the ensemble method approachable to the community.

1594M

Likelihood-based filtering of indels and structural variants by leveraging Mendelian inheritance and transfer learning. *H. Kang[1], D. Absher[2], A. Tan[1], C. Quick[1], A. Locke[1], Z. Chen[1], G. Jun[1], J. Vincent[3], C. Pato[4], M. McInnis[5,6], S. Zoellner[5], G. Breen[7], S. Levy[2], R. Myers[2], W. Iacono[8], M. McGue[8], L. Scott[1], S. Vrieze[1], M. Boehnke[1], G. Abecasis[1], The BRIDGES Consortium.* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Ann Arbor, MI; 2) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 3) Center for Addiction and Mental Health, Department of Psychiatry, University of Toronto, ON, Canada; 4) Department of Psychiatry and Behavioral Sciences, University of California, Los Angeles, CA; 5) Department of Psychiatry, University of Michigan, Ann Arbor, MI; 6) University of Michigan Depression Center, Ann Arbor, MI; 7) Medical Research Council Social, Genetics and Developmental Psychiatry Center, Institute for Psychiatry, King's College London, UK; 8) Psychology Dept, University of Minnesota, Minneapolis, MN.

Accurate detection and filtering of short insertions and deletions (indels) and structural variants (SVs) is very challenging. Unlike SNPs, which can be validated using array-based genotyping, indels and SVs have fewer reliable resources to evaluate quality. Assessment of the pattern of genotype calls between members of the same family can provide an alternative way to evaluate variant quality, as the vast majority of variants should follow Mendelian inheritance (MI) patterns. However, even at truly variant sites, genotype calls contain errors and uncertainties due to variable read depth or quality, limiting the ability to distinguish between true and false variants. We develop a variant filtering method (MiFT) by quantifying the statistical support for a MI pattern of each variant in family sequence data. First, we compare the likelihood of sequence reads under models with and without family structure to obtain a Bayes' Factor that quantifies the degree of support for MI. Second, we train a support vector machine (SVM) classifier to differentiate between variants with very high or low Bayes' Factors, using the variant features annotated by each caller. Third, we can transfer the SVM classifier to predict the quality of variants produced from an independent sequencing dataset without related individuals. We used MiFT to filter SNP, indel, and SV calls produced by GotCloud, UnifiedGenotyper (UG), HaplotypeCaller (HC), GenomeSTRiP, Pindel, and CBS on whole genome sequences of 281 Minnesota Twins samples over 84 families. We then applied the prediction model to 2,833 unrelated subjects sequenced in BRIDGES. We observed that existing SNP filtering methods are highly concordant to results from MiFT (e.g. 97% for UG). For indels, concordance of MiFT prediction with standard filtering was substantially lower (e.g. 78% for UG). When the filters disagree, variants exclusive to MiFT were significantly more concordant to independent validation results from Affymetrix exome arrays (e.g. OR=8.3, p=7.6×10-8 for UG). For structural variants, we observed that GenomeSTRiP effectively filters out Mendelian inconsistent variants, but also filters out 30% of the Mendelian consistent variants preserved by MiFT. MiFT provides a general framework to robustly filter SNPs, Indels, and SVs. MiFT allows us to evaluate variant call sets against a single standard and will facilitate systematic integration of multiple call sets while achieving high sensitivity and specificity.

1595T

Strength in Numbers: Efficiency and Quality Improvements in Clinical Whole Genome Interpretation. *E. Ramos, C. Mead, J. Sardina, A. Khouzam, D. Mote, S. Pond, A. Crawford, S. Ajay, J. Silhavy, S. Chowdhury, T. Hambuch.* Illumina Clinical Services Laboratory, San Diego, CA.

As the cost of whole genome sequencing rapidly decreases and technical limitations are surmounted, one of the most significant remaining challenges of clinical human whole genome sequencing (WGS) is the interpretation of newly discovered variants. Since October 2012, the Illumina Clinical Services Laboratory has sequenced and interpreted the genomes from approximately 500 primarily healthy adults; with interpretation focusing on 1600 genes associated with 1221 of the most commonly tested monogenic conditions. All single nucleotide variants were assessed to determine clinical significance by a team of trained geneticists and genetic counselors, in accordance with the American College of Medical Genetics & Genomics guidelines. This process includes a manual review of the literature for all variants detected within that subset of 1600 genes. As the Illumina Variant Database has grown to greater than 60,000 variants, including more than 650 variants with potential clinical significance, the average number of variants per genome that are novel to our laboratory has decayed significantly. Currently, of the approximately 5000 single nucleotide variants detected within the subset of 1600 interpreted genes, an average of 65 novel variants is found per genome. This is a six-fold decrease from the 360 novel variants identified per genome in May 2013, resulting in a substantial concomitant reduction in manual effort. This reduction has been accelerated through custom software developed to facilitate the clinical process of classifying, interpreting and reporting on variant disease relationships. These two factors have contributed to an overall decrease in time required for a single WGS sample, from 35 hours in May 2013 to less than ten hours today. Continued growth in the total number of variants within the database, as well as improvements to the software that aids in the annotation and interpretation process, is anticipated to result in a continued diminution of interpretation effort, which will improve the quality of ongoing interpretation and mitigate the cost of offering WGS in the clinic. Although challenges involving the clinical interpretation of whole genome sequence data will persist, these advances are anticipated to facilitate the implementation of WGS in the clinical laboratory testing arena.

1596S

Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *E. Smith[1], A.R. Carson[1], T. Solomon[1], H. Matsui[1], S.K. Braekkan[2,3], K. Jepsen[1], J.B. Hansen[2,3], K.A. Frazer[1,4,5].* 1) Department of Pediatrics and Rady Children's Hospital, University of CA, San Diego, 92093, CA; 2) Department of Clinical Medicine, Hematological Research Group, University of Tromsø, Tromsø, Norway; 3) Division of Internal Medicine, University Hospital of North Norway, Tromsø, Norway; 4) Clinical and Translational Research Institute, University of California, San Diego, USA; 5) Department of Clinical Medicine, University of Tromsø, Tromsø, Norway.

Genotypes generated in next generation sequencing studies contain errors which can significantly impact the power to detect signals in common and rare variant association tests. These genotyping errors are not explicitly filtered by the standard GATK Variant Quality Score Recalibration (VQSR) tool and thus remain a source of errors in whole exome sequencing (WES) projects that follow GATK's recommended best practices. Therefore, additional data filtering methods are required to effectively remove these errors before performing association analyses with complex phenotypes. Here we empirically derive thresholds for genotype and variant filters that, when used in conjunction with the VQSR tool, achieve higher data quality than when using VQSR alone. The detailed filtering strategies improve the concordance of sequenced genotypes with array genotypes from 99.33% to 99.77%; improve the percent of discordant genotypes removed from 10.5% to 69.5%; and improve the Ti/Tv ratio from 2.63 to 2.75. We also demonstrate that managing batch effects by separating samples based on different target capture and sequencing chemistry protocols results in a final data set containing 40.9% more high-quality variants. In addition, imputation is an important component of WES studies and is used to estimate common variant genotypes to generate additional markers for association analyses. As such, we demonstrate filtering methods for imputed data that improve genotype concordance from 79.3% to 99.8% while removing 99.5% of discordant genotypes. The described filtering methods are advantageous for large population-based WES studies designed to identify common and rare variation associated with complex diseases. Compared to data processed through standard practices, these strategies result in substantially higher quality data for common and rare association analyses.

1597M

Computational validation of NGS variant calls using genotype data. *M.A. Taub[1], S. Shringarpure[2], R.A. Mathias[3], R.D. Hernandez[4], T.D. O'Connor[5], Z.A. Szpiech[4], R. Torres[4], F.M. De La Vega[2], C.D. Bustamante[2], K.C. Barnes[3].* 1) Biostatistics, Johns Hopkins, Baltimore, MD; 2) Genetics, Stanford University, Palo Alto, CA; 3) Medicine, Johns Hopkins, Baltimore, MD; 4) Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA; 5) Medicine, University of Maryland, College Park, MD.

Variant calling from next-generation sequencing (NGS) data is susceptible to false positive calls due to sequencing, mapping and other errors. We present a method that uses machine learning techniques, specifically Random Forests, for computationally validating variant calls obtained from a sample of individuals. While existing methods use site quality information from known samples such as HapMap and dbSNP for training classifiers to distinguish between true and false variant calls, our method uses genotype data from the same samples to learn a more accurate classifier. We demonstrate our method on a set of variant calls obtained from 643 high-coverage African-American genomes from the The Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA), sequenced to high depth (~30X). On variant calls obtained using Illumina's single-sample caller CASAVA, our method has a True Positive Rate of 97.5% (at a False Positive Rate of 5%). On variant calls obtained from Real Time Genomics' multisample variant caller, our method obtains a True Positive Rate of 95% (at a False Positive Rate of 5%). Since most NGS sequencing data is accompanied by genotype data for the same samples, our method can be trained on each dataset to provide a more accurate computational validation of site calls than using generic methods. In addition, our method can allow for adjustment based on allele frequency (e.g., allow a different set of criteria to determine quality for rare vs. common variants) and thereby provide insight into sequencing characteristics that most clearly indicate data quality for variants of different frequencies. We have also applied our classifier to compare call sets generated with different calling methods, including both single-sample and multi-sample callers, and we have found that allele frequency is an important determinant of which calling method makes the most accurate calls.

1598T

Blood vs Saliva: Analysis of the Effect of Sample Type on Variant Calling Confidence for Human Whole Genome Sequencing. *M. Tayeb[1], A. Mijalkovic[2], M. Kovacevic[2], M. Popovic[2], S. Wernicke[2], C. Dillane[1], A. Del Duca[1], R.M. Iwasiow[1].* 1) DNA Genotek Inc., Ottawa, Canada; 2) Seven Bridges Genomics Inc, Cambridge, Massachusetts.

Saliva, obtained using the Oragene self-collection kit, as a source of genomic DNA has grown in popularity due to participant compliance and the high quality and utility of the collected DNA. Although Oragene/saliva is supported by over 1000 peer-reviewed publications, to date few papers have been published on Whole Genome Sequencing (WGS) from saliva. Here we present a systematic, multi-sample analysis of the effect of sample type (blood vs. saliva) on variant calling confidence and the effect of bacterial DNA in saliva on sequence alignment.

Blood and saliva were collected from seven individuals (two families). The bacterial content in each saliva sample was determined by 16S qPCR. DNA was extracted and sequenced on the Illumina HiSeq2000 using standard protocols from Illumina. Variant calling and filtering was performed on the Seven Bridges Genomics platform for bioinformatics analysis using a standard BWA and GATK pipeline in accordance with the Broad Institute's best-practices guidelines.

The percentage of bacterial DNA was found to closely correlate with the number of reads that could not be mapped to the human genome. Those reads were aligned to bacterial and viral genomes obtained from the Human Microbiome Project (HMP). A significant portion of unaligned reads mapped to the HMP reference (72% in blood, 32% in saliva).

No significant difference in the total number of variants called in blood and saliva were observed. Only a slight difference in concordance between sample types was observed (<0.15% for SNPs and <1% for INDEL). Downsampling blood samples to equal coverage virtually eliminated these differences suggesting that the difference in concordance was primarily due to coverage effects. Additional analysis was performed to check for existence of ambiguous regions that act as baits for bacterial reads causing saliva samples to accumulate false positive mutations in those regions.

The overall coverage loss due to bacterial content in saliva is small (approx 3% for every 5% bacterial DNA in the sample) and does not result in a difference in the number of variants called. Concordance differences between blood and saliva are minor and can be eliminated by ensuring adequate sequencing coverage.

1599S

High resolution HLA genotyping software for exome and whole genome sequencing data. *K. Kryukov, S. Nakagawa, T. Imanishi.* Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Kanagawa, Japan.

The MHC region on chromosome 6 contains human leukocyte antigen (HLA) genes, which are some of the most polymorphic genes in the human genome. Many HLA alleles are known to be associated with susceptibility to various infectious and autoimmune diseases. Genotyping HLA loci correctly is challenging due to abundant repeats, high sequence similarity among HLA genes, as well as the presence of highly similar pseudogenes. Normally, sequence-based genotyping is done by highly targeted sequencing of specific loci. Here we present a genotyping software that allows predicting HLA genotypes based on the whole exome or the whole genome sequencing data. Our method is based on sequence similarity searches of the raw reads to the database that combines HLA alleles from the IMGT/HLA database, the known MHC haplotype sequences from Sanger Center, and the entire reference human genome. This allows to accurately map reads to specific loci. We then focus on each locus and compare the mapped reads with the known allele sequences to determine the genotype. Our method has accuracy of 90% in 2-digit typing and 74% in 4-digit typing for Class-I and Class-II genes, on Ion Torrent sequence data targeted to specific HLA genes.

1600M

Using haploid human DNA to design and evaluate the HiSeq X data processing strategy. *M.O. Pollard, T.M. Keane, S.A. McCarthy, J.C. Randall, R.M. Durbin.* Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK.

The Illumina HiSeq X is a new sequencing platform that promises a reduction in sequencing cost to below $1000 per whole human genome. The dramatic increase in throughput that this platform will deliver has required changes to chemistry, software, and output. Of interest to dry lab scientists are several key changes in the data produced including an the increase in read length to 150 base pairs and the binning of base quality scores into eight distinct quality bins in order to reduce the quantity of data that is stored.

To take full advantage of this new platform, we have designed a series of systematic tests to evaluate the sequence data produced by the HiSeq X and determine the most appropriate subsequent processing protocols. We will test and compare different combinations of mapper, variant caller, and reference; using well characterised samples in order to examine the effects that this technology has on each and evaluating its utility for variant detection.

In this experiment, we plan to sequence the 1000 Genomes CEU Trio and two human haploid cell lines: CHM1hTERT and HAP1. We will evaluate the effects of various read alignment tools (BWA backtrack, BWA MEM, and SNAP), variant callers (samtools, GATK Unified Genotyper, and GATK Haplotype Caller), and choice of reference (GRCh37 vs GRCh38) on the sensitivity and specificity of SNP/InDel identification. The CEU Trio is a well-studied family, which has been used in many previous studies such as HapMap and 1000 Genomes providing a corpus of knowledge to compare against. Both the CHM1hTERT and HAP1 cell lines are effectively haploid, meaning that any heterozygous variants found are likely to be either a sequencing error or data processing artifact. In addition, when HAP1 is mixed with CHM1hTERT in equal proportions it should allow us to evaluate heterozygous calling and phasing performance. The findings of these experiments will be valuable for determining the optimal data processing strategy for the HiSeq X.

1601T

Genotyping Complex Markers for Drug Absorption, Distribution, Metabolism and Elimination with the Axiom® Genotyping Platform. *J. Gollub, C. Bruckner, M. Gallina, S. Malakshah, M. Mittmann, P. Roghani, A.H. Roter, F. Siddiqui, R. Varma, B. Wong, T. Webster, M.H. Shapero, J.P. Schmidt.* Affymetrix, Inc., Santa Clara, CA.

Many genomic markers (SNPs, indels, and copy number changes) relevant to drug absorption, distribution, metabolism and elimination (ADME) are difficult to genotype due to complex genomic contexts with high homology to other sites, and many nearby polymorphisms. Copy number (CN) changes in relevant genes such as GSTT1, GSTM1, UGT2B17, CYP2D6, and CYP2A6 can be difficult to assess because they are members of large gene families with highly homologous sequence, or because of the high prevalence of copy number states other than 2 in the population. We have developed algorithms for probe design and analysis on the Axiom® Genotyping Platform to assay these challenging regions. Copy number detection is based on signal response of SNP-genotyping probesets; considerations include probe selection for good dose response to changes in copy number state, and accurate summarization of signal values for classification. In the hybridization-based Axiom® assay, genotyping SNPs and indels in the presence of other, nearby polymorphisms requires probe design to account for alternative genomic contexts, as well as algorithmic identification of appropriate probes for each individual. Multiallelic markers pose a challenge as well, requiring the development of genotyping algorithms that allow for more than the usual three possible states per marker. Results indicate accurate copy number calls on the five key genes listed above, with good discrimination between homozygous copy loss (CN 0), hemizygous loss (CN 1), CN 2 and CN 3 where applicable. We likewise observe highly accurate genotype calls on many SNPs and indels in complex regions with multiple flanking variants. The ability to accurately identify pharmacogenomic polymorphisms in a highly multiplexed assay could dramatically impact future drug development processes. Here we show that advanced probe design strategies, and new algorithms, supported by the fundamental flexibility of the Affymetrix Axiom® technology, allow for successful genotyping of problematic ADME variants.

1602S

PHENOVAR: a phenotype-driven approach to facilitate routine utilization of clinical exome sequencing for the diagnosis of polymalformative syndromes. *C. Buote[1,4], F. Thuriot[1], M. Edmont[1], E. Gravel[1], Y. Trakadis[2], JF. Theriault[1,3], H. Larochelle[3], B. Maranda[1], S. Chénier[1], PE. Jacques[4], S. Lévesque[1].* 1) Dept. of Paediatrics, division of medical genetics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, Canada; 2) Dept. of Medical Genetics, McGill University Health Center, Montréal, Canada; 3) Dept. of Informatics, Faculty of Science, Université de Sherbrooke, Sherbrooke, Canada; 4) Dept. of Biology, Faculty of Science, Université de Sherbrooke, Sherbrooke, Canada.

BACKGROUND: Polymalformation syndromes consist in a large group of heterogeneous genetic disorders, for which our ability to identify the causative gene using conventional investigations remains limited. Exome sequencing offers a solution and is now available either on a research basis or in few clinical laboratories across the US. Routine utilization of exome sequencing is still hindered by our capacity to predict effectively the causative change(s) out of several thousands of variants. To facilitate exome analysis and accelerate implementation of exome sequencing in clinical practice, we have developed and recently published a software, named PhenoVar. This software integrates the patient's phenotype to the genotype data and suggests to the physician-user a short list of prioritized potential diagnoses for review. Here, we present the preliminary results of PhenoVar validation in patients affected with an undetermined polymalformation syndrome, in comparison to standard bioinformatics analysis. METHODS: A total of 25 patients with polymalformative syndromes of likely genetic etiology were accepted for exome sequencing after genetic counselling. Each patient has a normal array CGH and remains without a clear diagnosis after Sanger sequencing-based gene tests. To date, we completed the sequencing of 14 patients. A medical geneticist performed the analysis on these patients using PhenoVar, in parallel of the standard analysis done by the bioinformatics team. RESULTS: On average, PhenoVar reduced the number of potential diagnoses for manual review to 20 per patient in comparison to 63, for standard bioinformatics analysis. In both cases, we obtained a diagnostic yield of nearly 50 % (6 /14). Each time, the correct diagnosis was found in the top ten diagnoses of the PhenoVar's list. In patients who provided consent, we searched for variants in cancer genes, which led to the identification of a pathological variants in *BRCA2* in one of them. Remaining patients' exomes are being sequenced and analyzed, and will be presented. CONCLUSION: Our preliminary results suggests that exome sequencing combined with PhenoVar, using a phenotype-driven approach, led to a similar diagnostic yield than standard bioinformatics analysis, reduced the human resources and time required for bioinformatics analysis. Since it can be used directly by medical geneticists, this software could facilitate routine utilization of exome sequencing in clinical practice.

1603M

Single molecule reconstruction and variant detection of less than 1 genome in 1000. *K.R. Covington, M. Wang, D.M. Muzny, H. Doddapaneni, R.A. Gibbs, D.A. Wheeler.* Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX., USA.

Genetic mosaicism contributes to phenotypic diversity both in physiology and disease including phenotypic diversity in monozygotic twins, brain development and disease, and cancer. Conventional sequencing approaches to variant discovery rely on repeated sampling of variants, from the population of genomes in a given sample. Standard sequencing methods fail to address rare, subclonal mutations due to confounding influence of sequencing errors, even at high sequence depth. We address this problem by adapting published proof-of-concept methodologies for deep sequencing using single molecule reconstruction. Single molecule reconstruction relies on repeatedly sequencing the PCR products of single molecules of DNA and using the sequence data to reconstruct the sequence of the initial molecule. With this strategy we turn PCR duplicates into an advantage to repeatedly sample fragments which can be traced to a single DNA molecule. We used a combination of 6 to 24 Illumina barcode adapters pooled to individually tag sheared DNA libraries. DNA libraries were then captured using custom cancer capture designs. Libraries were sequenced using either Illumina MiSeq (for initial testing) or HiSeq platforms. Sequencing data was aligned to the human genome hg19 build. We developed custom software to reassemble sequencing reads into their original single molecule DNA fragments and for further analysis. Over 90% of non-reference bases sequenced were observed only in one read of a duplex set, indicating that these non-reference bases are merely sequencing errors. Our ability to recreate the initial DNA molecule allowed us to detect variant genomes at less than 1:1000 variant allele fraction in a series of spike-in tests. Interestingly, over 50% of consensus variants were called in only one molecule, even in normal tissues. These data suggest the presence of an extensive reservoir of somatic variation in normal tissues, which may contribute to human disease.

1604T

Detection of common and low frequency variants in cancer samples with SNPPET using NGS target enrichment data. *A. LNU, D. Joshi, C. Lecocq, J. Ghosh.* Agilent Technologies, Santa Clara, CA.

Identification of variants from cancer samples using NGS target enrichment is becoming increasingly popular in clinical research. Cancer samples are challenging for variant detection due to tumor heterogeneity, low abundance of variant bases, mixed genotypes, multi-allelic loci, copy-number variations, variable background mutation rates and contamination from normal cells. After evaluating available methods it was imperative to develop an algorithm in-house. SNPPET was developed in Agilent SureCall software to detect cancer variants with high confidence. SNPPET has two basic steps - first an efficient search for variants and then a careful examination of the variant's neighborhood to confirm the call. In the first step, the reads and bases are filtered for quality and coverage criteria such as mapping qualities, base qualities and read depths. SNNPET's initial model evaluates each locus and all non-reference alleles are assigned to sequencing error. Then a second model considers each non-reference allele to be a true variant. The models allow presence of more than one allele at each locus. A log-odds score is calculated between the two models and if the score passes a specified threshold, it is considered as a variant indicating the allele is more likely to be a variant than a sequencing error. If multi-allelic option is chosen, all potentially variant alleles are retained. Otherwise only the one with maximum allele fraction is retained. In the second step, a local re-assembly is done to evaluate all potentials variant combinations as haplotypes. The haplotypes are scored by a Bayesian model and all variants passing the threshold are retained. SNPPET allows one to examine loci of interest irrespective of whether a "true" variant is present or not. This allows one to ascertain whether there is low significance event plausible at a locus which did not pass the statistical tests, but still may be indicative of potential variation in a site of interest. Dilution series experiments were performed to benchmark the sensitivity and specificity of SNPPET and to compare the variant calls with other state of the art methods. A cancer cell line(HCT116) was mixed with normal HapMap(NA10831) at various concentrations. At concentration of <= 5%; SNPPET detected expected variants that are missed by GATK unified genotyper and SAMTools. SNPPET demonstrated greater sensitivity than any other methods compared, keeping the specificity at par with these methods.

1605S

Identification of common non-synonymous SNPs in proteomic datasets and their use to obtain measures of individualization and biogeographic background. *G. Parker[1,2], D. Anex[2], T. Leppert[3], L. Baird[3], N. Matsunami[3], M. Leppert[3], B. Hart[2].* 1) Utah Valley University, Dept of Biology, Orem, UT; 2) Forensic Science Center, Lawrence Livermore National Laboratory, Livermore, CA; 3) Department of Human Genetics, University of Utah, Salt Lake City, UT.

We have developed methodology to extract identifying genetic information from proteomic datasets. DNA-typing has revolutionized forensic practice and jurisprudence, however DNA often is degraded due to biological, chemical or environmental factors. Protein is considerably more stable, more abundant than DNA, and persists in the environment for a longer period. Protein also contains genetic information in its primary structure, the result of non-synonymous SNPs (nsSNPs). These single amino-acid polymorphisms are accessible to shotgun tandem mass spectrometry. We have identified nsSNP-containing peptides from 35 alleles in 26 genes expressed in the forensically informative hair shaft proteome. We obtained complex proteomic datasets from trypsin digests of the hair shafts of 54 validated European American individuals. Peptides corresponding to nsSNPs expressed in this protein population were identified and collated for each individual. The combined probability of each individual nsSNP profile was calculated using genotypic frequencies of each allelic combination in the European population (1000 Genomes Project) and the "product-rule". The power of genetic discrimination ranged from 1 in 1.002 to 1 in 9,000. The average power of discrimination was 1 in 280. The power of discrimination increased as a function of proteomic dataset quality ($r^2 = 0.624$, $n = 58$, $p < 0.0001$). When the power of discrimination is calculated using genotypic frequencies from the African population, increased powers of discrimination are achieved. This is consistent with an increased likelihood that the samples originate from a European relative to African origin. Likelihood measurements range from 1 to 780 with an average of 50, a median of 18, and a standard deviation of 116. ($n = 64$). Direct validation of the imputed status of each nsSNP allele was achieved with Sanger sequencing. A total of 430 genotype determinations were made from the proteomic data and 426 assignments were confirmed (specificity = 99.1%, FPR = 0.93%). The average sensitivity was 31%. We have established a framework for the use of proteomic datasets as a source of identifying genetic information, allowing measures of identity and biogeographic background to be made from forensic or anthropological protein. This study also demonstrates that protections should be enacted to ensure the privacy and confidentiality of human subjects when providing tissue for proteomic studies.

1606M

SNP and CNV Detection in Trisomy 21 Individuals Using a First-Principles Approach. *Y.A. Jakubek, D.J. Cutler.* Human Genetics, Emory University, Atlanta,GA.

Genotyping arrays are used to simultaneously detect single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) across the genome. Current analysis methods use empirical data to call SNPs; therefore, they are not well suited for studies with a small number of samples and for SNPs with a low minor allele frequency. These methods drop 20% to 33% of the targeted SNPs and are susceptible to batch effects. We have developed a SNP and CNV detection algorithm for Affy 6.0 arrays that is based on a low-level model of hybridization on the array. In it we model binding between all probe-DNA duplexes that form on the array. Furthermore, we explicitly model batch effects through a set of parameters that describe binding for each array experiment: DNA concentration, probe errors, salt concentration, wash stringency, and scanner settings. This biochemical approach allows for the individual analysis of each chip. Since we directly model DNA concentration, our algorithm can type SNPs and CNVs on chromosomes of any ploidy. Our method assigns a posterior probability to each SNP; we refer to this value as the quality score (QS) for the call. We analyzed data from Down syndrome and normal samples. 13% of targeted SNPs show significant cross-hybridization. 84% of SNPs on diploid chromosomes and 57% of SNPs on trisomic chromosome 21 had QS > 0.99. Our algorithm directly estimates DNA concentration for each probe and then uses these values to detect CNVs. This approach is fundamentally different from current approaches that rely on probe intensity. We called an average of 50 CNVs per individual and 68% of the CNVs called were in the database of genomic variants (DGV). This data was previously analyzed; validation was attempted for 64 CNVs of which 59 were validated. Our method only called the 59 validated CNVs. This method can be easily adapted to call SNPs within CNVs and for aneuploid chromosomes in general. Using only first-principles our method detects genetic variants with high accuracy.

1607T

Variant Detection and Validation in RNA-Seq Data. *F. Schlesinger, I. Khrebtukova, L.C. Watson, S.M. Gross, S. Pathak, T. Singer, T. Hill, G.P. Schroth, R. Kelley.* Illumina Inc San Diego, CA.

Shotgun sequencing of cDNA (RNA-seq) has become a widely used tool to measure gene expression. One benefit of RNA-seq is that, due to the full sequencing read-out, it allows the detection of genomic sequence variation at the same time as expression levels; ranging from SNVs and small insertions or deletions, to large scale structural events, such as gene fusions. This can be used to create 'exome-like' variant call sets quickly and cheaply, without the need for sequence enrichment, especially in non-model organisms. Perhaps more importantly, it also allows researchers to confirm and extend results from matched DNA sequencing data and support variant interpretation. Especially in the context of cancer samples such collection of matched DNA and RNA is becoming common, for example by The Cancer Genome Atlas (TCGA) project. However, while a lot of progress has been made on variant calling software for DNA sequencing, these tools typically do not work well with RNA-seq data, and specialized RNA callers are still limited in reliability, accuracy and run-time. In addition systematic evaluation of accuracy on different RNA-seq datasets, as well as best practice guidelines are missing. Here we present a new analysis pipeline for RNA-seq that integrates gene expression analysis with high-accuracy variant calling, based on one single-pass alignment process, greatly reducing computational cost. We use a comprehensive set of known variants (including SNVs and Indels), validated by analysis of inheritance patterns across a large pedigree, for a systematic, unbiased evaluation of RNA variant calls. On this data we apply machine-learning techniques to train an RNA-specific variant scoring model and measure sensitivity and specificity as a function of expression levels, including the impact of allele-specific expression on sensitivity. We assess performance across a range of RNA-seq data-types, including different sample preparations techniques and sequencing depth, showing false positive rates ranging from 1% to 5%. Integrating matched DNA and RNA data, we demonstrate how RNA-seq can be used to improve accuracy of DNA variant call sets by rescuing up to 5% of marginal calls. We derive allele-specific expression and between-exon phasing information to guide variant interpretation.

1608S

An Evaluation of Splice Prediction Software Accuracy Using in vivo Data from Patients with Osteogenesis Imperfecta. *J. Schleit[1], S.S. Bailey[1], T.T. Tran[1], D. Chen[1], S. Stowers[1], U. Schwarze[1], P.H. Byers[1,2].* 1) Pathology, University of Washington, Seattle, WA; 2) Medical Genetics, University of Washington, Seattle, WA.

Approximately 20% of pathogenic mutations in COL1A1, which encodes the proα(I) chain of type I collagen, alter mRNA splicing. The Osteogenesis Imperfecta (OI) phenotypes that result depend on the sequence and stability of the resulting mRNA(s). Stable mRNA transcripts produce more severe OI, while unstable mRNAs are associated with milder OI. When molecular diagnostics relies on DNA sequence analysis alone, the need to predict splice outcomes has increased. Many diagnostic labs use prediction software to interpret splice site mutations. To test the accuracy of these tools we sequenced cDNAs from patient-derived fibroblasts, treated with cycloheximide to stabilize mRNAs, from 38 individuals with OI. Our sample set included 15 samples from individuals with OI type I, 4 with OI type IV, 7 with OI type III individuals, and 12 with OI type II. We found that IVS[X]-1G>A changes always resulted in frameshifts when the first nucleotide of the adjacent exon was G, that other mutations in consensus sites were not as predictable, and that there were often multiple outcomes. We compared our measured findings to outputs from three splice prediction software: Human Splice Finder (HSF), Spliceport (SP), and Automated Exon Definition and Splice Site Analysis (ASSEDA). All three programs predicted that the sequence changes examined were pathogenic with greater than 80% accuracy. However, the programs differed in their ability to predict the actual splice outcome (HSF correctly predicted splice outcome in 42% of cases, SP in 55%, and ASSEDA in 74%). Determination that a variant is pathogenic is the first step. If splice site mutations in the same gene result in a range of phenotypic severity, a clinical severity score would be a valuable adjunct to outcome prediction. Because for COL1A1 mutations, the relation between splice outcome and phenotype has been well studied, we asked if the incorrectly predicted splice outcomes could, nonetheless, result in appropriate phenotypic predictions. For splice outcomes incorrectly predicted by HSF 68% resulted in phenotype errors, by SP 76% resulted in phenotype errors, and by ASSEDA 60% still resulted in phenotype errors. The type of error differed among programs. At this point, there is no substitute for the measure of splice outcome in many instances and care must be taken in the use of splice prediction software for clinical purposes.

1609M

Higher power and efficiency of whole genome sequencing over whole exome sequencing to detect SNVs in exonic regions. *A. Belkadi[1], A. Bolze[2], J.L. Casanova[1,2,3], L. Abel[1,2].* 1) Laboratory of Human Genetics of Infectious Disease, INSERM U1163, University Paris Descartes. Imagine Institute, Paris, France; 2) St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, the Rockefeller University, New York, NY, USA; 3) Howard Hughes Medical Institute, New York, NY, USA.

Whole exome sequencing (WES) is now a standard method to detect rare and common genetic variants. Whole genome sequencing (WGS) is becoming an attractive alternative approach with an affordable price. Here we compared WES and WGS in the detection of single nucleotide variants (SNVs) in the coding regions of the human genome using the most recent technologies. We performed WES (using the SureSelect Human All Exon Kit 71 Mb) and WGS (using the TruSeq DNA PCR-Free Sample Preparation Kit) in 7 European individuals. Both platforms sequence paired-end reads of ~100 base pairs. A total of 132,108 SNVs were detected by either WES or WGS within the regions targeted by the WES kit, 86.7% by both methods, 4.5% by WES, only and 8.8% by WGS only. While these variants were present in a total number of reads which were similar in WES (~5.8 millions) and in WGS (~4.7 millions), the distribution of the coverage depth was strikingly different between the two methods. In WES data, the distribution of depth was lognormal-like with a mode at 5x and a median at 26x, while the distribution was normal-like in WGS data with a mode/median at 38x, showing a much more homogeneous coverage by WGS. Among the 114,586 SNVs coded by both platforms, we noted discordant genotypes in 3,676 variants (3.2%) that were in all cases called as homozygous by WES and as heterozygous by WGS. We then filtered the variants on depth (> 8x), genotype quality (>20), and minor allele ratio in heterozygous genotypes (≥20%). Only 69.4% of WES variants passed the filter, while 97.1% of WGS variants did. After filtering, a total of 82,331 SNVs were called by both platforms (99.97% of those variants had the same genotype), 3,857 variants by WES only and 41,489 by WGS only. Refined analysis of variants exclusive to WES by the Integral Genomic Viewer tool showed that these variants are also present in WGS data but not called by Unified Genotyper. The utilization of an alternative/additional variant caller on WGS should call these variants. In contrast, variants exclusive to WGS were not present in WES data. These variants mapped to coding regions, UTRs, splice-sites and non-coding RNAs. We conclude that WGS is more powerful and more efficient than WES to detect SNVs even in exonic regions, as it calls about ~50% more high quality SNVs than WES. This work will have major impact to define the most cost-effective strategy to identify the genetic etiology of rare and common diseases.

1610T

Evaluation of the Illumina NextSeq500 for Rapid Whole Genome Sequencing. *S. Dames[1], J. Durtschi[1], R. Mao[1, 2], K.V. Voelkerding[1, 2].* 1) ARUP Labs, Salt Lake City, UT; 2) University of Utah Department of Pathology.

Introduction: Rapid sequencing of the whole genomes offers a new diagnostic tool in clinical settings, including newborns. A new next generation sequencing platform, the Illumina NextSeq500, generates 120 Gb of sequence in 29 hours using 2 X 150 base length reads. This represents 30-fold average coverage of the human genome. The current study evaluates the quality of whole genome sequencing data generated by the NextSeq500. **Methods**: HapMap sample NA12878 whole genome Illumina libraries were prepared using a Beckman Coulter SPRI-TE. The ARUP validated clinical pipeline comprised of BWA, SAMtools, GATK, and Annovar was employed. NextSeq500 NA12878 whole genome data was compared to NA12878 NIST and NA12878 exome data. For the gene structure variation and loss/gain analysis, alternate percentage and reference allele percentage reads per variant were plotted against chromosomal location to visualize CNVs that alter the expected allele read frequency. **Results**: Two independent runs of NA12878 yielded 68 Gb at 87.9 percent >Q30 and 134.5 Gb at 70.4 percent >Q30. Whole genome data filtered by exome bed file coordinates was intersected with NIST NA12878 data. 66,975 variants were detected in whole genome data, with 66,492 >4-fold (99.3 percent) and 65,232 >8-fold coverage (98.1 percent), average 21-fold coverage. The NIST high confidence data contains 55,176 variants in these regions, where 53,958 (97.8 percent) were concordant between whole genome data and NIST data, and 49,526 (89.8 percent) were concordant between the exome data and NIST data. The copy number gain/loss results were concordant with high density SNP array. **Conclusions**: Preliminary analysis demonstrates an approximate 98 percent concordance between variant calls derived from NextSeq500 whole genome sequencing of NA12878 and high confidence NIST NA12878 variant data More sophisticated analysis comparing different alignment and variant calling algorithms to improve alignment and variant call accuracy and speeds will be presented. Further analysis is underway to determine the cause of discordants between NA12878 NIST and whole genome data. Improvements in the speed of analysis and quality metrics will improve turn around time, and may preclude the requirement of Sanger verification for a subset of variants.

1611S

Benchmarking of DNA short read aligners on GCAT data sets. *R. Gupta, A. Kalbhor, P. Potla, S. Katragadda, V. Veeramachaneni, R. Hariharan.* Strand Life Sciences, Bangalore, Karnataka, India.

Background and Objectives: Next generation sequencing technology has led to the generation of millions of short reads at an affordable cost. Aligning these short reads to a reference genome is a crucial task for most of the downstream analysis. However aligning the short reads to a reference genome is a non-trivial task because of the large size of the data and many complex regions in the genome. These challenges necessitate the need for sophisticated algorithms that are both accurate and computationally efficient. In this work, we will briefly discuss Strand NGS (formerly Avadis NGS) alignment algorithm and present the benchmarking results on several simulated data sets and a real whole-genome data to compare it with other state-of-the-art algorithms. **Results:** Multiple aligners like Strand NGS, BWA, BWA-Mem, Bowtie2 and Novoalign3 are compared for accuracy and computational efficiency using 4 simulated data sets from the GCAT website and a real Illumina HiSeq 2500 whole-genome paired-end data of 1000 genomes CEU female sample, NA12878. Strand NGS and Novoalign3 showed comparable accuracy in terms of '% correctly mapped' reads and receiver operating curves (ROC). They also seem to outperform other algorithms especially on data sets with longer InDels. Our aligner, Strand NGS has been designed to detect longer indels by excluding the longest gap in a read match while computing the percent identity (alignment score) for the match. This makes it possible to give a higher cut-off for the percentage of gaps allowed, without reducing the percent match cut-off. For reads potentially originating from complex genomic locations like repeat regions, Strand NGS aligner produces a higher true positive rate compared to Novoalign3, when aligned reads are filtered based on the associated assigned mapping qualities. As for the performance comparison based on computational efficiency, other than minor differences, practically all the included algorithms showed comparable performance. **Conclusions:** Alignment of millions of short reads to a large reference genome with many complex regions is still a hard problem and almost all current algorithms adopt some form of strategy to trade-off accuracy and computational efficiency. The benchmarking results presented in this study suggest that Strand NGS aligner is a powerful approach for DNA short read alignment and either compares well or even outperforms other state-of-the-art algorithms.

1612M

A comparison of commonly used alignment algorithms using 15 whole genome sequences. *B.D. Pickett[1], M.E. Wadsworth[1], K.L. Hoyt[1], J.D. Duce[1], J.S.K. Kauwe[1], M.J. Clement[2], P.G. Ridge[1].* 1) Department of Biology, Brigham Young University, Provo, UT; 2) Computer Science Department, Brigham Young University, Provo, UT.

Due to the large amount of relatively small reads produced from next-generation sequencing (NGS), assembling a genome is a time-consuming and complex task. Aligning reads accurately and in a timely manner is critical for successful research. Many algorithms exist to align NGS reads and each claims to be the most accurate and has the data to "prove" it in publications describing the software. However, these comparisons are flawed for several reasons: (a) The developers of novel software packages most often are the researchers performing the comparison and report only the results demonstrating that their software is superior to existing software; however, these authors have an obvious conflict-of-interest, and indeed, their comparisons are frequently not repeatable by researchers. (b) Many of these comparisons have been performed using either very small data sets or simulated data that fails to capture the complexity of real biological data. (c) Each comparison uses different data and metrics to assess software performance. Until now, no comparison has been conducted to compare these software packages that overcomes these three challenges. We perform an unbiased comparison of various aligners for NGS to (a) identify overall effectiveness, strengths, and weaknesses of each software package and (b) make the data available as a standard efficacy test for newly created software. Our analysis provides other researchers with evidence suggesting which algorithm or series of algorithms will be best for their research with unique combinations of variations in their data. Our tests are run on 15 Illumina-sequenced whole genomes. Each genome has an average read depth of 37. For our research we test the following algorithms: Bowtie2, BWA, GSNAP, Mosaik, Novoalign, Soap2, and Stampy. Each of these is free for academic use and is currently maintained. Ultimately, this will lead to more accurate experimental and analytical results.

1613T

A comprehensive comparison of RNA-seq and microarray in transcriptome profiling of rat livers exposed to a broad range of agents. *C. Wang[1], B. Gong[2], P.R. Bushel[3], J. Thierry-Mieg[4], D. Thierry-Mieg[4], J. Xu[2], H. Fang[2], H. Hong[2], L.J. Lancashire[5], C. Furlanello[6], L. Shi[7], R.S. Paules[3], S. Auerbach[3], W. Tong[2].* 1) Center for Genomics, Loma Linda University, Loma Linda, CA; 2) National Center for Toxicological Research, Jefferson, AR; 3) National Institute of Environmental Health Sciences, RTP, NC; 4) NCBI, NLM, National Institutes of Health, Bethesda, MD; 5) Thomson Reuters, London, UK; 6) Fondazione Bruno Kessler, Trento, Italy; 7) Fudan University, Shanghai, China.

Next-generation sequencing technologies have revolutionized the genomic research and allow the genome and transcriptome of any organism to be explored without a priori assumption. Compared to microarrays, RNA-seq is able to provide single-nucleotide resolution, strand specificity, and short-range connectivity through paired-end sequencing. However, emerging transcriptomics technologies should be evaluated by ways in which they create opportunities to advance the understanding of complex biological systems. An extensive and systemic investigation based on a comprehensive study design was performed to investigate the strengths/weaknesses and comparability of the two platforms in biologic elucidation. RNA extracted from livers of rats exposed to 27 agents eliciting a broad range of transcriptional response, comprised of seven modes of action (MOAs) with varying degree of biological complexity, was used to profile gene expression using Illumina RNA-seq and Affymetrix microarrays and to assess concordance. We found that RNA-seq detected more differentially expressed genes (DEGs) than microarray and hence enriched more biological pathways. While RNA-seq was superior in detecting DEGs with low abundance and could also detect differential expression of transcripts, non-coding RNAs, and exon-exon junctions in a treatment-dependent manner, prediction performances of classifiers derived from both platforms were comparable. Overall, the cross-platform concordance in terms of DEGs, enriched pathways, or modes of action is highly correlated with treatment effect size, gene-expression abundance, and the biological complexity of the mode of action.

1614S

Implementing an NGS Bioinformatics Pipeline: Making the Transition from Research to Clinical. *L. Watkins, K. Hetrick, D. Snyder, H. Ling, S. Griffith, J. Goldstein, M. Mawhinney, J. Romm, E. Hsu, G. Lowe, K. Roberts, K. Doheny, B. Craig.* Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality sequencing and genotyping services and statistical genetics consultation to investigators working to discover genes that contribute to disease. The CIDR NGS bioinformatics analysis pipeline, continually developed since 2009 to keep pace with the rapid changes in the predominant sequencing analysis tools, is designed and tuned for large-scale research projects with huge numbers of analyses that must be done in parallel and fault-tolerant fashion. In winter 2013-14 CIDR partnered with a molecular diagnostic lab at our institution to migrate a traditional genetic test panel to NGS, wherein CIDR would run the NGS bioinformatics analysis and they would handle everything else. Despite a limited budget and tight timeline the effort was successful, with the first clinical results reported out in March 2014. In some regards the requirements for a clinical bioinformatics pipeline are similar to that of production research (e.g., the need for reliability and strict quality control) while others can be quite different (e.g., the small numbers of samples, absolute need for quick turn-around time, strict adherence to well-defined regions of interest, an imperative to never report a false positive, the specific and limited relevant annotations), in particular the requirement to lock down all aspects of the pipeline after validation and to document it in detail and track any associated changes in accordance with new CAP guidelines for NGS bioinformatics analysis. This militates against sharing bioinformatics pipelines and infrastructure between research and clinical, so the decision was made to establish completely separate systems. This poster describes the practical aspects and related considerations of the subsequent effort, including getting separate equipment set up and tested/validated, the logistics of remote access and execution and data transfer (with integrity checks) between separate locations, the specifics of the software tools used in the final pipeline and relevant settings, filtering parameters and QC metrics chosen and how they're used, the informatics activities required to meet CAP requirements, the cumulative person-hours required for all these activities and how that effort was distributed, as well as plans for the future. The intent is to convey lessons learned and recommendations for labs pursuing similar efforts.

1615M

Amplicon based 16S ribosomal RNA Sequencing and Species Identification. *J. Dickman, J. Risinger, M. Toloue.* Bioo Scientific Corporation, Austin, TX., USA.

Next generation sequencing analysis of 16S ribosomal RNA is commonly used to identify bacterial species and perform taxonomic studies. Bacterial 16S rRNA genes contain nine hyper-variable regions with considerable sequence diversity among different bacterial species and can be used for species id. Rapid determination of highly complex bacterial populations through targeted amplification can provide an accurate gauge of diversity at taxonomic hierarchies as low as the genus level. A single 16s rRNA hyper-variable domain does not have enough sequence diversity to distinguish contain all bacterial. With increased read lengths of MiSeq chemistry, Bioo Scientific has expanded the common analysis of the fourth hypervariable domain (V4) of prokaryotic 16S rRNA to V1, V2 and V3 regions simultaneously. Optimized preparation through a streamlined standardized procedure allows for high-quality, reproducible libraries. This optimization can be applied to different windows of 16S rRNA as well as other relevant prokaryotic taxonomic markers.

1616T

Super-resolution imaging technique mbPAINT for DNA optical mapping. *J. Chen[1], L. Kisley[2], A. Bremauntz[2], C.F. Landes[2].* 1) Department of Chemistry & Biochemistry, Ohio University, Athens, OH; 2) Department of Chemistry, Rice University, Houston, TX.

Optical mapping has been used to assist the de novo assembly of whole genomes. This fluorescent imaging technique measures the restriction maps of very long DNA molecules that are stretched on a solid support. The resolutions of the traditional techniques are limited by the diffraction limit of light. Here I am reporting a new optical mapping strategy that allows sub-diffraction limit imaging of genome optical maps using a super-resolution optical nanoscopy based on single-molecule localization method, which has been named motion blur Point Accumulation for Imaging in Nanoscale Topography (mbPAINT). The strategy is demonstrated with model target and probe DNA sequences, and an optical map of a randomly chosen repeated 7-basepair sequence in the phage λ-DNA.

1617S

Greatly improved de novo assemblies of eukaryotic genomes using PacBio long read sequencing. *E. Antoniou[1], M.C. Schatz[1], P. Deshpande[1], J. Gurtowski[1], S.M. Eskipehlivan[1], M. Kramer[1], H. Lee[1], S. Goodwin[1], C. Heiner[2], G. Khitrov[2], W.R. McCombie[1].* 1) Genome Center, Cold Spring Harbor Laboratory, Woodbury, NY; 2) Pacific Biosciences, Menlo Park, CA 94025.

High-throughput short read sequencing has revolutionized re-sequencing applications. However, they have had a more limited impact on generating high-quality de novo assemblies of eukaryotic genomes, principally because the short reads are insufficient for disambiguating complex repeats in the genomes. However, in re-sequencing of human genomes with short reads, structural variations may be missed or be ambiguous because of the need to compare to a reference without the structural variants. A combination of rapid advances in Pacific Biosciences long-read sequencing chemistry and improved assembly methods are converging to make very high quality whole genome assemblies possible. We recently sequenced multiple strains of S. Cerevisiae and S. Pombe. By using the BluePippin (Sage Sciences) platform to produce long fragment libraries and the P5 enzyme / C3 chemistry combination, we obtained over 80x coverage of reads exceeding 10,000 bp, and maximum read lengths extending beyond 35,000bp.We are able to produce near-perfect assemblies of the genomes such that every chromosome was assembled into a single contig or a very small number of contigs. We also made an assembly of the IR64 strain of rice (430Mbases genome) using only Pacific Biosciences long read data. The N50 contig size of this assembly is 600kb, doubling our previous assembly N50 obtained with a mix of Illumina short read and PacBio long reads.

1618M

Naturally Hyperactive Transposase and its Application to High Coverage Whole Genome Sequencing With Low DNA Input. *A. Belyaev[1], C. Hansen[1], J. Fox[1], B. Arezi[1], B. Novak[2], F. Useche[2], C. Pabon[2], H. Hogrefe[1].* 1) Agilent Technologies, Inc., San Diego, CA; 2) Agilent Technologies, Inc., Santa Clara, CA.

Transposases are enzymes that facilitate movement (transposition) of mobile genetic elements (transposons) within and between genomes. High frequency transposition is believed to be detrimental to survival as transposases encoded in prokaryotic and eukaryotic genomes are notably inactive. For example, the best characterized transposase, hyperactive Tn5, required extensive engineering before it could be applied to *in vitro* DNA manipulations, such as cloning, mutagenesis and next generation sequencing library preparation. Based on our knowledge of Tn5, we were surprised to find that a transposase encoded in the genome of bioluminescent marine bacterium *Vibrio harveyi* is at least as active in vitro as hyperactive Tn5 transposase. Unlike Tn5, Vibrio harveyi transposase was expressed in *E.coli* to a high level of about 25% of soluble protein of total cell lysate. We purified the enzyme to homogeneity, proven it to be stable at -20°C for over 3 years and determined that it could be immobilized to a solid support. *Vibrio harveyi* transposase is included in the SureSelectQXT library prep kit. The method of this kit generates consistent fragmentation and uniform sequencing coverage from 50ng of genomic DNA compared to traditional time-consuming adapter-ligation methods that typically require 1-3µg input. In less than 1.5 hours (with only 30 minutes hands-on time) and using an automation-friendly workflow, one can prepare dual-indexed libraries for sequencing on the Illumina HiSeq or MiSeq instruments. Advantages of this transposase as a model for transposition studies and in applications for single-cell analysis, microfluidics and microarrays are discussed.

1619T

Exome Capture with Accelerated Hybridization Provides Same-Day Generation of High Performance Target-Enriched NGS Libraries. *M.C. Borns[1], B. Rogers[1], B. Arezi[1], M. Corioni[2], M. Visitacion[2], C. Pabon[2], D. Roberts[2], H. Hogrefe[1].* 1) Agilent Technologies, Diagnostics and Genomics Group La Jolla, CA; 2) Agilent Technologies, Diagnostics and Genomics Group Santa Clara, CA.

Assays of high sensitivity and minimal turn-around time that generate unambiguous results from minimal sample input are becoming essential. Coupled with the single-molecule resolution afforded by next-generation sequencing, target enrichment provides a cost-effective solution to enable confident analysis of various genomic targets by increasing assay sensitivity. In this study, we enable an accelerated workflow to generate exome-enriched libraries while providing uncompromised variant calling performance. Briefly, fragmented and adapter-tagged DNA libraries are produced in a single step with only 50ng genomic DNA using Agilent's transposase-based library prep technology. DNA libraries are then taken through a novel 1.5-hour fast hyb process, that provides accelerated capture of targets using Agilent's long 120-mer cRNA baits. Combining transposase-based library preparation with this novel in-solution capture technology enables the preparation of up to 16 dual-indexed (post-capture) libraries for sequencing on the Illumina HiSeq or MiSeq platforms in less than 8 hours. Optimized for transposase-generated libraries, SureSelect fast hybridization technology is fully automatable and can be used with bait libraries as small as 0.2Mb up to full exome + UTR libraries, achieving true "exome in a day" results. With this streamlined protocol, we routinely obtain high performance libraries with >60% on-target, >98% coverage at 1x depth, >75% at 20x depth, and <15% duplicates with the Human All Exon V5 and V5+UTR bait sets. Further performance metrics such as library complexity, SNP sensitivity and concordance will be presented for additional bait sets and competitor kits.

1620S

A comprehensive comparison of commercially available hybridization and amplification based exome enrichment methods. *T. Guettouche, K. Xu, L. Dong, L. Tian, C. Kim, H. Hakonarson.* Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

Whole exome sequencing (WES) has become a commonly used method for both research and clinical applications. The most widely used enrichment types are in-solution hybridization and amplification based methods for exome sequencing. WES allows the detection of causative genetic changes and candidate genes in Mendelian and complex disorders as well as germline genetic alterations in cancer. We report a comparison of solution based capture methods (Agilent SureSelect v4 and v5, NimbleGen SeqCap EZ 3.0, Illumina Nextera Rapid and Expanded Capture) as well as an amplification based method (Life Technologies Ion AmpliSeq Exome) and a restriction digest mediated capture method (Agilent Haloplex Exome). The kits included in this study are commonly used and up to date capture methods. All captured samples were sequenced on an Illumina HiSeq2000 except for the Ion Ampliseq exome samples which were sequenced on a Life Technologies Ion Proton system. We used a well characterized family trio from the Center for Applied Genomics Biorepository and the Genome in a bottle (NIST) human HapMap sample NA12878 available from Coriell to compare enrichment and genetic variant metrics of the different methods. We found significant differences in target coverage, on target bases, uniformity of coverage, detected SNVs and indels, and trio discordance. While generally all methods included in this study performed well, our results have important implications and should be considered especially for applications of clinical nature.

1621M

Korean Reference Genome Project: Design and Population Genome Variants. *K. Hong, K.S. Jung, H.Y. Jo, H. Choi, H.J. Ban, M. Chung, S. Cho, B.G. Han.* Center for Genome Science, KNIH, KCDC, Cheongwongun, Chungbuk, South Korea.

The Center for Genome Science (CGS) in Korea National Institute of Health (KNIH) drew up the Korean Reference Genome project (KRG) in 2012. Total 622 DNA samples of study subjects for the KRG were obtained from three different resources. The whole genome sequencing was conducted by Illumina Hiseq2000 sequencer. The sequencing reads trimmed by the Sickle trimming and mapped to human reference genome (hg19). The mapped sequences were analyzed by SAMtools to calling the variants. VCFtools were used to merged the population variant sites, and compute the variants stats. Annovar were used to understand the proportion of the Known variants in the dbSNP138 and 1000G and the variant positions of gene regions. Over 97% of the SNVs identified in each genome were found in dbSNP database or 1000 Genome Project variants. Over 55% of the indels in each genome were found in the 1000 Genome project. The insertion/deletion ratio was 0.75 and approximately 0.1% of the indels were located in the coding region. In this presentation, we will describe the sequencing subjects, raw sequencing stats, and variant calling results. The all variants presented in this paper can be found from our web-browser.

1622T

Resolving the 'Dark Matter' in Human Genomes through Long-Read Sequencing. *J. Korlach.* Pacific Biosciences, Menlo Park, CA.

Second-generation sequencing has brought about tremendous insights into the genetic underpinnings of biology. However, there are many functionally important and medically relevant regions of genomes that are currently difficult or impossible to sequence, resulting in incomplete and fragmented views of genomes. Two main causes are (i) limitations to read DNA of extreme sequence content (GC-rich or AT-rich regions, low complexity sequence contexts) and (ii) insufficient read lengths which leave various forms of structural variation unresolved and result in mapping ambiguities.

Long, unbiased sequence reads from single molecule, real-time (SMRT) sequencing overcome these limitations to provide a more comprehensive and contiguous view of genomes. The sequence data were used for the first long-read *de novo* assembly of a human genome, with over an order of magnitude greater contiguity than previous second-generation sequencing approaches, and a higher level of completeness. Base calls of uniform quality over the broadest DNA GC% and complexity spectrum, the majority of which are contained in reads greater than 11,000 bases, and exceeding 25,000 bases in many cases, provide the necessary performance to resolve many genomic regions previously considered 'difficult' or 'inaccessible.' The long reads also allow long-range phasing of allelic variation. I will highlight numerous examples demonstrating this capability, including the resolution of complex repeats, tandem duplications, insertions, deletions, palindromes, and repeat expansions in the human genome, as well as closing gaps in reference genomes and resolving complex splice isoforms.

1623S

A new method for low-input, PCR-free NGS libraries with exceptional evenness of coverage. *L. Kurihara, J. Laliberte, C. Couture, S. Sandhu, R. Spurbeck, L. Banks, S. Chupreta, V. Makarov.* Swift Biosciences, Inc., Ann Arbor, MI.

PCR-free NGS libraries demonstrate improved evenness of coverage relative to NGS libraries that have undergone PCR amplification. However, most PCR-free methods require 0.5-1 μg of input DNA which restricts such analysis when DNA sample quantity is insufficient. To expand the applicability of PCR-free sequencing, we developed a novel adapter attachment chemistry that is more efficient than existing methods. This new technology results in libraries that capture a higher complexity of each sample, enabling sequence-ready library preparation from 100 ng of input DNA with no PCR required. Another advantage of this technology is that it creates libraries with exceptional evenness of coverage over a broad range of base composition. This was demonstrated by Picard Collect GcBiasMetrics, by improved relative coverage of challenging human sequences such as the 1,000 bad promoters defined by the Broad Institute with 79%; GC content (Ross et al, Genome Biology 2013), as well as sequencing of Plasmodium falciparum (19%; GC) where coverage plots aligned to theoretical distribution. To further broaden the applicability of this method, libraries can be prepared from as little as 10 pg of input DNA by addition of limited PCR amplification. The broad dynamic range of this approach makes it possible to use one kit and one workflow for sample inputs over five orders of magnitude. Libraries prepared using this new method are compatible with both amplicons and hybridization capture and the improved library prep efficiency enables better recovery from FFPE and other samples with limiting input quality.

1624M

Understanding and adopting updates in the human reference genome assembly (GRCh38). *V.A. Schneider[1], T. Graves-Lindsay[2], P. Flicek[3], R. Durbin[4] for the Genome Reference Consortium.* 1) National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD; 2) The Genome Institute at Washington University, St. Louis, MO; 3) European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, UK; 4) The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

The availability of a high quality human reference genome assembly, coupled with advances in sequencing technology, has facilitated basic research leading to new insights into genetic variation and revealed a previously unrecognized level of diversity in the human population. Today, these findings are driving advances in the field of clinical genomics. They have also contributed to the identification of assembly errors and demonstrated that the initial linear representation of the human reference assembly insufficiently represents human genetic variation. Since 2009, the Genome Reference Consortium (GRC), the group responsible for updating the human, mouse and zebrafish reference genome assemblies, has used an assembly model that allows for representation of allelic variation in the human reference genome assembly. Linear chromosome models continue to provide haploid sequence representations, while scaffold sequences known as alternate loci provide alternate sequence representations for genomic regions too complex to represent by a single path or at which there are variants that have sequence not represented on the chromosome. The previous human reference assembly GRCh37 (hg19) contained just 9 alternate loci scaffold at 3 genomic regions. Released in December 2013, GRCh38 (hg38), the current reference assembly, now contains 261 alternate loci scaffolds at 178 distinct genomic regions. These alternate loci provide the only reference assembly representation for more than 3 Mb of sequence, including more than 100 genes. We will present analyses showing how use of alternate loci in alignment target sets can reduce off-target alignments, which may improve variant calling and reveal a need for the development of alignment and variant calling tool chains that make use of the full reference assembly. We will highlight other key improvements in GRCh38, including base updates, sequence additions and the use of single haplotype resources to retile complex regions, which corrected errors and improved annotation. We will describe modeled centromere sequences that have been added to the chromosomes and are expected to serve as mapping targets for α-satellite DNA and benefit studies of human variation, epigenomic regulation and centromere biology. We will present tools and resources developed at NCBI for migrating data to and utilizing GRCh38 and discuss the evolution of the human reference genome assembly as we enter the era of personal genomics.

1625T

The Ion PGM™ Hi-Q™ Sequencing Polymerase: Reducing Systematic Error, Increasing Accuracy, and Improving Read-length. *P.B. Vander Horn, A. Kraltcheva, G. Luo, M. Landes, S. Chen, K. Heinemann, T. Nikiforov, J. Shirley, E. Tozer, D. Mazur.* Ion Torrent, Thermo Fisher Scientific, Life Sciences Solutions Group, Carlsbad, CA., USA.

High accuracy and sequencing uniformity are the hallmarks of a successful DNA sequencing platform. These qualities benefit greatly various scientific and medical research applications. Here we show that the new Ion Torrent™ Hi-Q™ sequencing chemistry provides substantial improvements in overall accuracy, read-length, and systematic error relative to the previous sequencing chemistry. Using the Hi-Q system we see a 50% decrease in overall error, mainly by reducing InDels, and up to a 10-fold reduction in systematic error. In addition, we show that with optimized emulsion PCR cycling conditions, we have increased coverage uniformity of the *Rhodobacter sphaeroides* genome from 88% to 95%. We also observed improvement in uniformity and representation of GC-rich amplicons from various human libraries. By incorporating these improvements, the Ion PGM system is now poised to enable a broader range of research applications, such as enhanced de novo genome assemblies, Human Leukocyte Antigen (HLA) sequencing, bacterial identification, and meta-genomic analysis.

*For Research Use Only. Not for use in diagnostic procedures.

1626S

Third generation sequencing and analysis of complete mitochondrial genomes. *E.P. Hoffman[1], K.B. Gettings[2], K. Kiesler[2], P.M. Vallone[2], L.B. Davenport[1], S. Dadgar[1], K. Panchapakesan[1], S. Knoblach[1], J.M. Devaney[1].* 1) Research Ctr Genetic Medicine, Children's Natl Medical Ctr, Washington, DC; 2) National Institute of Standards and Technology, Gaithersburg, MD.

Next-generation sequencing (NGS) has enhanced investigators' ability to conduct biomedical research examining genetics. Mitochondrial genetics has benefitted from NGS because technology now allows for screening of all 16,569 base pairs of the mitochondrial genome simultaneously for single nucleotide polymorphisms and low-level heteroplasmy. We utilized Single-Molecule Real Time (SMRT) sequencing on a PacBio RS II (Pacific Biosciences) to sequence NIST Standard Reference Material (SRM) 2392 component A (lymphoblastoid cell culture line CHR) and component B (lymphoblastoid cell culture line 9947A), and SRM 2392-I (promyelocytic cell line HL-60). These SRMs are intended to provide quality control when sequencing of human mitochondrial DNA (mtDNA) for forensic identification, medical diagnosis, phylogenetics, and anthropology. We amplified with three overlapping PCR primers sets (6kb PCR products) the entire mtDNA sequence. Then we processed the three mtDNA SRMs using a DNA Template Prep Kit 2.0 (3Kb to 10Kb), Polymerase Binding Kit P4, and DNA Sequencing Kit 2.0. SMRTbell libraries were constructed from pooled, full-length 6kb PCR amplicons tiling the entire 16.6kb mtDNA genome. Each pooled sample was placed on SMRT cell V3 and a 180-minute movie was performed. The data was processed using SMRT Portal with the revised Cambridge Reference Sequence (rCRS). Average read lengths for each SRM were around 3,300 bp with an average coverage of 6,673X. For 2392 component A, we discovered 36 variations from the rCRS including three insertions of C's at positions 309.1, 315.1, and 16193.1. For 2392 component B and 2392-I, we were able to detect 20 variations and 33 variations, respectively, in each SRM from the rCRS including insertions of C's at mtDNA positions 309.1 and 315.1. The number of variants discovered in the three SRMs using the PacBio RS II matched previously published data. We are now investigating methods of analysis to detect heteroplasmy that is occurring in these mtDNA SRMs. This study shows that a third-generation sequencing system can be used to generate entire mtDNA sequences using a three-primer system for amplification. The current NGS approach used in this abstract has the potential to be beneficial in a forensic setting or for medical diagnostics.

1627M

Full-length, single molecule whole transcriptome sequencing reveals alternative 5'- start sites, splicoforms, and poly(A) addition signal sequences. *D.J. Munroe, T. Skelly, C. Raley, R. Stephens, Y. Zhao, S. Gowda, C. Xiongfong, K. Talsania, S. Ratnayake, D. Soppet, X. Wu, B. Tran.* Frederick National Laboratory, Frederick, MD.

Transcriptome analysis has illustrated that the splicing of individual mRNAs can be highly divergent, often resulting in different protein isoforms. Further, mRNA 5' and 3' untranslated regions (UTR), which can be variable, have been demonstrated to play an important role in post-translational gene regulation. The Pacific Biosciences SMRT DNA sequencing platform (PacBio) with its long-read length and single molecule sequencing characteristics is an ideal system for the analysis of these transcript variations. We have utilized this platform to examine divergent splice-forms, 5'-UTRs, and 3'-UTRs in a full-length whole transcriptome analysis of a variety of human samples. As expected, we have observed a plethora of splice-form variations. Somewhat un-expectedly, we also see significant 5'-UTR variation and differential poly(A) addition variation which results in alternative 3'-UTR sequences. Here we report and catalog these variations including alternative splicing, 5' start sites, and poly(A) signal sequences (both AAUAAA and non- AAUAAA). In addition, we have compared our PacBio sequencing results with those observed with the Illumina sequencing platform.

1628T

Preliminary analysis for the evaluation of risk prediction methods using SNP-based genomic profiles data. *R. Arguello[1, 2], F. Hernandez[3], E. Ramirez[1], F. Gonzalez[2], M. Duarte[2], H. Arredondo[1], A. Sanchez[2], Y. García[2], D. Arellano[1], F. Gonzalez[1].* 1) Faculty of Medicine, Autonomous University of Coahuila, Torreon, Mexico; 2) Institute of Science and Genomic Medicine, Torreon, Mexico; 3) School for Biological Sciences, Autonomous University of Coahuila, Torreon, Mexico.

Different methods for predicting genetic risk factors in human diseases have been developed in the last decade. However, many of these techniques have been primarily applied to individuals from Caucasian ancestry. Thus, there is a lack of analysis carried out in populations from other ethnic groups. In this study, we analyzed fifty randomly-selected individuals from the North Area of Mexico that were SNP genotyped using the Illumina array technology with more than 600,000 SNPs. Additionally, 1934 SNPs associated to 139 diseases extracted from the National Human Genome Research Institute GWAS Catalog were used to evaluate existing methods described in the literature for estimating genetic risk factors in human diseases. To do this, we implemented some of these algorithms in an in-house software package. As expected, these methods appear to be less precise when a large number of SNPs is associated to be linked to that particular disease but these can be used as an indicator in the assessment of human diseases by the medical and scientific community.

1629S

Development of a comprehensive, ontology-driven phenotyping system and web-based patient registry for Fanconi anemia. *A.D. Auerbach, S. Balakrishnan, S. Mollah, F.P. Lach, E. Sanborn, A. Smogorzewska, E. Barbour.* The Rockefeller Univ, New York, NY.

The International Fanconi Anemia Registry (IFAR) was established at The Rockefeller University in 1982 to serve as a longitudinal phenotyping system, source of DNA for genotyping studies, and a biorepository. In 1998, a separate public database was created to curate variants in all known FA genes, now in LOVD v.2.0 (www.rockefeller.edu/fanconi/mutate/). Here we describe our newest web-based application whereby clinical and molecular data for all 16 currently identified FA genes will be able to be queried via the internet. The ultimate goal of this new IFAR application is to help move FA research forward by collecting as much genotypic and phenotypic data as possible, displaying these data in a comprehensible, user-friendly manner, and making the data set available to other FA clinicians and researchers. The enhanced IFAR application uses Java for the Application Server programming, chosen because it is fast, secure and reliable. We are using Oracle as the database to store and retrieve related information. The web interface for the client was developed using JSP, HTML and JavaScript, allowing for multiple users to access the system concurrently. The special FA ontology was developed using Protégé for data modeling with the files in OWL format. The “Fanconi Anemia Ontology” is publically available in the NCBO BioPortal (http://bioportal.bioontology.org/ontologies/IFAR?p=summary). This Ontology was created using an OWL file provided by Dr. Ada Hamish and Francois Schiettecatte at the Centers for Mendelian Genetics, with their permission. Novel classes pertaining to FA were added and modifications were made using HPO, OMIM, NCI, and SNOMED. The ontology, while intended for the IFAR specifically, can be easily modified for other disease systems. We are currently migrating all of the clinical and molecular data from the original IFAR database created in FileMaker Pro, to this new ontology-based application. After this is accomplished, access to the de-identified data will be available to others, following rules set by the Rockefeller University Hospital regarding patient privacy. There are currently more than 1300 FA patients with this rare disease enrolled in the IFAR. The new display screens and navigational tools make the application easy to use, and should eventually enable researchers and physicians worldwide to learn more about genotype/phenotype correlations in FA and related diseases, and the role of the FA proteins in DNA damage repair and cancer.

1630M

RD-Connect platform and standardized exome-phenome analysis pipeline: application to 20 use cases. *S. Beltran[1], D. Salgado[2,3,4], V. de la Torre[5], J. Paschall[6], S. Laurie[1], J.P. Desvignes[2,3], A. Topf[7], M. Calissano[7], I. Zaharieva[8], F. Muntoni[8], M. Roos[9], P. Lopes[10], M. Girdea[11], C. Kingswood[1], M. Vázquez[5], J.M. Fernández González[5], R. Thompson[7], M. Brudno[11], V. Straub[7], H. Lochmüller[7], M. Bellgard[12], J.L. Oliveira[10], P.A.C. 't Hoen[9], A. Valencia[5], C. Béroud[2,3,13], I. Gut[1].* 1) Centro Nacional de Análisis Genómico (CNAG), Barcelona, Catalonia, Spain; 2) Aix-Marseille Université, Marseille, France; 3) Inserm, UMR_S 910, Marseille, France; 4) EMBL Australia, Australian Regenerative Medicine Institute (ARMI), Monash University, Clayton, Victoria, Australia; 5) Centro Nacional de Investigaciones Oncológicas , Madrid, Spain; 6) European Molecular Biology Laboratory , European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Cambridge, United Kingdom; 7) Institute of Genetic Medicine, MRC Centre for Neuromuscular Diseases, Newcastle University, UK; 8) Dubowitz Neuromuscular Centre, UCL Institute of Child Health and Great Ormond Street Hospital for Children, London, United Kingdom; 9) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 10) DETI/IEETA, University of Aveiro, Portugal; 11) Centre for Computational Medicine, Hospital for Sick Children and University of Toronto, Toronto, ON, Canada; 12) Centre for Comparative Genomics, Murdoch University, Perth, Western Australia; 13) APHM, Hôpital TIMONE Enfants , Laboratoire de Génétique Moléculaire, Marseille, France.

Around 300 million people worldwide are estimated to suffer from one of the 6000+ known rare diseases. Rare disease research faces particular challenges because patient populations, clinical expertise, and research communities are small in number and highly fragmented both geographically and in terms of medical specialty. Therefore, discovery is slowed down due to limited access to sufficient patients, high quality information and results and would benefit from greater coordination of patient registries, biobanks, data repositories, and best practices in deployment of new -omics technologies and analysis methods. The EU FP7-funded RD-Connect project is building a platform to harmonize and securely integrate databases, registries, biobanks, clinical bioinformatics and -omics data generated with standardized pipelines. Platform design has consisted of a formal use-case evaluation process performed in collaboration with potential clinical research users and the two main associated EU FP7-funded projects Neuromics (rare neuromuscular and neurodegenerative diseases) and EURenOmics (rare kidney diseases). These projects have adopted the Human Phenotype Ontology to describe their patients and are generating genomic, proteomic and metabolomic data that is being deposited at the European Genome-phenome Archive (EGA). The data is then accessed and processed by RD-Connect's standard analysis pipelines to generate comparable results that are integrated in the platform. We present here the first version of the RD-Connect exome sequencing analysis, annotation and prioritization pipeline, which includes widely used prediction tools such as SIFT, PolyPhen, phyloP or MutationTaster together with novel ones such as Human Splicing Finder and UMD predictor. Tools that use phenotypic, expression and regulatory information have also been used for variant and gene prioritization. We report here the outcome of applying the pipeline to an initial set of 20 use cases and the preliminary version of the platform back-end and front-end that allows storage, access and filtering of the genomic and phenotypic information. The integration of resources being conducted by RD-Connect is a working example of the emerging global rare disease bioinformatics ecosystem.

1631T

PATHWAY APPROACHES TO STRENGTHEN GENETIC VARIATION ANALYSIS. *E. Cirillo, M. Kutmon, C. Evelo.* Department of Bioinformatics - BiGCaT, Maastricht University, The Netherlands.

Introduction. Pathway and network analysis is used to understand genomics data from transcriptomics, proteomics and metabolomics experiments. It uses prior knowledge of biological relationships and can evaluate related effects that strengthen each other. Genetic variants analysis (such as SNPs, indels..) would also benefit from this approach because of the evaluation of related variations in different genes, and integration of variant data with other omics results. Currently, genetic variations cannot easily be combined in pathway representations. It is also not clear how to better visualize and interpret variation data once we connected it to pathway content programmatically. In this project we take up the challenge to integrate genetic information in the open source pathway analysis program PathVisio (pathvisio.org). Design and analysis ideas. We connected variants with related genes, using bioinformatic mapping approaches (see bridgedb.org). With this we can visualize variants on a pathway. We now aim for meaningful presentations showing some design studies: - Variants information for selected genes is shown in a separate panel of PathVisio page. - Genetic variations with statistical values from large genetic datasets (e.g. GWAS) can be highlighted for the annotated variants (colors in the panel and clickable values). - Specific information about variants like HapMap environment or function predictions can be shown as popups or via linkouts to genome browsers. We will also map variants and genes for statistical approaches using larger sets of biological pathways (e.g. from WikiPathways), to find those where many variants occur in specific phenotypes. This is comparable to the over representation and gene set enrichment analyses. Finally, we will investigate the use of existing advanced tools (like sift and polyphen) to integrate predictions of biological functional changes. Pathways as a bridge to networks. Pathway and network biology can be easily combined: our tools convert one into the other and we can use the same mappings between database identifies in both. Variations are evaluated in pathways as networks, increasing network extensions (e.g. with miRNA or drugs) and topology studies. This will combine network connectivity, betweenness and shortest path analysis with the occurrence of deleterious variants in disease phenotypes.

1632S

Mutiple-trait genomic selection and phenotype prediction. *A. Dahl[1,2], J. Marchini[1,2].* 1) Department of Statistics, University of Oxford, Oxford, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom.

Random effect models have been used in genetics to control for aggregate genetic structure in association studies, estimate heritability and predict the latent genetic contribution to phenotypic variation. Unfortunately, the ordinary random effect model does not generalize well to multivariate phenotypes; the resulting high-dimensional parameter estimates are both statistically unstable and computationally intractable. Moreover, missing observations plague mulitvariate data, as typical methods must throw away all samples with even one missing phenotype. Motivated by these concerns, we propose a new random effect model to decompose genetic and environmental contributions to multivariate phenotypic variance. The method naturally incorporates missing data and, in particular, can be used to impute unobserved phenotypes for subsequent analyses. In terms of imputation, this new method outperforms all tested competitors, including standard univariate methods from genetics and state-of-the-art imputation tools from mainstream statistics, on both simulated data and real human glycomics data. A second application is to estimate the latent genetic contribution to phenotypes, or breeding value of study individuals. On simulated data and real data from chicken breeding, the proposed method produces the best estimates of the future breeding values. The model is fit with variational Bayes, enabling this method to be run on thousands of individuals and tens of phenotypes.

1633M

Scaling up genomic data management, indexing, and analysis for a million genomes. *F. De La Vega[1], S. Young[2], J. Wiley[2], A. Patel[2], M. Pae[2], R. Hayek[2].* 1) Annai Systems, Inc., Burlingame, CA; 2) Annai Systems, Inc. Carlsbad, CA.

As sequencing technologies continue to evolve and the cost of genome sequencing drops, several large-scale population sequencing projects with sizes ranging from tens to hundreds of thousand samples are now being started. This sample size imposes difficult demands in data management, metadata indexing, and analysis to which current paradigms of data distribution and access are inadequate. Cloud computing technology is enabling sharing of high-performance computational resources and inspiring new models for data analysis, but still suffer by lack of interoperability and data transport latency. Additional requirement include maintaining secure access to meet informed consent obligations and metadata storage and indexing to facilitate meta-analyzes. Here we present a cloud storage and analysis architecture that satisfies the above requirements. In this architecture we integrate technology we deployed as part of the Cancer Genomics Hub (CGHub) and which currently supports the TCGA, TARGET, and the CGCI programs, and hosts more than 44,000 data files totaling more than 500 Terabytes with capacity to grow to 5 Petabytes. For identity management we use the NIH InCommon identity provider (IdP) via the Shibboleth middleware and for metadata indexing the SRA Metadata XML Schema 1.5. The GeneTorrent (GT) protocol allows data migration supporting preauthentication and authorization keys, SSL encrypted data channels, and genomic metadata integration. GT provides inherent parallelization, and its provision for relatively painless horizontal scaling, allowing federation. We further integrated SeqWare, an open source computational scheduling infrastructure which facilitates automation of pipelines based on metadata, and Sage Synapse, a workspace that aggregates, describes, and allows sharing of analytical pipelines enabling reproducible bioinformatics. We are implementing a pilot of this architecture in support of the ICGC PanCancer project, which will analyze whole genome data from 2,000 tumor-normal pairs federated across six data centers, and for the ICGC-TCGA DREAM somatic mutation calling challenge.

1634T

ClinGen database for curation of clinically relevant genomic variants. *X. Feng[1], T.P. Sneddon[2], H. Chen[1], A.R. Jackson[1], S. Paithankar[1], L.J. Babb[3], C. Bizon[4], D. Maglott[5], E.M. Ramos[6], H.L. Rehm[3,7,8], J.M. Cherry[2], C. Bustamante[2], S.E. Plon[1], A. Milosavljevic[1].* 1) Baylor College of Medicine, Houston, TX; 2) Stanford University School of Medicine, Stanford, CA; 3) Partners Healthcare Personalized Medicine, Cambridge, MA; 4) Renaissance Computing Institute, Chapel Hill, NC; 5) National Center for Biotechnology Information, Bethesda, MD; 6) National Human Genome Research Institute, Bethesda, MD; 7) Harvard Medical School, Boston, MA; 8) Brigham and Women's Hospital, Boston, MA.

The Clinical Genome Resource (ClinGen) is an NIH-funded program dedicated to creating a database of clinically relevant genomic variants to inform genome interpretation in a variety of clinical contexts. ClinGen is collaborating closely with the NCBI ClinVar database and is supporting algorithmic and expert curation of data from a variety of sources, including hundreds of thousands of test results shared by clinical genetics laboratories through ClinVar and other research data and bioinformatic predictions. A core component of ClinGen is the development of ClinGenDB, a workspace with a dedicated user interface (UI) for curators to collect and track additional data on variants in ClinVar. It will draw data from a variety of clinical and research databases, including several genomics initiatives, and enable expert review of variants as well as resolution of discrepant classifications in ClinVar. ClinGenDB will support ongoing data updates and versioning and will be accessible through REST Application Programming Interfaces (REST APIs) for algorithmic processing. In ClinGenDB, the data model driven UI presents curators with aggregated and computationally derived variant information. The data is structured to support variant evaluation according to configurable guidelines such as the new American College of Medical Genetics and Genomics (ACMG) Interpretation of Sequence Variants (ISV) guidelines. Expert curation is supported by a semi-automated decision-making system. The system will apply the configurable guidelines to infer provisional conclusions about variants. The conclusions and justifications provided by the system will then be reviewed by curators with ongoing versioning of assertions. We anticipate that implementation of the data model and decision-making system in ClinGenDB will greatly assist curation of clinically relevant genomic variants. Toward this goal, we will provide several case studies of variant curation using a working prototype of ClinGenDB and the variant curation UI, focusing on the subset of 56 genes identified in a recent report by the ACMG on the return of incidental findings (Green et al., 2013).

1635S

Efficient sharing of exome/genome variant and phenotype data between diagnostic labs. *I.F.A.C. Fokkema, J.T. den Dunnen, P.E. Taschner.* Dept Human Genetics S-4-P, Leiden Univ Medical Ctr , Leiden, Netherlands.

Gene variant databases (LSDBs) using the LOVD platform (Leiden Open-source Variation Database, http://www.LOVD.nl) collect and share information about genes, variants and phenotypes (diseases) from research and diagnostic labs. LOVD version 3 also facilitates exome and genome sequence data analysis. LOVD can be queried and adapted for many different purposes. The Dutch and Belgian working group for Breast Cancer DNA Diagnostics (LOB) has decided to share >7500 variants detected in the BRCA1/2 genes in breast cancer families since 1997. For this, the data, almost evenly split among both genes, have been submitted to the LOVD 3 shared gene variant database installation (1,2). Others include: world-wide BRCA variant data sharing (Human Variome Project and Global Alliance for Genomics and Health); Insight Consortium colon cancer variant database; country-specific data views (Finnish Disease (FinDis) portal). LOVD system advantages: simple standardized submission of new data, instant updates after curation, easy maintenance and automatic backups. Although most data are publicly accessible online, some data (detailed phenotype information) are shared by consortium members only using the new LOVD3 access level, designated “collaborator”. Others can see whether such information is available (password protected file links), giving them the option to contact the submitter for further details. Members can contribute their opinions about variant classification, increasing its consistency, but being aware of potential misinterpretation they have reservations sharing this information. Data are stored variant-by-variant and connected to each individual patient and submitting diagnostic lab. Users can perform queries per gene or individual, use other linked resources of interest, view data tracks in genome browsers and use web services to access variants stored in other gene variant databases. For databases curated by us both phenotype descriptions and/or gene variants can be submitted with the request to assign the so-called VIP-status for both variants and phenotypes demanding specific attention. Within 3 weeks, this match-making feature successfully brought researchers into contact, cracking rare disease cases and resulting in a high-impact publication. 1) http://databases.lovd.nl/shared/genes/BRCA1 2) http://databases.lovd.nl/shared/genes/BRCA2 The Human Variome Project has granted LOVD the recommended system status for variant collection.

1636M

Design and Implementation of an Informatics Infrastructure for Clinical Genomics Analysis and Reporting. *J. Hirsch, A. Hsu, T. Loeser.* Syapse, Palo Alto, Ca.

This talk will discuss challenges and recent developments in implementing informatics infrastructure to support clinical use of genomics, specifically in the area of oncology. While much focus has been placed on the ability to properly call variants, the “last mile” in taking processed variant calls and producing a useful, annotated clinical report with therapeutic recommendations is critical for enabling the clinical use of genomics data. Generating, delivering, and updating a clinical genomic sequencing report that a physician can use in routine care has proven to be a challenge, driven by reliance on outdated software tools. To bring genomics data to point of care, labs and clinics must be able to handle three types of data with very different properties and requirements: evolving biomedical knowledge, complex medical data, and high-volume genomics data. This talk will present a software solution, based on semantic computing principles, that enables the integration of all three data types for the purposes of clinical reporting. We will present the results of clinical pilots focused on maintenance of an evolving variant interpretation knowledge base, sharing of variant interpretation across institutions, and the use of this knowledge base to automate clinical genomic report generation and updating, as knowledge changes. We will describe the interpretation of identified variants through a shared knowledge base, automated assembly of annotated clinical reports, and delivery of reports to clinicians at point of care. Results of a pilot program to integrate these technologies into practice within an integrated community health care system will be presented. We will also present the results of physician adoption of a web-based, interactive clinical genome report format that incorporates clinical care guidelines with genomic report data.

1637T

Functional interpretation of noncoding somatic variants from cancer genomes. *E. Khurana, Y. Fu, M. Gerstein.* Yale University, New Haven, CT.

Whole-genome sequencing generally reveals thousands of somatic variants in individual tumors. Identification of key variants responsible for tumor growth and progression is a major challenge. This is especially the case for noncoding variants that are less well understood than coding variants. Discovery of somatic mutations in telomerase reverse transcriptase (TERT) promoter across many different cancer types has shown that regulatory variants may constitute driver events. However, most current methods for identification of cancer driver mutations focus primarily on protein-coding genes. We previously developed a method for identification of candidate noncoding drivers by integrating patterns of selection from 1000 Genomes data with functional annotations from ENCODE (FunSeq). Besides analyzing loss of transcription factor binding sites, our method now identifies creation of new binding sites by somatic mutations in promoters and enhancers of coding genes. We have also integrated function-based FunSeq scoring scheme with patterns of recurrence of somatic variants across hundreds of patients from multiple studies. Tumors exhibit differential expression signatures of thousands of genes. Our unified approach allowed us to explore the fraction of gene expression changes that may be explained by regulatory mutations vs. mutations in coding genes across many different cancer types.

1638S

Update and expansion of Human Variation Database in Japanese Database Integration Program. *A. Koike[1], H. Sawai[2], I. Inoue[3], S. Tsuji[4], K. Tokunaga[2].* 1) CRL, Hitachi, Ltd., Kokubunji, Japan; 2) Dept. Hum. Genet, Univ. Tokyo; 3) Nat. Inst. Genet; 4) Dept. Neurology, Univ. Tokyo.

The Human Variation Database (http://gwas.biosciencedbc.jp/cgi-bin/hvdb/hv_top.cgi) is a repository database set up to achieve continuous and intensive management of Japanese GWAS data and variations identified by NGS and other experimental methods. We have widely called for data submission to aim for data-sharing among researchers. Variations include short/long insertions and deletion and structural variations as well as SNPs/SNVs. Since vast amount of knowledge about disease-variation relationships are varied in the scientific literature, these data have been extracted by manual curation to organize the disease-related variations and improve our understanding of disease mechanisms and disease-disease relationships. More than 30,000 disease-related variation entries are currently registered with allele/genotype frequencies and statistical data. Pathway data are also provided to facilitate the understanding of epistasis and relationships between similar diseases. Since some disease-related variations are recognized to be different among populations, the reference genomes are built in each population and are registered.

The recent improvements of the database are as follows. 1) The scope of the registry was extended to include drug response-related variations and virus resistance-related variations. 2) HLA-DB was constructed because of the characteristics of the HLA region. 3) Links to driver mutations archived in widely used cancer somatic mutation databases were added in order to clarify the relationships between the driver mutations in various cancers and germ-line mutations, although only variations in germ-line cells are targeted in the database. In this presentation, we overview the database structure and discuss the differences of deleterious variations between germ-line cells and somatic cells from the view points of physicochemical properties, functional properties, and evolutionary conservation properties.

Acknowledgement: This work was supported by the contract research fund 'Database Integration Program' from the Japan Science and Technology Agency.

1639M

Exploring the genome-wide roles of transcription factors and their complexes in chromosome interaction. *MJ. Li[1,2], LY. Wang[1], PC. Sham[3], MQ. Zhang[4], JS. Liu[2], JW. Wang[1].* 1) Biochemistry, The University of Hong Kong, Hong Kong; 2) Statistics Dept, Harvard University, Boston, MA, USA; 3) Psychiatry Dept, The University of Hong Kong, Hong Kong; 4) Molecular and Cell Biology Dept,, The University of Texas at Dallas, Dallas, TX, USA.

The tight regulation of genes in different cells is governed by temporal and spatial biological signals. It is very important to pinpoint the pattern of transcription factors (TFs) and their complexes in looping interactions and to detect TF complexes as well as the underlying cis-regulatory modules (CRMs) in different human cell types. Existing studies on analysis of TFs and their complexes were only performed at one dimension and not at genome-wide scale. Recently, the unbiased chromosome conformation capture, Hi-C, can detect the genome-wide chromatin interactions, but has restrictions on its resolution due to the variable cell-to-cell chromosome structures and inadequate sequencing depth. In this study, we provide a comprehensive analysis on TFs regulatory pattern within chromosome looping by combining Hi-C and ENCODE ChIP-Seq data from three human cell types (GM12878, H1-hESC and K562). We first devised a strategy to map ChIP-Seq peaks of each TF to a normalized 10kb Hi-C contact matrix and construct an interaction matrix for each participant TF. We observed tight correlation for TFs participant activities in high resolution chromosome looping between biological replicates, which indicate the TF activities is more stable than local DNA interactions. To check the enrichment of different chromatin marks and genomic features in the interaction region of each participant TF, we performed enrichment test on several histone modifications marks (H3K4me1, H3K4me2, H3K4me3, H3K27me3, H3K36me3, H3K9ac, H3K27ac), CAGE, DHSs, DNA methylation (RRBS) and conservation signals. We found chromatin markers and genomic features of each TF is highly correlated between replications. We also observed that grouped TFs shared consistent patterns of chromatin marks and genomic features, which indicate their similar roles in gene regulation. Furthermore, we developed a method to detect the candidate TF complexes by reducing the dimension of huge sparse interaction matrix for each participant TF followed by community clustering. We successfully detected several known complexes mediating long range enhancer-promoter interaction in different cell types. Using the combinatory motifs scanning, we can predict the genome-wide CRMs for specific TF complex with known motifs. Finally, by correlating gene expression profiles of TF and its targets , we can pinpoint the role of detected TF complexes (active or repressive) in controlling cell type specific gene regulation.

1640T

Comparing blood and brain gene expression networks in Huntington's Disease by semantic analysis. *E. Mina, P.A.C. 't Hoen, W. van Roon-Mom, M. Roos.* Department of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.

Huntington's Disease (HD) is a neurodegenerative disease with the most prominent pathology in the brain. However, human brain tissue is not easily accessible for molecular biology studies and it cannot be isolated from living patients. The widespread pathology of HD indicates that the phenotype is not limited to brain dysfunction. Several studies have shown that understanding alterations in peripheral tissues could be valuable for monitoring disease progression. Considering that transcriptome technologies have successfully been used for biomarker discovery and the study of physiological and pathophysiological mechanisms, it is evident that studying transcriptional changes in peripheral tissue can provide new insights that can lead to the development of new therapies and markers to monitor disease progression. In addition, novel sequencing technologies and the wide availability of data and published articles, create opportunities for novel, more effective ways of data exploitation. Our methodology identifies and prioritizes disease signatures from blood that are also associated with abnormalities in brain, in a robust way by a combination of large scale network analysis, integration of heterogeneous datasets, and incorporation of prior knowledge mined from literature. Our preliminary data show that modules in gene co-expression networks are significantly more reproducible than individual gene markers between blood and brain. We additionally link significant HD signatures to biological and pharmacological knowledge sources to prioritize biomarkers based on their potential as drug targets using a 'Linked Data' approach. This allowed us to identify drug targets by querying across four different data sources, including drugbank and pathway data to target pathways that are impaired in HD. For instance, we found Imatinib, associated with the gene ABL1 that targets autophagy and is known to delay production of beta-amyloid. Moreover, in the HD context, we prioritize drug targets according to their structural or functional similarity to known targets that can pass the blood-brain barrier. In conclusion, our approach is a powerful method to identify genes that are potentially involved in physiological and pathophysiological processes in HD. Using the 'Linked Data' approach, our results can be easily extended and will become available to the scientific community as a resource for further investigation for understanding disease pathology in HD.

1641S

Genetic Risk Prediction and Neurobiological Understanding of Alcoholism. *A. Niculescu[1, 12], D. Levey[1], H. Le-Niculescu[1], J. Frank[2], M. Ayalew[1], N. Jain[1], B. Kirlin[1], R. Learman[1], E. Winiger[1], Z. Rodd[1], A. Shekhar[1], N. Schork[3], F. Kiefer[4], N. Wodarz[5], B. Muller-Myhsok[6], N. Dahmen[7], M. Nothen[11], R. Sherva[8], L. Farrer[8], A. Smith[9], H. Kranzler[10], M. Rietschel[2], J. Gelernter[9], GESGA Consortium.* 1) Psychiatry, Indian Univ Sch Medicine, Indianapolis, IN; 2) Central Institute of Mental Health, Mannheim, Germany; 3) Department of Human Biology, The J. Craig Venter Institute, La Jolla, California, USA; 4) Dept. of Addictive Behavior and Addiction Medicine, Central Institute of Mental Health, Medical Faculty Mannheim / Heidelberg University. Germany; 5) Dept. of Psychiatry, University Medical Center Regensburg, Univ. of Regensburg, Germany; 6) Dept. of Statistical Genetics, Max-Planck-Institute of Psychiatry, Munich, Germany; 7) Dept. of Psychiatry, Univ. of Mainz, Germany; 8) Boston University School of Medicine, Department of Medicine (Biomedical Genetics); 9) Yale University School of Medicine, Department of Psychiatry, Division of Human Genetics; and VA CT Healthcare Center; 10) Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, and Philadelphia VAMC; 11) Dept. of Genomics, Life & Brain Center, Univ. of Bonn; Inst. of Human Genetics, Univ. of Bonn, Germany; 12) Indianapolis VA Medical Center, Indianapolis, Indiana, USA.

We have used a translational convergent functional genomics (CFG) approach to discover genes involved in alcoholism, by gene-level integration of genome-wide association study (GWAS) data from a German alcohol dependence cohort with other genetic and gene expression data, from human and animal model studies, similar to our previous work in bipolar disorder and schizophrenia. A panel of all the nominally significant p-value SNPs in the top candidate genes discovered by CFG (n=135 genes, 713 SNPs) was used to generate a genetic risk prediction score (GRPS), which showed a trend towards significance (p=0.053) in separating alcohol dependent individuals from controls in an independent German test cohort. We then validated and prioritized our top findings from this discovery work, and subsequently tested them in three independent cohorts, from two continents. In order to validate and prioritize the key genes that drive behavior without some of the pleiotropic environmental confounds present in humans, we used a stress-reactive animal model of alcoholism developed by our group, the DBP knock-out mouse. A much smaller panel (n=11 genes, 66 SNPs) of the top CFG-discovered genes for alcoholism, cross-validated and prioritized by this animal model, showed better predictive ability in the independent German test cohort (p= 0.041). The top CFG scoring gene for alcoholism from the initial discovery step, synuclein alpha (SNCA), remained the top gene after the stress-reactive animal model cross-validation. We also tested this small panel of genes in two other independent test cohorts from the United States, one with alcohol dependence (p=0.00012), and one with alcohol abuse (a less severe form of alcoholism) (p=0.0094). SNCA by itself was able to separate alcoholics from controls in the alcohol dependent cohort (p= 0.000013) and the alcohol abuse cohort (p= 0.023). So did 8 other genes from the panel of 11 genes taken individually, albeit to a lesser extent and/or less broadly across cohorts. SNCA, GRM3 and MBP survived strict Bonferroni correction for multiple comparisons. Taken together, these results suggest that our stress-reactive DBP animal model helped to validate and prioritize from the CFG-discovered genes some of the key behaviorally relevant genes for alcoholism. These genes fall into a series of biological pathways involved in signal transduction, transmission of nerve impulse (including myelination), and cocaine addiction.

1642M

Analyst Portal - a real-time, distributed web query tool that streamlines data search at a genomics center. *H. Qiu, C. Kim, F. Mentch, L. Hermannsson, H. Hakonarson.* Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA.

In the setting of a large genomics research center, we constantly have requirements to have up-to-date detailed workflow information as well as efficiency in querying available data in seamless and integrated way to help scientific discoveries. Instead of taking the conventional data warehouse approach to extract data from various sources into a centralized data store, which is both time consuming and hard to co-ordinate with the real-time workflow status, we use distributed database query facility to dynamically query different source databases. The resulting suite of web applications on top of the distributed real-time query layer is being referenced as Analyst Portal. Data modalities searched by Analyst Portal include service order intake, biorepository sample management, SNP-array genotyping, gene array expression data, next generation sequencing, DICOM neuroimaging, phenotype data from hospital EMR system and from collaborators, catalog of genetic variants (SNP, CNV, indel) . For individual modality, Analyst Portal supports three levels of detailed data queries: single subject information lookup, all records in the modality, and a search limited to an uploaded subject list. In case of multiple domains search, Analyst Portal supports cohort selection by combining user defined search terms from different modalities and generating complex distributed SQL query. For example, a query like “get a list of Caucasians, males only, 6-10 years old subjects who have at least 2 ADHD diagnoses over the past 3 years, and on at least one medication such as Adderal, Concerta or Ritalin, and have blood DNA samples in our biorepostiory, and already genotyped on Illumina or Affymetrix SNP chips, and not NGS exome sequenced yet” can be done real-time easily in Analyst Portal. Among 50 typical data queries scientists and researchers have used during previous research studies, the Analyst Portal has ability to address 70-80% of them instantly. Here, we review the software architecture, performance tuning, user experience and areas for future optimization.

1643T

Phenotype terminologies in use for genotype-phenotype databases: A common core for standardisation and interoperability. *P.N. Robinson[1], S. Aymé[2], L. Chanas[2], A. Hamosh[3], A. Rath[2], International Consortium for Human Phenotype Terminologies.* 1) Institute for Medical Genetics, Charite-Universitätsmedizin, Berlin, Germany; 2) INSERM, US14, Paris, France; 3) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

The community needs to be provided with terminology standards in order to achieve interoperability between databases intended for clinical research. Phenotypic information is crucial to interpret genomic rearrangements as well as high-throughput sequence data including whole exome and whole genome sequencing. The aim of our work was to develop a core terminology of phenotypes interoperable with all terminologies in current use. Relevant terminologies in use by different communities to describe phenomes were cross-referenced: PhenoDB (2846 terms), London Dysmorphology Database (LDDB; 1318 terms), Orphanet (1243 terms), Human Phenotype Ontology (9895 terms, 22/08/2102), Elements of Morphology (AJMG; 423 terms), ICD10 (1230 terms), as well as medical terminologies in use: UMLS (7,957,179 distinct concept terms), SNOMED CT (>311,000 concepts), MeSH (26,853 concepts) and MedDRA (69,389 concepts). We established a strategy to compare the terminologies to find commonalities and differences, using ONAGUI as a tool to identify exact matches. The non-exact matches were verified manually by an expert. A core-terminology of 2,300 terms was derived and analyzed by a panel of experts (International Consortium for Human Phenotype Terminologies - ICHPT). The resulting consensual terminology will be freely available in a dedicated website (www.ichpt.org) and at the International Rare Disease Research Consortium (IRDiRC) website. Mappings from ICHPT to other terminologies will be given in order to enable interoperability between databases that use the various terminologies to annotate phenotypic data, with the goal of enabling data exchange between all major genetic databases.

1644S

Data exploration through stark visualizations in gene expression profile of down syndrome. *J. Rualo[1], P. Skinner[1], D. Weaver[1], N. Economou[1], DG. Knowles[2], E.G. Couto[2].* 1) Translational Medicine, PerkinElmer, Inc., Waltham, MA; 2) Bioinformatics, Integromics, Waunakee, WI.

Life Scientists are increasingly challenged by huge datasets resulting from High-Throughput Experiments that contain an enormity of information. Making sense of the big data produced is essential to understanding complex biological systems and doing so without wasting unnecessary resources is key. Frequently this problem is addressed using different software solutions for each technology and handling many steps manually or with ad-hoc scripting. This generates problems when comparing results from different technologies in a reliable way and makes it difficult to represent the data for analysis when attempting to answer biological questions of interest. In this work we evaluate if the use of interactive visual exploration of the data can provide the researcher with the ability to better evaluate their data and extract relevant biological conclusions. By taking advantage of visual data representations and the human capacity to immediately recognize patterns in visual data, we demonstrate how this approach accelerates data exploration and helps to reveal information easily missed using the traditional data analysis paradigm. In the traditional approach the data visualization is only used as the end product of the analysis. The initial steps and the exploratory analysis use limited visual aids that are frequently highly summarized visualizations focused on a specific question; thus, important trends are easily missed when not specifically searched. In order to demonstrate how visualizations can help not only in the representation of the final results but also during the entire analysis process we have reanalyzed the expression data from a recent study published in Nature (Letourneau A. et al. Domains of genome-wide gene expression dysregulation in Down's syndrome. Nature 2014 Apr 17;508(7496):345-50) together with other expression datasets related to Down Syndrome deposited in the GEO database. We show how by using interactive and visual exploration at the different steps of the analysis in combination with the standard tools used during the traditional data analysis we can reveal several interesting features of the data that can be easily missed and ignored when using only the conventional analysis & reporting.

1645M

Evaluating global enrichment of trait-associated variants in epigenomic features. *E. Schmidt[1], J. Chen[2], W. Zhou[2], J. Zhang[2], E. Chen[2], C. Willer[1,2].* 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI.

Genome wide association studies have identified hundreds of non-protein-coding variants with modest effect sizes, making alterations in transcript levels difficult to detect. This trait-associated genetic variation likely plays a role in transcriptional regulation via epigenomic regulatory features, a diverse set of which have been experimentally defined by ENCODE in various cell types. We develop GREGOR (**G**enomic **R**egulatory **E**lements and **G**was **O**verlap algo**R**ithm) to address 3 aims: 1) elucidate the important tissue types in which genetic variation impacts transcription for a particular trait, 2) narrow our focus of the regulatory features underlying transcription disrupted by trait-associated variants, and 3) use positional overlap with selected regulatory domains to identify potential functional candidate variants at trait-associated loci.

Taking into account linkage disequilibrium, we evaluate enrichment of 5 sets of metabolic-related trait-associated variants in a range of regulatory features by comparison to control variants from 1000 Genomes. We match each GWAS SNP with control SNPs based on 3 properties: i) number of variants in high LD, ii) minor allele frequency, and iii) distance to nearest gene. We find enrichment of aggregated sets of trait-specific variants in DNase hypersensitive sites in biologically relevant tissue types. Examples include significant enrichment *P*-values and fold change of variants associated with blood pressure in blood vessel ($P=1\times10^{-9}$; 1.5), body mass index in olfactory neurosphere-derived cells ($P=4\times10^{-5}$; 1.7), coronary artery disease in heart ($P=2\times10^{-5}$; 1.7), lipids in liver ($P=2\times10^{-14}$; 1.6), and unexpectedly, type 2 diabetes in heart ($P=1.4\times10^{-6}$; 1.7). We further evaluate the position of index SNPs and their LD proxies relative to histone marks and ChIP-seq TF binding sites, as well as functional chromatin states.

We prioritize individual non-coding variants at GWAS loci and experimentally test candidate SNPs at ≥ 5 GWAS lipid loci using a luciferase reporter to measure transcript activity in human HepG2 cells. In all, our tool systematically employs the wealth of regulatory data available from ENCODE to establish known biological connections with trait-associated variation, find potentially novel biological mechanisms, and guide experimental follow-up to identify functional variants at GWAS loci.

1646T

A bioinformatic protocol for the study of rare diseases. *F. Tobar Tosse, E. Ocampo Toro, P. Hurtado Villa.* Department of basic Science for Health, Pontificia Universidad Javeriana Cali, Cali, Valle del Cauca, Colombia.

Introduction: Rare diseases (RDs) involve several genetics and molecular mechanisms, which describe diverse and specific phenotypes and conditions; which difficult the concentration of efforts to research and propose new therapies for the patients. However, there is a relevant feature in these kind of diseases; that is the genetic context, where a single variant in a gene could define the whole context of the disease, for example, in the lysosomal storage diseases. Additionally, there is a very important fact, that is the consequences in the genome; these mean, although several mechanisms are associated to each RDs, these can be generally described as aberrations of the human genome, whose organization is disrupted and therefore the dynamic is changing too. Results and discussion: Accordingly, at the present work we present a bioinformatics workflow for the association of genomic information related with RDs. We improved and applied methods of data mining and nodes theory, considering the phenotypic and genetics descriptions for several RDs. This approach allow us to identify, molecular and genomics associations among several rare diseases. Interestingly, physiological and phenotypic descriptions could be associated to the combination of genomic patterns, and could describe focused of genomic susceptibility. Additionally, it was associated to affectation of critical metabolic or physiological processes. For example, we identify several syndromes related to adrenal Insufficiency with genomics patterns (Based on density of repetitive elements), that phenotypically are related to hydroxylase deficiency, and the consequences in the lipid metabolisms. The relational network shows this metabolic process as critic node in the genomics associations identified in Rare diseases, therefore genomics context could define unknown related factors important for the exploration of new genetics, or new targets for disease exploration. Conclusion: Although Rare Disease are described as diverse molecular and genetics anomalies, by bioinformatics approaches could be possible the identification of genomics and metabolic connections, that allows the definition and implementation of new methods for specific target in several RDs.

1647S

EnhancerDB: a database of human enhancers and their putative targets. *P. Wang, L.Y. Wang, M.Q. Zhang, Q. Lian, J. Wang.* The University of Hong Kong, Hong Kong, Hong Kong, Hong Kong.

As a major class of distal cis-regulatory elements, enhancers are crucial to the regulatory system orchestrating gene transcription. Thereby detection of enhancers has been extensively studied and numerous enhancers have been detected by experiments and predicted by computational models. Here we present a database containing ~3,000,000 human enhancers. 3,598 enhancers are experimentally detected and collected from 190 publications, and the rest are predicted based on transcription factor binding motifs, high conservation, histone modification marks or eRNA expression. Meanwhile, we incorporated data of transcription factor binding motifs and ChIP-Seq peaks to annotate the enhancers. In addition, we used available long-range chromatin interaction data, including ChIA-PET and Hi-C in 10 human cell lines to present putative target genes of the enhancers.

1648M

A highly efficient and scalable compute platform for massive variant annotation and rapid genome interpretation. *J. Warren[1], E. Colak[1], A. Kiani[1], J. Li[2], S. Bhattacharya[3], N. Bani Asadi[1,2], S. Barr[1], A. Butte[3], G. Nolan[4], R. Chen[5], W.H. Wong[6,7], H.Y.K. Lam[2].* 1) Department of Engineering, Bina Technologies, Redwood City, CA; 2) Department of Bioinformatics, Bina Technologies, Redwood City, CA; 3) Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford CA; 4) Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford California 94305; 5) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 6) Department of Statistics, Stanford University, Stanford, CA; 7) Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA.

After obtaining the variants from next generation sequencing data, researchers and clinicians still face the undertaking of interpreting the results. Despite the availability of numerous public databases, using this collective information is an arduous task due to inconsistent data, multiple versions and nonstandard formats. Even after aggregating the data and annotating the variants, it remains a laborious exercise to identify the causative variants associated with the disease in question.

To address this challenge we present a highly efficient data pipeline that leverages big data technologies to integrate annotations from a large range of biological databases. The pipeline takes variant callsets, annotates all samples, and indexes the variants to support real-time queries and analytics. Specifically, it uses Hadoop MapReduce to perform extensive precomputation on the annotation data; builds indexes of the annotated variants; versions and manages data in both relational and NoSQL databases; and horizontally scales to increase throughput. With this approach, a 5-node cluster can annotate a typical whole exome sequencing (WES) sample with ~300 million annotation records (>200GB) in under 30 minutes or a whole genome sequencing (WGS) sample under an hour. Annotating a sample and indexing its variants are computationally demanding steps, but these are one-time costs. After the indices are generated, users can concurrently and repeatedly perform ad-hoc queries to identify the variants that meet given criteria. On average, it takes less than one second for any complex query to return the results on a single sample.

Researchers can easily query the samples in a database-centric view or a genome-centric view. In one application we analyzed the WGS data from the DNA of *Ata*, the skeletal remains of a 6-inch human found in the Atacama Desert, Chile. The annotation and analysis took less than an hour for >3M total variants and identified 37 functional changing variants located in genes highly associated with dwarfism and skeletal dysplasia. Four of these variants were not found in dbSNP and may be of interest for further investigation. Such filtration greatly reduces the search time and space for researchers. With predefined statistical workflows, researchers can also carry out genetic studies for case-control, trio, cohort, and tumor-normal studies.

1649T

Investigating the genetic architecture of pulmonary arterial hypertension shared with other diseases. *L.A. Yancy[1,2], V.A. de Jesus Perez[3,4], R.T. Zamanian[3,4], A.J. Butte[1,2].* 1) Biomedical Informatics Training Program, Stanford University School of Medicine, Stanford, CA; 2) Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA; 3) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA; 4) Vera Moulton Wall Center for Pulmonary Vascular Disease, Stanford University School of Medicine, Stanford, CA.

Pulmonary arterial hypertension (PAH) is a rare and fatal disease with a median survival of approximately 4 years for those afflicted. Many genome-wide association studies (GWASs) have attempted to determine the variants predictive of PAH, but often due to small sample sizes, studies have not led to highly predictive markers. However, GWASs have been successful in determining variants associated with many other diseases. We have an ongoing effort to curate these variants into a database called VARiants Informing MEDicine (VARIMED). VARIMED contains results from over 17,000 peer-reviewed genetic epidemiology studies (e.g. GWAS), covering over 460,000 variants (in over 18,000 genes) associated with over 6,500 phenotypes. In this study, we leverage VARIMED to determine if PAH exhibits a shared genetic mechanism with other diseases. Finding other diseases that have a significant enrichment of PAH associated variants may provide insight into (novel) variants or genes that play a significant role in PAH (and are otherwise not detectable in a PAH GWAS due to small sample size) or even novel therapies for PAH that can be "borrowed" from diseases with similar genetic architecture. Here, we perform a GWAS on exomes (~88,000 variants) from PAH subjects, rank those variants based on p-values and odds ratios (leaving ~350 "priority" variants), and identify genes associated with those prioritized variants (~300 genes). Using those "prioritized" genes, we then query VARIMED and identify diseases with a similar genetic architecture (i.e. identify those diseases with a significant enrichment of top-ranked genes from our PAH GWAS). Through this, we found significant similarities between PAH and a number of immune-related diseases, with the most significant finding being an unusual acquired kidney disease. This work can be extended to investigate the shared genetic architecture of other diseases (rare or otherwise understudied) by taking advantage of the large spectrum of diseases now with characterized genetic architecture.

1650S

Integrative genomic analysis of exome and RNA-seq data from multiple tissue sources obtained from a single breast cancer patient. *A. Vladimirova, R. Goold, S. Sanga, T. Klingler.* Station X Inc., San Francisco, CA.

The ability to analyze genomic data from differing sequencing methodologies and multiple tissue types in a single cancer patient provides a tantalizing opportunity to explore the capabilities of an integrative genomic study to elucidate potential tumor drivers and mechanisms. Simply because of sample availability, researchers are typically limited to analyses of primary tumor data. Access to data from both RNA-Seq and DNA-Seq datasets from adjacent normal tissue, blood and metastatic tissue samples in a single patient is rare, yet could greatly increase signal and reduce noise in these analyses. Technological advances and cost reductions mean truly integrative genomic analyses, in which mutational signatures can be reviewed in concert with transcriptional profiles across relevant samples from multiple tissues, could be used to understand tumorigenesis and potentially guide intervention. We present here a proof of concept study of a single breast cancer patient where we integratively analyze exome and RNA-Seq data obtained from four sample types - primary tumor, blood, adjacent normal tissue and metastasis. We use GenePool™, a commercial software application for biomarker discovery and validation in large patient cohorts (Station X, San Francisco, CA) to quickly prioritize variants and genes found in the tumor and metastasis by comparison to the blood and the adjacent normal breast tissue. Similarly, we analyze the primary tumor relative to the metastasis to identify differences in mutational signatures and transcriptional profiles. We present the results of this single patient integrative analysis and show how GenePool can extend these types of analyses to cohorts of patients, becoming a useful tool in the understanding and treatment of tumors.

1651M

Limitation of multiple testing through the integration of TCF7L2 DNA occupancy and SNP association data reveals *GIP* and *CPPED1* as novel type 2 diabetes loci. *M.E. Johnson[1], J. Zhao[1], J. Schug[2], S. Deliard[1], Q. Xia[1], V. Guy[1], J. Sainz[3], K.H. Kaestner[2], A.D. Wells[4], S.F.A. Grant[1,2,5].* 1) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Genetics and Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Institute of Biomedicine and Biotechnology of Cantabria (IBBTEC), Spanish National Research Council (CSIC), Santander, Spain; 4) Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA; 5) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

The *TCF7L2* locus is strongly implicated in the pathogenesis of type 2 diabetes (T2D). We previously mapped the genomic regions bound by TCF7L2 using ChIP-seq in the HCT116 cell line, revealing an unexpected highly significant over-representation of specific GWAS loci categories. Given that only a minority of the predicted genetic component to most complex traits has been identified to date, plus the fact that a number of GWAS-implicated transcription factors have now been shown to exhibit statistically significant preferential binding to loci associated with complex traits, we investigated if restricting association analyses to just the genes uncovered from our ChIP-seq approach in order to substantially reduce multiple testing could yield novel T2D loci. When investigating all the known GWAS loci bound within 5kb by TCF7L2 (most likely to be functional), derived from HCT116, apart from the known *TCF7L2* locus itself (rs7901695), the coronary artery disease associated T allele of rs46522 within the *UBE2Z-GIP-ATP5G1-SNF8* locus yielded significant and novel DIAGRAM-derived association with T2D risk (OR=1.07; $P=3.20\times10^{-4}$); indeed, the occupancy site was approximately 4kb from the transcription start site for GIP in an intergenic region known to be a hub for binding proteins, H3K27Ac histone marks and open chromatin via a DNAse I hypersensitive site. Indeed the locus encoding the receptor for GIP (*GIPR*) has already been reported in relevant GWAS settings to be associated with BMI and glucose and insulin response metrics. Furthermore, when we analyzed tag-SNPs within genes not previously implicated by GWAS but bound within 5kb by TCF7L2 in HCT116, we observed significant association within the same dataset of the A allele of rs4780476 within *CPPED1* with T2D risk (OR=1.1, $P=4.10\times10^{-5}$). Furthermore, the TCF7L2 occupancy site was in the immediate *CPPED1* promoter region. This is an equally notable observation, as only one paper has been published to date on this gene product and shows that down-regulation of *CPPED1* expression improves glucose metabolism in vitro in adipocytes. Our limitation of multiple testing is based on biological plausibility, where a GWAS-implicated transcription factor is clearly pointing us to genes that are genetically associated with complex disease more often than expected by chance and thus may also be pointing us to novel genes where their strength of the association was at the level of noise at the genome wide scale.

1652T

Metadata-driven tools to access data from the ENCODE project. *E.T. Chan[1], C.A. Sloan[1], V.S. Malladi[1], J.S. Strattan[1], L.D. Rowe[1], B.C. Hitz[1], N.R. Podduturi[1], F. Tanaka[1], B.T. Lee[2], K. Learned[2], M. Maddren[2], S. Miyasato[1], M. Simison[1], E.L. Hong[1], W.J. Kent[2], J.M. Cherry[1].* 1) Genetics, Stanford University, Palo Alto, CA; 2) Center for Biomolecular Science and Engineering, University of California at Santa Cruz, Santa Cruz, CA.

The Encyclopedia of DNA Elements (ENCODE) project is an ongoing collaborative effort to create a comprehensive catalog of functional elements in the human and mouse genomes. Its current data corpus exceeds 4000 experiments across more than 400 cell lines and tissues using a wide array of experimental techniques to survey the chromatin structure, regulatory and transcriptional landscape in human and mouse genomes. All ENCODE experimental data, metadata and associated computational analyses are submitted to the ENCODE Data Coordination Center (DCC) for validation, tracking, storage, and distribution to community resources and the scientific community.

As the volume of data increases, the identification and organization of data sets becomes challenging to present to users. Here, we describe the web interface, search tools and underlying database the DCC have built for simple and intuitive access to ENCODE data and metadata. Extensive and structured metadata describing experimental variables, such as the biological samples, specific reagents, and protocols necessary to replicate the assays and their subsequent analysis collected from ENCODE data producers drive a powerful faceted browsing interface, allowing users to filter and retrieve particular slices of the large data corpus. Elasticsearch-driven real-time indexing also allows users to perform full-text searches to directly access specific data of interest. Upcoming features planned include access to data standards, quality metrics for data files, data visualization tools, uniform processing and analysis pipelines as part of the revamped ENCODE public portal. Data and metadata from the ENCODE project can currently be accessed at https://www.encodedcc.org.

1653S

Accessing ENCODE project data using a REST API and JSON. *C.A. Sloan[1], E.T. Chan[1], V.S. Malladi[1], L.D. Rowe[1], J.S. Strattan[1], B.C. Hitz[1], N.R. Podduturi[1], F. Tanaka[1], K. Learned[2], B.T. Lee[2], S. Miyasato[1], M. Simison[1], E.L. Hong[1], W.J. Kent[2], J.M. Cherry[2].* 1) Genetics Department, Stanford University, Palo Alto, CA; 2) School of Engineering, UC Santa Cruz, Santa Cruz, CA.

The Encyclopedia of DNA Elements project (ENCODE) has been producing data for over eight years to investigate DNA and RNA binding proteins, chromatin structure, transcriptional activity and DNA methylation on a variety of human and mouse tissues and cell lines primarily. As the complexity and diversity of the data grows, the tools required to organize, search and access the data in meaningful ways need to be more sophisticated. The ENCODE Data Coordination Center (DCC) has incorporated a representational state transfer application programming interface (REST API) with JSON objects to facilitate the access of ENCODE experimental metadata using a web portal. Meta-data can be accessed and data can be searched for at http://www.encodedcc.org/ using the HTTP access commands of get and post. We further expand on the access capability by allowing filtering of the meta-data with the use of search urls. This system allows external researchers to write their own interfaces to access, analyze and visualize the ENCODE data. It also facilitates the integration of the ENCODE data with other similar large-scale data sets like Epigentics Roadmap and modENCODE. Here we will present our JSON schemas, examples of the REST API and use-cases for the search functions. Our goal is for the genomics community to use the released ENCODE data available through these methods for data mining and integration.

1654M

Beyond Flat Files: Creating a web-based data API to simplify parsing and distribution of GTEx data. *T. Sullivan, D. DeLuca, K. Hadley, K. Huang, J. Nedzel, A. Segre, E. Gelfand, T. Young, G. Getz, K. Ardlie, the GTEx Consortium.* The Broad Institute, Cambridge, MA.

Large-scale genomics projects such as GTEx produce vast amounts of data, and the goals of many of these projects include creating publicly-accessible databases of genetic information. Often, to perform meaningful analysis, one must download very large text files, which must then be merged, integrated, and queried. The programming overhead the user faces during these integration and querying steps can be disproportionate to the amount of information sought by the analyst. While web pages can provide some querying capability, we argue that the heart of the solution to this problem lies in providing a web-based API. We outline here the major ways in which the GTEx API development has contributed to the ultimate goals of GTEx data distribution. These benefits fall into three categories: querying, software integration, and insulation. **Data querying:** The GTEx API provides the user with the ability to ask specific questions. For example, querying with an isoform instantly provides isoform specific expression levels across all GTEx tissues. **Integration in any software:** A major advantage that a data API has over a website is that the queries can be programmatically integrated into client systems. GTEx data can be accessed within scripts on demand. Non-GTEx web portals can query GTEx data at any time for integration and display within their webpages. **Insulation of data storage from data access:** The question of which technologies to use to store and query large datasets is non-trivial and often evolving. The API insulates the end user from these issues by providing a domain specific outward face to the data that is intuitive to use. As a result, database technologies could be modified or completely replaced without the API user having to modify existing code.

Thus far the major beneficiary of the GTEx API has been the GTEx Portal itself. Efforts to develop GTEx data visualization tools rely heavily on the GTEx API. The API delivers data efficiently for multi-tissue eQTL data, gene and isoform expression data, and sample metadata. In the future, the availability of public data APIs across large-scale genomics projects will allow researchers to easily integrate data from multiple types of studies and quickly find what measurements have been made about their genes of interest, be it in the context of a variety of normal tissues, cancer types, or disease states.

1655T

Association data in dbGaP and Minimum-Required Information for Data Sharing. *Z. Wang, M. Feolo, Y. Jin, N. Popova, M. Xu, S. Sherry.* Information Engineering Branch, National Center for Biotechnology Information, NIH, Bethesda, MD 20892.

Association statistics for each locus are important end products from studies that investigate the genomic contribution to disease/trait manifestation. Therefore since its inception, the dbGaP has been a central repository for archive and dissemination of these data. dbGaP currently provides over 3800 data sets for tests of association, linkage and somatic mutation analyses. Over 90 percent of these data sets are constructed with SNP markers, the rest are results for specific genes or whole genome copy number variations. Although these association data cover a variety of phenotypic traits, diseases and pharmacological responses, they have been provided by only a small portion of the studies in dbGaP. The dbGaP team is renewing its call to encourage researchers to upload their association results, including those that do not reach genome wide significance, to dbGaP. These results will be permanently archived and made fully accessible to authorized researchers. After potentially participant-identifiable information being removed, these data are also included to the open-access dbGaP browser and PheGenI web resource (http://www.ncbi.nlm.nih.gov/gap/phegeni) , where they are linked to NCBI resources and updated with subsequent genome builds and dbSNP annotations. To facilitate the sharing of these data, we propose a guideline for Minimum Info Required for Association Data (MIRAD). The MIRAD is composed of following data elements: 1) locus identifier (rs#, geneID, dbVarID); 2) variation summary (sample size, genotype quality and variant counts of each group) within the locus; 3) Statistics (p-value, FDR); and 4) Coding (risk) allele, and effect size. Collection of these minimal elements will aid other researchers to evaluate supporting evidence and independently verify discoveries from other research efforts in a consistent fashion. Association results are strongly encouraged for submission, even in cases where individual level data is not suitably consented for broad data sharing, as these data will allow the research community to directly use analysis results for meta-analyses, to increase statistical power, or for the development of hypotheses. This presentation will describe the current association analysis results data archived at dbGaP, and outline and discuss MIRAD.

1656S

Genome in a Bottle: So you've sequenced a genome, how well did you do? *J.M. Zook, H. Parikh, M. Salit, Genome in a Bottle Consortium.* Biosystems and Biomaterials Division, National Institute of Standards and Technology, Gaithersburg, MD.

To help clinical and research labs understand the trustworthiness of variant calls from Next Generation Sequencing, the Genome in a Bottle Consortium (www.genomeinabottle.org) is developing well-characterized whole human genomes. We recently published our methods to integrate and arbitrate between datasets from multiple sequencing technologies to develop high-confidence SNP, indel, and homozygous reference genotypes for our pilot NIST human genome Reference Material based on Coriell DNA NA12878 (Zook et al, Nat. Biotech. 2014). Since then, we compared our arbitrated multi-dataset calls to orthogonal methods developed by Illumina Platinum Genomes and Real Time Genomics using phased pedigree information to develop high-confidence calls. We found that our arbitrated calls are conservative but have exceptionally few errors and high quality homozygous reference calls, and that phased pedigree methods do particularly well in finding systematic sequencing errors such as homopolymer errors but have some systematic alignment errors that inherit properly. We combine the strengths of each of these datasets to develop the most comprehensive highly accurate benchmark set of SNP, indel, and homozygous reference calls. Genome in a Bottle is also developing similar multi-dataset arbitration and pedigree methods for structural variants. In addition, to enable users of our benchmark to assess sequencing and bioinformatics performance in a uniform manner, the Genome in a Bottle Consortium has developed a specification for performance metrics to be output by benchmarking algorithms using our Reference Materials. Finally, we will describe how labs are assessing their performance using our high-confidence SNP, indel, and homozygous reference genotypes for our pilot NIST human genome Reference Material, and our progress developing future Reference Materials.

1657M

Quantity or quality that is the question: integrative genome-wide association. *A.M. Mezlini[1,2], C. Baciu[1], S. Morrissy[1], M. Taylor[1], D. Malkin[1], A. Goldenberg[1,2].* 1) SickKids Research Institute, Toronto; Canada; 2) Computer Science, University of Toronto, Toronto, Ontario, Canada.

The causal factors behind complex genetic diseases are numerous and heterogeneous. Many studies have identified and functionally validated associated protein-coding variants. Recently, variants modulating gene expression levels and their relation to disease have attracted a lot of attention.This raises the question of how often gene products (proteins) quality versus quantity abnormalities are causing disruption in biological processes in relation to disease? Both gene products' quantity and quality might be important for any given disease thus, analyzing both data types simultaneously can provide better models of disease and characterize larger sets of patients. To address this issue we present an integrative method for combining two complementary data types: exome sequencing data as an indicator of protein quality and gene expression data which is indicative of protein quantity. Note that our approach is very different from eQTL analysis, since we are treating expression data as a predictor of a disease, rather than an outcome. Using variant harmfulness as a prior, we obtain a score per gene per patient and by exploiting gene interaction information such as protein-protein and/or regulatory networks, our method identifies sets of genes whose aberrant states potentially contribute to the disease in a large proportion of the patients. The significance of the gene set is assessed via permutations. Our analysis allows 1) identify a small set of relevant rare and common coding variants; 2) zero-in on potential regulatory aberrations. In the regulatory case, the “drivers” can be retrieved in a subsequent targeted analysis of regulatory SNVs, CNVs and epigenetic regions in a cost-effective manner for the candidate patients. Varying the proportion of the regulation-attributed vs coding variant-attributed genome aberration in a simulated setting we show that gene-based variant aggregation methods such as CAST, Calpha, RWAS fail in scenarios where less than 60% of the causal variants are coding and that our method is able to retrieve the disease mechanisms even in scenarios where as few as 20% of causes are coding variants. Our method is on par or outperforms all others in terms of precision and recall in all considered scenarios. We applied our methodology to a cohort of medulloblastoma patients and uncovered significant gene sets for different subtypes of medulloblastoma. Our results confirm and extend previously reported gene-disease associations.

1658T

Cross-species genome and epigenome visualization on WashU EpiGenome Browser. *X. Zhou[1], R.F. Lowdon[1], D. Li[1], I. Smirnov[2], Y. Cheng[3], P.A.F. Madden[4], R. Hardison[5], J. Costello[2], T. Wang[1].* 1) Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO; 2) Brain Tumor Research Center, Department of Neurosurgery, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, CA; 3) Department of Genetics, School of Medicine, Stanford University, Stanford, CA; 4) Department of Psychiatry, Washington University in St. Louis, St. Louis, MO; 5) Department of Biochemistry & Molecular Biology, Pennsylvania State University, University Park, PA.

Integrative comparison of genomes and epigenomes across mammalian species is of critical importance to human genetics research. While genome sequence comparison has pinpointed conserved elements in the human genome, patterns and rules of epigenomic conservation are not yet well defined. The same genome encodes multiple epigenomes. Epigenomes are cell type-specific, change with age, and respond to environment. Epigenetic variation and dynamics may be the foundations of great biological complexity across species and between individuals. For instance, human-mouse conserved genomic regions may have different epigenetic modification patterns, which may explain human-specific traits that are otherwise difficult to explain using the genome alone. Such an analysis requires representing each species’ epigenomic information in the context of genome sequence comparison. It thus poses a great challenge to data access and visualization. We address this challenge through a major innovation in the WashU EpiGenome Browser (http://epigenomegateway.wustl.edu/browser/). We developed a new mechanism that allows investigators to compare genomes of two species by visualizing their genome alignments at different resolution, smoothly transiting from gapped-alignment at single-base resolution, to syntenic blocks and rearranged genomes at megabase-scale. In the context of this alignment, vast epigenomic data from both species can be displayed and directly compared. Our Browser breaks the “one-reference-genome-only” status quo of genome browser tools, and contributes a powerful tool for the research community to access the cross-species genomic resources (i.e., Roadmap Epigenomics, ENCODE, modENCODE) in comparative genomics and epigenomics framework.

1659S

The Bio-LarK Patient Archive - Systematic phenotype data collection for Rare Disease Genomics. *A. Zankl[1,2], C. McNamara[3], T. Groza[3].* 1) Academic Department of Medical Genetics, Sydney Children's Hospitals Network (Westmead), Sydney, NSW, Australia; 2) Genetic Medicine, Sydney Medical School, University of Sydney, Sydney, NSW, Australia; 3) School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, Australia.

Many applications in Clinical Genomics are hindered by a lack of good phenotype data. This is particular true for rare diseases, where patients to phenotype are hard to find. Only global consortia can assemble sufficiently large patient cohorts to collect high quality phenotype data. However, a number of practical issues stymy current phenotyping efforts: 1. Phenotype data collection relies mainly on busy clinicians who are not specifically paid for the task. Phenotype data collection therefore has to be as user friendly as possible and needs to be incentivised to achieve high participation rates. 2. Medical terminology is very complex and poorly standardised. This makes combining phenotype descriptions from different sources very difficult. 3. Aggregating phenotype data from multiple institutions raises concerns regarding patient privacy, but also regarding data ownership. The Bio-LarK Patient Archive (Bio-LarK PA) was specifically built to address all these issues. Bio-LarK PA is a web-accessible database that allows clinicians to store rich phenotype data on their patients, including clinical summaries and images (photos, xrays etc.). Clinical summaries and image descriptions (e.g. xray reports) can be entered as free text or even pasted in from other sources, thus reducing the data entry burden. Clinical summaries and image descriptions are automatically annotated with suitable terms from the Human Phenotype Ontology (HPO), thus standardising the entered information and making it easily retrievable. Patient data can be kept private, shared with one or more individuals or shared with one or more groups. Sharing can be limited to the HPO annotation only or can include the full text and/or images. Shared cases can be discussed via a built-in discussion forum, for example to review a case with a group of colleagues or to obtain expert advice from a specialist overseas, a powerful incentive for clinicians to enter data into the database. Bio-LarK PA is built on Semantic Web technologies and integrates with other Bio-LarK projects such as the Bio-LarK Knowledge Base. All data is stored in RDF format and made available through a SPARQL endpoint. Other ontologies such as the Orphanet Rare Disease Ontology can be easily integrated. All Bio-LarK software is Open Source and distributed free of charge. We believe adoption of the Bio-LarK Patient Archive could leapfrog phenotype data collection for many rare diseases initiatives.

1660M

Genetic predictive modeling of diabetes based on circulating glycemic measures. The Long Life Family Study (LLFS). *A.T. Kraja[1], M.K. Wojczynski[1], J.H. Lee[2], I. Miljkovic[3], B. Thyagarajan[4], S.J. Lin[1], I.B. Borecki[1], T.T. Perls[5], K. Christensen[6,7], A. Newman[3], P. An[1], M.A. Province[1].* 1) Division of Statistical Genomics, Department Of Genetics, Center for Genome Sciences & Systems Biology, Washington University School of Medicine, St. Louis, MO, USA; 2) Department of Epidemiology, the Sergievsky Center, Columbia University Medical Center, New York, NY, USA; 3) Graduate School of Public Health, Department of Epidemiology, Center for Aging and Population Health, University of Pittsburgh, Pittsburgh, PA, USA; 4) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA; 5) Division of Geriatrics, Department of Medicine, Boston University Medical Center, Boston, MA, USA; 6) Danish Aging Research Center, Epidemiology, University of Southern Denmark, Odense, Denmark; 7) Department of Clinical Biochemistry and Pharmacology and Department of Clinical Genetics, Odense University Hospital, Odense, Denmark.

The LLFS is a cohort study of longevous families recruited from 4 centers Boston, New York, Pittsburg and Odense, Denmark. The LLFS family members have lower prevalence (7.1%) of type 2 diabetes (T2D) than the US European-ancestry population (10.2%) 20 years and older. We compare genetic associations of literature validated loci for glucose, insulin, insulin resistance and diabetes in LLFS versus top ranked results of LLFS. We hypothesize that this approach may identify genes that associate with insulin and glucose metabolism that do not operate under disease conditions. We conducted an additive genetic model genome-wide association scan (GWAS) for fasting glucose (FG), insulin (FI), HOMA-IR, HOMA-B, and glycated hemoglobin HbA1c (N=4,550). The Illumina Omni 2.5 chip was used and ~9.2M (MAF > 0.01, r^2>0.3) variants were imputed. From public databases, GWAS variants unique per gene with p-values at $p \leq 10^{-7}$ were selected (n=279). Predictive models via Graphical Trees and Support Vector Machine for T2D/pre-diabetes were constructed. Analysis of this list of candidate genes using GeneGO software revealed involvement in "Protein folding and maturation insulin processing" pathway ($-\log_{10}$(p)-FDR=10.4) and "Diabetes" ($-\log_{10}$(p)-FDR=77.9). The best LLFS SNPs of the 279 candidate genes performed poorly in LLFS in associations with five features compared with other top ranked variants present in LLFS. Best LLFS SNPs of 279 candidate genes showed respectively a Range, Mean, Median and SD $-\log_{10}$(p) for FG 0.3-6.7, 2.2, 2.2, 1.0; FI 0.3-5.6, 2.0, 2.0, 0.8; HOMA-IR 0.4-5.0, 2.0, 2.0, 0.8; HOMA-B 0.2-5.0, 2.0, 2.0, 0.9; and HbA1c 0.4-5.0, 2.0, 2.0, 0.8; compared to LLFS best SNPs (p<0.00001) for FG 5.0-7.8, 5.4, 5.2, 0.5; FI 5.0-7.7, 5.6, 5.4, 0.6; HOMA-IR 5.0-7.6, 5.5, 5.3, 0.5; HOMA-B 5.0-6.3, 5.4, 5.3, 0.3; and HbA1c 5.0-6.6, 5.5, 5.4, 0.5. Between LLFS own top ranked candidates and literature's top candidate gene list, a few overlap (LLFS own Genes/New, FG 43/37, FI 35/34, HOMA-IR 41/40, HOMA-B 33/33 and HbA1c 35/34). The new LLFS findings may point to loci important for normal glycemic metabolism conditional on age. Among new findings rs78375027 (chr6:19304147, $-\log_{10}$(p) =7.7, MAF=3.5%, n= 3,943) near *LOC100506885* associates significantly with FI. This locus has been previously associated with myocardial infarction (rs9460319, $-\log_{10}$(p) =3.5), hypertension (rs11755216, $-\log_{10}$(p) =5.1), LV mass (rs10498696, $-\log_{10}$(p) =5.7) and stroke (rs9295438, $-\log_{10}$(p) =3.8).

1661T

Changing patient behavior through comprehensive risk analysis with genomic and health data. *H. Fakhrai-Rad, P. Menon.* BaseHealth, Redwood City, CA., Select a Country.

70% of diseases are preventable, and common complex diseases and conditions like Heart Disease, Cancer, Stroke, Diabetes, Arthritis, and Obesity account for 7 in 10 deaths each year in the US. Changes in behavior can dramatically lower patient risk of developing diseases, even if one has a genetic predisposition. The first wave of consumer-focused genomics took fragmented approaches that focused strictly on genomic data, or health assessment data, and didn't provide the full view of a person's health, and ultimately did not lead to patient behavior change at high rates. We've since learned that genomic data for common complex diseases is only part of the puzzle, and should be used in concert with all of a patient's risk factors to get the most personalized and complete view of a patient's health available. How then, with the care of a physician, can genomic data best be delivered to an individual, alongside comprehensive health data, with clear ways to turn insights into enduring action? At BaseHealth, we have created a platform that leverages big data analysis to synthesize many data sets (genomic data, lab tests, lifestyle factors, family history, and other medical information) to output a comprehensive health assessment unique to each patient. Through multidimensional data modeling and meta-data analysis driven by published and curated genetic data and scientific data, health databases, and metadata analysis, we can capture complex relationships between diseases and risk factors and give physicians a better understanding of their patients' unique risks for keycommon complex diseases. The platform covers more than 40 common complex diseases that combine both modifiable and nonmodifiable risk factors and engages patients and their physicians in a highly collaborative way. With analysis provided by our platform, physicians can make meaningful recommendations that lead to patient behavior change. In this presentation, we will show how statistical modeling can meaningfully increasing a patient's chance to prevent a high-risk disease. Furthermore, we'll examine what happens when physicians use a comprehensive approach that combines all data types that have an effect on health to engage their patients to improve health and lower disease risk over time. In addition, we will share results some of the pilot programs from among the 50 physicians and 250 patients that are currently using Genophen.

1662S

Discover and access human genome sequence with new NCBI services. *S. Sherry, C. Xiao, D. Slotta, K. Rodarmer, M. Feolo, M. Kimelman, G. Godynskiy, C. O'Sullivan, E. Yaschenko.* National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD.

Next Generation Sequencing (NGS) activities currently generate a massive amount of data for use in human disease research, population genetics, and methods development. Most of this data is currently generated on research subjects who, through their informed consent, have placed restrictions on the use and redistribution of their data. This presentation will introduce new NCBI services that facilitate the discovery and computational access these data for dbGaP's current collection of 750,000 individuals.

Discovery is facilitated by "Beacons" developed in collaboration with the Global Alliance for Genomic Health (GA4GH) that identify if a specific allele is observed in any of the repository's data. NCBI has developed a public beacon at www.ncbi.nlm.nih.gov/projects/genome/beacon/ and a dbGaP beacon at https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=beacon. The public beacon returns a true/false reply, and the dbGaP beacon provides individual level details of SRA allele coverage and limited phenotypes for study approved users.

Improved methods for access have been introduced into the SRA toolkit that support encryption during delivery and at rest for secure transport and local management of dbGaP data. This is an appropriate technology for NIH-approved cloud-based computational environments that conform to the revised dbGaP data security policy. A demonstration API of these new iterator classes in Java and C++ illustrates easy programatic access to NGS data. It is available at http://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/ngs/package/ and uses example queries to 1000 genomes data stored in public SRA and dbGaP to help developers create analysis pipelines that are interactive with NCBI services using an easily configured encryption environment that is transparent to the user. These methods can significantly reduce the local storage footprint for NGS data when hardware costs are a consideration.

1663M

GenomeBrowse: A Comprehensive Community-Driven Visualization Platform for NGS data and Public Annotations. *G. Rudy, S. Gardner, G. Christensen.* Golden Helix, Bozeman, MT.

Next generation sequencing data has become a prolific tool in genetic research. As such, tools supporting that interpretation must provide an intuitive and powerful user experience that integrates research data with public resources. Current tools such as IGV focus on only local data, while web-based browsers that have public annotations are cumbersome to use with private data. We introduce Golden Helix GenomeBrowse 2, a free visualization tool for all common NGS file formats such as BAM, VCF, BED, WIG, FASTA, BigBed and BigWig while also hosting an instantly streamable public repository of annotations for 43+ species and 250 tracks. With over 3,000 registered users, GenomeBrowse has a healthy community contributing their own reference genomes and annotations to the repository. Annotations can be imported from any binary or text format into a columnar store database like format we created called Tabular Storage Format (TSF). It embeds both a genomic index for quickly visualizing genomic features at a given interval as well as different levels of aggregate binned data to enable views of large sources at genomic scale. While these features are shared with the BigBed/BigWig format, TSF also includes column-based per-field compression and lexicographical indexing to enable searching large sources like dbSNP for an RSID. Locating a RSID on the TSF converted dbSNP file (1.2GB TSF with 61 million records) takes 70ms in GenomeBrowse versus 48 seconds to locate the record with grep on the 10GB VCF or 1.6 seconds to start loading the RSID on NCBI's dbSNP website. GenomeBrowse enabled collaboration on projects by integrating with the Evernote cloud-based notebook store. You can link GenomeBrowse to your free Evernote account, open, create and edit notes on your project, and copy into your notes contextual data from your samples or public annotations. Notes can contain hyperlinks to jump back to a given genomic context, as well as custom formatted screenshots. Notes can be shared through Evernotes built-in notebook sharing, as well as through a single-button publish feature that provides a cloud-hosted hyperlink of a note. We demonstrate how GenomeBrowse can integrate public data sources such as ClinVar, COSMIC, dbNSFP 2.4 and ESP6500 to aid in the validation and interpretation of putative causal variants in a Mendelian study and document the findings in Evernote.

1664T

A mutation in TBC/LysM associated domain containing 1 (Tldc1) causes craniofacial abnormalities in mice. *R. Zeng[1], V.W Keener[2], B. Dawson[2], B. Lee[2], C.A Bacino[2], J. Shendure[3], M.J Justice[1,2].* 1) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 2) Department of Molecular Physiology and Biophysics Houston, Baylor College of Medicine, Texas, US; 3) Department of Genome Sciences, University of Washington, Seattle, Washington, US.

Craniofacial 3 (Crf3; MGI: 2671818) is a highly penetrant recessive trait that was previously isolated in a series of large-scale mouse N-ethyl-N-nitrosourea (ENU) mutagenesis screens on a mixed strain background consisting of C57BL/6J and 129S6/SvEvTAC. Crf3 homozygotes are small in size with visible craniofacial phenotypes including short snout, microcephaly and upright standing ears. We used a novel soft-tissue scanning technique to provide a non-invasive, reliable and quick method to accurately distinguish, quantify and classify the craniofacial abnormalities in mutant mice, and link them to craniofacial defects of human syndromes. The Crf3 locus has been mapped to mouse chromosome 8, linked to D8Mit92 (Z=5.7). To identify causal DNA variants, we sequenced one homozygous Crf3 sample by using whole-exome capture and massively parallel DNA sequencing. Sequencing of paired-end libraries was performed on an Illumina Hiseq 2000 system (100bp reads). Sequence reads were mapped to the mm10 (GRCm38) mouse reference genome by BWA and variants were called by GATK. Hard-filtering for the variants was based on a minimum SNP quality of 20, read depth of 5 and variant confidence of 1.5. Annotation of variants and prediction of impact of protein sequence were performed by Annovar. Variants were filtered by removing common polymorphisms, which were present in sequence from 14 mouse lines that had been isolated in our lab on the same strain background and in sequence from 18 different mouse strains from the Wellcome Trust Sanger Institute mouse SNP database. After filtering, only one homozygous variant remained that was located on mouse chromosome 8: a T to C nucleotide change in a splice donor site of Tldc1 (NM_028883: exon4: c.378+2T>C). Eight DNA samples presumed to be Crf3/Crf3 homozygotes based on phenotype were homozygous for this mutation, whereas none of seven DNA samples from unaffected, presumed heterozygous or wild type littermates were homozygous for this allele. To date, the function of human TLDC1 on HSA 16q24.1 is unknown. TLDC1 lies near the FOXF1 cluster, and is deleted in some cases of the 16q24.1 microdeletion syndrome, and lies within human duplications associated with disease. Mice with mutations in Tldc1 may prove useful as models to determine if craniofacial defects are a part of the syndrome in such patients.

1665S

Completing CpG methylation statuses in human and vertebrate genomes by integrating SMRT sequencing kinetic data. *S. Morishita[1], Y. Suzuki[1], J. Korlach[2], S. Turner[2], W. Qu[1], T. Tsukahara[3], H. Yurino[1], J. Yoshimura[1], Y. Takahashi[4], J. Mitsui[4], S. Tsuji[4], H. Takeda[3].* 1) Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Japan; 2) Pacific Biosciences, Menlo Park, CA, USA; 3) Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan; 4) Department of Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

Determining the methylation information for regions with low or high copy numbers using second-generation sequencing is challenging because the read length is insufficient, especially when the repetitive regions are long and nearly identical to each other. To resolve these problems, single-molecule real time (SMRT) sequencing shows promise because it is not vulnerable to GC bias, it has long read lengths, and its kinetic information is sensitive to DNA modifications. However, raw kinetic information at a single CpG site contains some noise, and characterizing the DNA methylation for large size genomes demands prohibiting coverage of SMRT reads. Since hypo-/hypermethylated CpG dinucleotides are often contiguous over a long span in vertebrate genomes, we propose a novel algorithm that combines the kinetic information for neighboring CpG sites and increases the confidence in identifying the methylation statuses of those sites when they are correlated. The sensitivity and specificity of our algorithm were both of >90% for the genome of an inbred medaka (*Oryzias latipes*) strain within a practical read coverage of <30-fold. With this method, we newly characterized the methylation status of repetitive elements (*e.g.,* the occurrences of ~6-kb-long interspersed nuclear elements (LINEs)), regions of duplicated genes (*e.g.,* HIST2 cluster) in the human genome, and nearly identical living transposons of length 4682 bp in the medaka genome, which were difficult to observe using bisulfite-treated short reads.

1666M

Sequencing of 50 rhesus macaques facilitates identification of new genetic models of human disease. *G. Fawcett[1], D. Rio Deiros[1], R.A. Harris[2], M. Raveendran[1], B. Ferguson[3], Z. Johnson[4], S. Kanthaswamy[5,6,7], E.J. Vallender[8], N.H. Kalin[9,10,11], R.W. Wiseman[12,13], D.M. Muzny[1], R.A. Gibbs[1], J. Rogers[1].* 1) Dept Mol & Human Gen, Baylor College Med- HGSC, Houston, TX; 2) Dept. Mol. & human Gen, Baylor College Med, Houston, TX; 3) Division of Neurosciences, Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, OR; 4) Yerkes National Primate Research Center, Emory University, Atlanta, GA; 5) Molecular Anthropology Lab, Department of Anthropology, UC Davis, Davis,CA; 6) California National Primate Research Center, UC Davis, Davis, CA; 7) Department of Environmental Toxicology, UC Davis, Davis, CA; 8) New England Primate Research Center, Harvard Medical School, Southborough, MA; 9) Depts. of Psychology and Psychiatry, Univ. of Wisconsin-Madison, Madison, WI; 10) Waisman Laboratory for Brain Imaging and Behavior, Univ. of Wisconsin- Madison, Madison, WI; 11) HealthEmotions Research Institute, University of Wisconsin-Madison, Madison, WI; 12) Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, WI; 13) Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI.

Rhesus macaques (*Macaca mulatta*) are the most commonly used non-human primate model of disease, and are particularly important as models of psychiatric, neurological and infectious diseases. Therefore, characterizing genetic variation within research colonies of rhesus monkeys will advance the study of a variety of different diseases. We sequenced the exomes of 50 unrelated rhesus macaques from 5 NIH-funded primate research centers. Single nucleotide variants (SNVs) are defined as variants with a minimum of 4 reads covering each position with a minimum quality score of 30, whereas validated SNPs also have both alleles observed in at least two animals. Prior studies suggest rhesus macaques have greater genetic diversity overall than humans. Yuan, et al (2012) reported that while rhesus macaques exhibited ~3 SNPs/Kb, the majority of these variants were in intergenic regions, and that coding regions were similarly polymorphic in rhesus macaques and in humans. Estimates of human SNP density in predicted exonic regions (Bainbridge, et al 2011) show 1.45 SNPs/Kb in exons. Our data is similar to the human data with an exonic SNP density of 1.15 SNPs/Kb. However, our SNP density in off-target intronic regions is significantly greater in macaques (4.8 SNPs/Kb). Our results are consistent with Yuan et al's expectation of greater non-coding but similar coding variation density in macaques relative to humans. This result is unsurprising since coding variants are likely to be under strict negative selective pressure. We also sought to identify specific new potential models of disease. Among the 50 exomes, we identified: 133 transcription stops gained, 17 transcription starts lost, 225 splice sites changed, and 11,768 non-synonymous SNPs. We observed variants in genes from many different ontological categories (DAVID analysis) including neurologically relevant and immune system genes. For example, there was one non-synonymous variant in *CR2*, four in *DNAH11*, one in *ATP6AP2* and two in *FGFR3. CR2* is important for immune system function, and *DNAH11* and *ATP6AP2* are neurologically functional genes, while *FGFR3* is associated with a variety of morphological disorders and cancer. Our results indicate that this species carries functionally significant mutations in many genes relevant to human disease risk.

1667T

Identifying mouse models related to human disease. *S. Rockwood[1], S.M. Bello[2], J.T. Eppig[2], J.E. Richardson[2], M. Sasner[1], C.L. Smith[2], L. Donahue[1].* 1) Genetic Resource Science, The Jackson Laboratory, Bar Harbor, ME; 2) Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME.

Access to model organisms is of paramount importance to advancing the research objectives of the bio-medical research community. To facilitate this access, the Mouse Repository at The Jackson Laboratory (JAX) has served as a centralized resource for the distribution, cryopreservation and development of high health status mouse models for over 50 years. Hundreds of new strains are added each year to one of the largest collections of characterized mouse strains available. The increasing magnitude and diversity of current day mouse resources present a challenge to researchers who attempt to identify and obtain mouse models most appropriate for their work. In addition to newly designed search interfaces at the JAX website (www.jax.org), other web-based search resources are available to assist with this task. The Mouse Genome Informatics group at JAX maintains two of the most useful: the International Mouse Strain Registry (IMSR) and the Human/Mouse: Disease Connection site (HMDC). The IMSR (www.findmice.org) is a searchable online database of mouse strains and mutant ES cell lines available worldwide. The HMDC (www.diseasemodel.org) provides a seamless human-to-mouse data traversal, enabling clinical and translational researchers to take advantage of the wealth of data and annotations from mouse models; as well as allowing mouse researchers to connect their findings directly to genetic associations reported in human disease. This poster provides several examples illustrating how these resources can be used to quickly identify and obtain mouse models related to human disease research applications. (also see related posters by Smith CL et al and Eppig JT et al). Donating a strain to the Jax Mouse Repository is an easy way to fulfill the NIH's requirements for sharing mice. Researchers wishing to have strains considered for inclusion in the Repository are encouraged to submit their strains at: www.jax.org/donate-a-mouse The JAX Mouse Repository is supported by NIH, The Howard Hughes Medical Institute, The Ellison Medical Foundation and several private charitable foundations.

1668S

FlyNet: a genome-wide functional gene network for Drosophila melanogaster as a platform to explicate human GWAS gene candidates. *J. Shin, C. Kim, H. Shim, H. Kim, S. Hwang, T. Lee, I. Lee.* Dept. of Biotechnology, Yonsei University, Seoul, South Korea.

Drosophila melanogaster (fruit fly) is a historically long-established model organism and its availability encompasses almost every field in modern biology, such as genetics, development and neurobiology. Moreover, D. melanogaster can provide many human disease models including diabetes and neurological disorders. Recent progress of genome-wide association study (GWAS) in humans has elucidated thousandish disease-associated genetic elements. Nevertheless, model organisms are still indispensable for functional validation and characterization of those candidate disease genes. In this study, we present a genome-wide functional gene network for D. melanogaster, FlyNet, and demonstrate its utility in pathway interpretation and functional characterization of the candidate genes from human GWAS, facilitating genetic dissection of human diseases in D. melanogaster.

1669M

Benchmarking of isoform quantification tools for GTEx using long read technologies. *D.S. DeLuca[1], A. Battle[2], T. Sullivan[1], T. Lappalainen[3], M. Sammeth[4], S. Montgomery[2], F. Reverter[4], T. Nance[2], E. Tsang[2], E. Pantaleo[5], J. Maller[1], C. Brown[6], B. Engelhardt[7], G. Getz[1], K. Ardlie[1], The GTEx Consortium.* 1) The Broad Institute, Cambridge, MA; 2) Stanford University, Stanford, CA; 3) New York Genome Center, and Department of Systems Biology, Columbia University, NY; 4) Center for Genomic Regulation, Barcelona, Spain; 5) University of Chicago, IL; 6) University of Pennsylvania, PA; 7) Duke University, NC.

One of the hallmark benefits of RNA-seq over traditional gene expression quantification methods is the ability to capture the sequences of exon-exon junctions, which are indicative of the transcript isoform structure present in the samples. Beyond qualitatively measuring the presence of alternative isoforms, methods exist to infer isoform abundance. However, assessing the accuracy of isoform quantification methods remains extremely challenging due to a lack of ground truth. Orthogonal technologies such as qtPCR or NanoString have been used for validation but any approach that targets only portions of each isoform is limited in its applicability to validate the range of alternative isoforms produced by RNA-seq. Here we evaluate a panel of algorithms for the quantification of isoforms, including Flux, Cufflinks, eXpress, RSEM and Sailfish in combination with multi-tissue validation experiments from the Genotype Tissue Expression (GTEx) project. As orthogonal technologies we are employing long fragment, long read Illumina RNA-seq as well as deep 1-2 kb read PacBio sequencing. We have generated 66 TruSeq libraries from a panel of 13 different GTEx tissues with an average fragment length of 400 base pairs. These libraries were sequenced to 100 million 250bp paired-end reads. The tissue panel consists of adipose, artery, cerebellum, frontal cortex, lymphocytes, heart, lung, muscle, pancreas, testis, thyroid and whole blood. The dataset to be evaluated consists of the standard GTEx sequencing runs (2x76bp short fragment TruSeq) derived from the same samples, as well as a subsampled version of the evaluation set in which the read lengths have been truncated to 76bp to simulate standard runs. The opportunity provided by the multi-tissue GTEx sample bank is critical to this evaluation considering that many isoforms are expressed in a tissue specific manner. These data will provide insight into the strengths and limitations of current algorithms as well as into the process of isoform specific expression in general.

1670T

Droplet Digital PCR: A new tool for quantitative analysis of alternatively spliced mRNAs and pre-mRNA processing. *G. Karlin-Neumann, C. Litterst, D. Maar, S. Marrs, S. Tzonev, G. McDermott.* Digital Biology Center, Bio-Rad Laboratories, 5731 W. Las Positas, Pleasanton, CA.

Understanding the transcriptome's regulatory complexities including co- and post-transcriptional modifications such as splicing (Bentley, 2014), can be of key importance. Examples range from defining the post-transcriptional regulatory role of Cytoplasmic Intron-Sequence Retaining Transcripts (CIRTs) in the firing of hippocampal neurons and in platelet activation (Khaladkar et al, 2013), to assessing the effects of exon-skipping drugs in mitigating disease in Duchenne muscular dystrophy (Arechavala-Gomeza et al, 2012). Here we demonstrate the ability of droplet digital PCR's (ddPCR's) EvaGreen detection capability to distinguish and quantify intron-containing RNA intermediates and splice variants. As an example of differences in pre-mRNA processing, spliced and unspliced *GUSB* transcripts were quantified in total RNA from 20 human tissues. Among these tissues, testes showed the greatest fraction of unspliced transcript (~32%) whereas skeletal muscle showed the lowest (~1.3%), with testes having ~3X the concentration of spliced RNA (~1900 copies/ 4 ng total RNA) as skeletal muscle (660 copies/ 4 ng total RNA). Similar results were seen for other genes, e.g. *GAPDH* and *ACTB*. Simultaneous detection and quantification of exon-skipping was demonstrated in a similar fashion for three genes with numerous brain isoforms: *CAMTA1*, *TPM3* and *ABLIM1*. Most tissues tested showed one log (e.g. *ABLIM1*) to two logs (e.g. *TPM3*) more of the unskipped variant (containing the intervening exon) than the skipped; however, some tissues such as brain, heart and thyroid showed nearly a log-fold greater abundance of the skipped variant than in most other tissues. Both of these study results were confirmed with Taqman 5' nuclease probe assays using the same gene primer pairs as in the EvaGreen analyses. Thus, we demonstrate that ddPCR enables rapid and sensitive, multiplexed quantification of RNA processing intermediates and splicing isoforms with or without probe-based assays.

1671S

Approach for limited cell ChIP-Seq on a semiconductor-based sequencing platform. *S. Ghosh[1], K. Giorda[1], R. Marcus[2], M. Taylor[1], E. Farias-Hesson[1,2], D. Bluestein[2], G. Meredith[1], S. Leach[2], B. O'Conner[2].* 1) Thermo Fisher Scientific, South San Francisco, CA 94080; 2) National Jewish Health, Center for Genes, Environment & Health, 1400 Jackson St., Denver, CO 80206.

Dendritic cell (DC) lineages coordinate immune system activity through functional specialization. Irf4, a transcription factor, is required for CD11b+ DC lineage development from bone marrow stem cells and has been implicated in multiple inflammatory diseases such as asthma. The epigenetic consequences of immune specialization in CD11b+ DCs and relation to inflammatory diseases remain largely unexplored partly due to the difficulty of using highly purified, and typically, limited populations of cells in ChIP-seq (chromatin immunoprecipitation followed by sequencing) assays.

A robust ChIP protocol - using an input control - was developed using limited amounts of K562 cells. ChIP-seq* on the Ion Proton™ - followed by analysis using MACS2 - of the transcription factor CTCF and histone modification marks (H3K9me3-methylation, H3K27ac-acetylation) strongly correlate with ENCODE datasets - with confirmation rates of 69%(CTCF), 73%(H3K27ac), 89%(H3K9me3). Furthermore, CTCF Proton™ data show significant enrichment for insulator domains at 41%; the moderate presence of DNA methylation concomitantly with H3K27me3 - a signature for polycomb repression - is attested to by the ENCODE-Broad ChromHMM. Also, 55% and 35% of the acetylation marks enrich for the enhancers and promoters respectively. These results are based on chromatin derived from 1 million(M) cells, making it viable for generating data from limited number of primary cells. This is in contrast to the 10M cells recommended by the ENCODE consortium.

This methodology was used to compare Irf4 genomic binding sites generated from flow-sorted population of 1, 3, 5, and 25M CD11b+ lineage murine DCs. We observed comparable Irf4 ChIP-seq results from 5 versus 25M cells. Moreover, results indicate that as low as 5M flow-sorted cells can be used to acquire high quality(FDR: 10^{-19}) ChIP-seq data. We identified genomic Irf4 binding sites proximal to genes, whose activity is consistent with CD11b+ DC lineage activity and/or known to contribute to inflammatory disease. To examine Irf4 functional regulation of the identified gene targets, we performed RNA-seq analysis with CD11b+ DCs and a related lineage, CD103+ DCs. Integrating expression analysis with ChIP-seq indicates a unique CD11b+ DC gene expression program concordant with Irf4 loci association in comparison to CD103+ DC. **For Research Use Only*..

1672M

GEM.app: using hadoop to empower the revolution of large-scale collaborative analysis and data-sharing in the genomic age. *M. Gonzalez, R. Acosta, R. Schule, S. Zuchner.* Human Genetics and Genomics, Unviersity of Miami Miller School of Medicine, Miami, FL.

Next-generation sequencing technologies have revolutionized how the human genetics research community explores genetic variation. The use of NGS has led to the identification of numerous genes, that when mutated, can give rise to disease phenotypes. However, due to high rates of genetic/clinical heterogeneity and the abundance of rare variation, it is now evident that large cohorts will be necessary in order to achieve enough power to detect biological/genetic pathways that are enriched in disease processes. The management and interpretation of such data is vital to continue the progress of human genomics research. To this end, we have developed, GEnomes Management Application (GEM.app), in order to make large-scale genomic analysis and data sharing accessible to users of all technical backgrounds. GEM.app uses the combination of a highly distributed computing cluster and a user friendly graphical web interface to provide users the capabilities to run real time analyses of the ~6,000 exomes and ~900 genomes currently uploaded in GEM.app. In addition, we have developed tools within GEM.app to facilitate the creation of collaborative data-sharing groups, where researchers are in full control to share data-sets instantly. Within these collaborative data-sharing groups, researchers from around the world can contribute data that can be concurrently analyzed by all users in the data-sharing network. GEM.app utilizes the latest variant annotations (conservation, amino acid substitution predictions, ClinVar, EVS, etc.) to allow flexible and powerful analysis options. We provide five major analysis modules: 1) Variants within Families - basic filtering of variants, 2) Genes Across Families - filtering for genes that are recurrently mutated, 3) Allele Sharing - filtering for alleles that are enriched, 4) Cancer Analysis - quickly analyze tumor/normal pairs, and 5) PathFinder - hypothesis driven analyses using Gene Ontology and various pathway databases to identify pathways enriched for mutations. Over 60 genes have been identified by GEM.app users, including DDHD1, DDHD2, GBA2, CYP2U1, RTN2, B4GALNT1, BICD2, FBXO38, ANKRD11, PNPLA6, SARM1, DNAJC5, etc. Combining powerful distributed IT technologies with bioinformatics will lead to efficient tools to explore the complexities of human disease. As we see consortia and collaborative networks adopting GEM.app for analysis, we observe a radical reduction in time to discovery and increasingly complex questions resolved.

1673T

A Data Driven Approach to Precision Medicine. *P. Lum.* Ayasdi, Menlo Park, CA.

The pharmaceutical industry is currently in the midst of a large data problem: there is far too much data to be analyzed in order to promote a viable compound, and there are too many different types of data to fuse in order to get started. There is an urgent need for software that can analyze clinical and genomic data and derive valuable insights in one place. Ayasdi's application, Ayasdi Cure makes it easy for biologists, chemists, and physicians to consume and derive value from complex data by revealing patterns and subgroups in a visualization-based user interface. This product is built on the Ayasdi Platform which leverages hundreds of machine learning algorithms through the framework of Topological Data Analysis (TDA) to deliver operational insights from data in minutes. The product contains specific features that are best suited to pharmaceutical discovery in research including biological pathways, a gene variant viewer, and a compound visualization tool. However, this application represents a larger theme in the pharmaceutical industry: the movement towards data-driven approaches to precision medicine. With Ayasdi Cure, it is easy to incorporate patient information with chemical information to create more effective and targeted treatments, truly turning data into therapies. For example,a top 5 pharma company used Ayasdi Cure to salvage a chemical compound during clinical trials which was proven to be effective for a specific sub-population. This talk will focus on the trends in data-driven approaches to drug discovery, the types of data ecosystems that need to exist for effective analyses, and a few case studies that illustrate how Ayasdi Cure has helped pharmaceutical companies derive insights from complex data.

1674S

Improved Small RNA Library Preparation Workflows for Next-Generation Sequencing. *S. Shore[1], N. Paul[1], M. Salcedo[1], G. Zon[1], C. Olsen[2], K. Qaadri[2].* 1) TriLink BioTechnologies, San Diego, CA; 2) Biomatters, Inc. Newark, NJ.

MicroRNAs (miRNAs) are a group of small noncoding RNAs that regulate transcriptional and post-transcriptional gene expression. They regulate gene expression by binding to the 3'-untranslated region (3'UTR) of mRNAs, which leads to mRNA degradation or destabilization. Studies have demonstrated the effect of miRNAs on methylation machinery and protein expression in disease progression via epigenetic mechanisms. With the improvement of sequencing technologies, miRNAs are increasingly being characterized using deep sequencing. One of the challenges of deep sequencing small RNAs (small RNA-Seq) is in the sample preparation workflow. This workflow involves the ligation of fixed sequences (adaptors) onto the 5' and 3' ends of the starting RNA library. These adapters are prone to mispairing with one another without a segment of the library in between, resulting in adapter-dimer formation. Most library workflows use affinity capture to remove these adapter dimers, which is inefficient due to the close size similarity between adapter dimers and adapter-tagged small RNA libraries. Alternatively, a gel purification step can be used, which in consequence, can deplete low abundance sequences from the starting library. Furthermore, adapter-dimers can predominate when input library concentrations are low, thus suppressing formation of the adapter-tagged library. In this poster, we present a novel approach to small RNA library sample preparation using chemically modified adapters to disfavor adapter-adapter ligation while allowing for efficient joining of adapters onto the 5' and 3' ends of the library. Using this technology, we demonstrate improvements in specificity and yield for small RNA library preparation.

1675M

Enhanced fetal aneuploidy detection using hardware accelerated alignment. *M. Sykes[1], C. Roddey[2], M. Ruehle[2], R. McMillen[2], P. Whitley[1].* 1) Sequenom Inc., San Diego, CA; 2) Edico Genome Inc., La Jolla, CA.

Noninvasive prenatal testing can be performed by massively parallel sequencing of DNA from maternal plasma. This method has been shown effective in the detection of fetal aneuploidies of chromosomes 13, 18, 21 and the sex chromosomes. Accurate classification of these aneuploidies requires, in part, alignment of sequencing reads to the human genome, calculation of chromosome fractions based on these alignments and calculation of z-scores for each chromosome based on these fractions. The success of these steps relies upon the choice of aligner and algorithm used to determine the chromosome fractions.

Here we present reclassification of a dataset of 1269 samples previously analyzed using bowtie 2 as the aligner. In this study alignments are generated by the DRAGEN processor, a hardware-accelerated sequencing analysis system developed by Edico Genome. We report systematic differences between the two aligners but equivalent performance in terms of chromosome fraction variability and thus chromosome quantification.

Both the bowtie 2 and DRAGEN based analyses successfully identified all known T13, T18 and T21 cases in the dataset. The sensitivity and specificity were both > 99.9% in each classification. At the same time the DRAGEN system provides speed increases of greater than thirty-fold relative to bowtie 2 running with 6 threads on a 3.5 GHz Xeon CPU, allowing a single computer to replace the efforts of a small cluster.

These results demonstrate that the classification algorithm for fetal aneuploidy is robust and resistant to localized changes in the alignment profile. Furthermore the DRAGEN system provides equivalent performance to bowtie 2 with a significant increase in speed.

1676T

Platform comparison between Ion Proton and Illumina HiSeq 2500 on a 759-gene disease panel across 248 samples. *A.V. Uzilov[1,2], H. Shah[1,2], M. Mahajan[1,2], D. Starcevic[1,2], R. Sebra[1,2], R. Chen[1,2], E. Schadt[1,2].* 1) Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Targeted panels allow making next-generation sequencing libraries enriched for genomic regions of interest, providing more cost-effective sequencing for the needed depth. Commercial panels and custom panel design services exist for both Ion Proton and Illumina HiSeq instruments. However, little publically available comparative data exists to guide choice of appropriate platform for large sequencing projects. We sequenced blood-extracted DNA from 248 individuals on Ion Proton and Illumina HiSeq 2500 using a custom targeted panel covering 4413 genomic regions and full exons from 759 genes selected for their genetic disease relevance. Agilent's SureSelect hybridization-capture chemistry (covering 2.8 Mb) was used for Illumina and amplicon-based AmpliSeq (covering 4.4 Mb) was used for Proton. We look at coverage uniformity and variant concordance to compare the two panel versions in their 2.6 Mb overlap. This is the largest comparative analysis between the two chemistry/platform combinations both in terms of sample volume and panel size. 18.1% of variant calls are unique to the Illumina panel, 6.6% are unique to the Proton panel, 1.0% are conflicting calls at the same site, and 74.5% are concordant (median percent across all samples). We observe that Proton concordance depends on sequencing depth at depths seen in our study (median 118X for Illumina and 100X for Proton). As Proton depth increases, concordance increases and "unique in Illumina" calls decrease, suggesting that raising Proton depth leads it to make more calls found by Illumina. However, there is no improved concordance at higher Illumina coverage, suggesting we are operating at or above optimal depth there. To find optimal depth for Proton, we sequenced NA12878 on two Proton instruments to high depth (one barcode per 318-type chip, 972X and 766X median coverage). We show how subsampling reads impacts variant calling accuracy versus similar high-depth replicate runs on Illumina (323X and 289X median coverage). At highest depths, sensitivity for GIAB gold-standard set of known variants was 96.6% for both Illumina replicates and 89.0% and 89.4% for the two Ion replicates. We describe factors influencing accuracy, concordance, and callability and discuss if they can be controlled by adjusting depth or variant caller settings or whether they are a property of the panel or sequencer. Our findings will be useful to investigators when selecting technology appropriate for their NGS study.

1677S

Sparse sufficient dimension reduction and matrix subset selection methods for big image data analysis in cancer. *N. Lin, J. Jiang, S. Guo, M. Xiong.* University of Texas School of Public Health, Houston, TX.

With the advancements in modern biomedical engineering, biomedical imaging has become one of the most reliable tools for disease diagnostics and the evaluations of the progress of different treatments. A Key issue for image analysis is high dimension data reduction and feature selection. The most widely used dimension reduction and feature selection methods have two serious limitations. One limitation is that these methods often use unsupervised dimension reduction method, which selects dimensions without taking information on class labels or disease status into account. The reduced data will then lose important class information that is hidden in the original data. The second limitation is that the algorithms for searching features are unable to deal with large number of features. To overcome these limitations, we introduce concepts of sufficient dimension reduction (SDR) and coordinate hypothesis which project the original high dimensional data to very low dimensional space while preserving all information on clinical outcome information, and formulate feature selection problem as a subset selection problem for matrices. We develop matrix approximation theories and error estimation methods as powerful tools for optimal image feature matrix column selection. We develop a randomized algorithm that is designed to select image features with provable guarantees that image reconstruction by selected image features can reach prespecified high accuracy. To further reduce dimensions, we extend one dimensional principle component analysis to two dimensional principle component analysis (2DFPCA) and use 2DFPCA scores to represent image data. The proposed method is applied to 188 kidney cancer histology images from different individuals. Of these 188 images, 121 of them are confirmed to have kidney cancer and the rest of them are served as the control in our study. The randomized based sparse SDR method was applied to 2DFPCA score to select features. We used support vector machine (SVM) as classifier which took selected features as input. By 5-fold cross validation, the average classification accuracy in the test dataset was 89.3% and in the training dataset was 96.4%. If we use only 2DFPCA score of the original image data as features and the same SVM as the classifier, we only can reach average accuracy 65% in the test dataset and 67.2% in the training dataset. This strongly demonstrates that the proposed method substantially outperforms other methods.

1678M

A Genomics Analysis Pipeline for Cloud Computing. *R.J. Mashl[1], K. Ye[1], N. Nutter[1], D.C. Koboldt[1], D.E. Larson[1], K. Chen[2], L. Ding[1].* 1) The Genome Institute, School of Medicine, Washington University, St. Louis, MO; 2) Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX.

As medicine continues to enter the era of personalized medicine, the amount of raw genomics data is expected to increase substantially. The traditional analysis model, wherein data is downloaded from secured internet sites, e.g., data coordination centers, to local storage for processing, is likely to result in an accumuation of unprocessed data if data transfer rates, storage capacities, and computing rates cannot keep apace. Individual investigators or those working in small collaborative groups may face additional logistical challenges in terms of cost, compute resources, data storage and security, and computer management expertise. One possibility for addressing these problems is to use cloud computing, which provides access to large amounts of storage and computation as a service. Technologies that enable data analysis to be performed at the data storage site, i.e., those that "bring the tools to the cloud," are therefore beneficial. In this study, we describe an easy-to-use web interface for configuring a genomics analysis pipeline featuring software packages that are among the state-of-the-art in variant calling tools for next-generation sequencing methods, namely, VarScan, BreakDancer, and Pindel. This combination of tools allows users to analyze for germline and somatic single-nucleotide polymorphisms (SNPs) and insertions and deletions (indels), copy-number variations, and structural variations that include translocations and inversions. The configured computations can be deployed either on a physical compute cluster or on cloud/virtual compute clusters. The interface utilizes the StarCluster cluster-computing toolkit for interacting with Amazon's Elastic Compute Cloud. The software is available through the Turnkey Variant Analysis Project (TVAP) web site at http://tvap.genome.wustl.edu. Additional information is available at the ASHG 2014 Annual Meeting workshop "iSeqTools to Demistify the Cloud and Genomics Analysis for Researchers Seeking Ways to Analyze High-Throughput DNA Sequencing Data." This work is supported by the National Human Genome Research Institute grant U01HG006517 (to L.D.).

1679T

A Comprehensive Bioinformatics and Data Management Platform to Enable High Powered Genomic Discovery. *J. Kaufman[1], F.M. De La Vega[1], T. Schlumpberger[1], S. Young[2], J. Wiley[2], A. Patel[2], M. Pae[2], R. Hayek[2].* 1) Annai Systems Inc., Burlingame, CA; 2) Annai Systems Inc., Carlsbad, CA.

As the amount of sequencing data that is available in the public domain continues to increase, and the cost of a whole human genome sequence declines, the desire to perform large-scale analyses has grown considerably. By combining large public data sets with one's own data, the statistical power of a study can be increased in a cost effective, streamlined manner. However, the availability of a compute, storage and data-sharing platform that can facilitate these types of "big data" analyses is relatively limited. The options that are available for these studies are typically lacking in one key aspect or another such that identifying an optimal solution often involves some level of pulling together disparate, sub-optimal components. Here we describe a singular environment that includes high performance computing, data storage, data sharing and access to high value public data sets. Annai ShareSeq™ Genomic Resource is a bioinformatics and data management platform that provides a highly scalable, secure and computationally powerful solution for genomic researchers, in particular, those interested in cancer related data. We will describe the various aspects of ShareSeq as well as present examples of how this resource can be used for data analysis. A workflow manager can be used to configure an analysis pipeline by taking advantage of a variety of analysis tools that are resident in ShareSeq. Computing resources can be accessed in an on-demand manner by initiating a virtual machine that can be scaled appropriately, depending on the analysis job size. Other robust tools can be used to transfer sequencing files, create a storage repository to house them, and later, leverage metadata that has been automatically indexed, for flexible querying of the collection of files. These capabilities, in combination with easy access to large data sets, create an invaluable resource for researchers and clinicians that are using genomic data.

1680S

Validation of a Series of Genomic StripAssays® to Salivary DNA Collection Using the DNA•SAL™ Device. *P.D. Slowey[1], C. Oberkanins[2], A. Berndt[2], G.A. Thomas[1].* 1) Oasis Diagnostics Corporation, Vancouver, WA US; 2) ViennaLab Diagnostics Vienna, Austria.

PCR hybridization based StripAssays® [ViennaLab Diagnostics] provide a rapid convenient method for research and diagnostic detection of mutations associated with cancer, genetic disorders, genetic predispositions and pharmacogenomics. The assays consist of four steps: gDNA isolation, multiplex PCR, hybridization, and detection. Using DNA isolated from blood, conventional StripAssays® detect a combination of wild type and mutant probes for the identification of a variety of genetic disorders and predispositions including Gaucher disease, thalassemias, CVD and others. While blood is a common specimen matrix, invasive collection limits participant compliance particularly with small children and its utility for remote point of care applications. Due to a number of factors including the non-invasive nature of collection, ease of use and subject compliance, saliva has become an increasingly important alternative to blood for nucleic acid based research and diagnostics. While saliva is simple to collect, the proportion of epithelial cells (buccal,) leukocytes and non-human nucleic acid from microbial sources varies between individuals and can confound analysis and restrict adoption of saliva based testing. In the present study, the performance of gDNA extracted from saliva collected using the DNA•SAL™ Salivary DNA Collection Device [Oasis Diagnostics®, Vancouver WA US] and isolated using the Mini•SAL™ DNA Isolation Kit [Oasis] was evaluated in 5 different commercially available PCR assays from ViennaLab: α and β thalassemia (4-130, 4-160), Gaucher disease (4-260), CVD T (4-360) and CVD A (4-370) chosen from the ViennaLab portfolio as tests with the highest level of complexity and sensitivity to DNA quality and quantity. For each test, 5 μl of DNA isolated at a concentration from 2-19 ng/μL using the Mini•SAL™ protocol was used in the StripAssay®. Bands on the StripAssay® test strips were inspected visually and using ViennaLab's own proprietary software package [Evaluator™]. All tests yielded results directly comparable to results obtained using DNA isolated from blood samples. The data indicate that gDNA extracted from saliva via the Mini•SAL™ DNA Isolation Kit following collection using the DNA•SAL™ device is a robust alternative to blood DNA when used with the ViennaLab StripAssays.® Together they provide a simple-to-use system with the potential for truly realizing point of care genetic testing.

1681M

Direct to PCR Genomic Analysis Using Saliva Derived Samples. *G.A. Thomas[1], C. Oberkanins[2], A. Berndt[2], P.D. Slowey[1].* 1) Oasis Diagnostics Corporation, Vancouver, WA; 2) ViennaLab Diagnostics, Vienna, Austria.

Saliva has become an increasingly important sample matrix as an alternative to blood for nucleic acid based research and diagnostics utilizing PCR, hybridization, microarrays and next-gen sequencing due to the non-invasive nature of the collection, ease of use and participant compliance. While saliva may be readily collected, the proportion of epithelial cells (buccal,) leukocytes and non-human nucleic acid from microbial sources varies between individuals and can confound analysis. Recently, the utility of DNA extracted from saliva has been validated as a robust alternative to blood for use in PCR based strip assays. The StripAssay® technology [ViennaLab] consisting of four steps: gDNA isolation, multiplex PCR, hybridization and detection provides a rapid convenient method for research and diagnostic detection of mutations associated with cancer, genetic disorders and pharmacogenomics. In the study, DNA extracted from saliva using the DNA•SAL™ Salivary DNA Collection Device [Oasis Diagnostics®] and isolated using the companion Mini•SAL™ isolation kit [Oasis] performed identically to blood DNA in 5 different multi-probe StripAssays®. To test the suitability of DNA•SAL™ collected saliva samples for direct PCR analysis without DNA isolation, the performance of independently collected saliva samples in two genetic disorder tests from ViennaLab [FV-PTH-MTHFR (4-260) and Hemochromatosis (4-210)], was evaluated as a function of input sample dilution. Both StripAssays® examine multiple wildtype and mutant probes (4-260: factor V G1691A, prothrombin G20210A and MTHFR C677T. 4-210: HFE H63D, S65C and C282Y). Normal test results were obtained for all samples when simply diluted with water within the range of 1:5 to 1:100. Half of undiluted samples failed presumably due to inhibition of PCR activity, while all samples failed at a dilution of 1:500 presumably due to excessive dilution of the target. The data demonstrates that saliva collected using the DNA•SAL™ Device when diluted within the range of 1:5 to 1:100 can be used directly in PCR assays without substantial signal loss. The elimination of the requirement for DNA isolation when using the DNA•SAL™ collection tool in tandem with PCR hybridization StripAssays® greatly facilitates the potential for truly realizing remote point of care genetic testing.

1682T

Spatially Encoded Assays. *M.S. Chee, J.A. Altin, W. Delport, B. Min, P. Capek, E. He, C.M. Dambacher, D.A. Routenberg.* Prognosys Biosciences, Inc., San Diego, CA.

Detailed information on spatial patterns of gene expression in complex tissues is needed to enable new insights into the molecular differences between normal and disease states. Although in situ hybridization has excellent resolution, it is limited in multiplexing. Conversely, microarrays and RNA-seq multiplex well but are not designed to provide spatial information. Laser capture microdissection enables highly multiplexed analysis of well-defined sites in a tissue section but is relatively low throughput and laborious. We developed a scalable assay system that is both spatially resolved and enables high multiplexing, overcoming limitations of the methods listed above. The assay involves binding of probes to target RNAs in the sample, and encoding the probes by spatially-addressed delivery of sequence-encoded adapter molecules. The sequences of the resulting assay products and their frequency provide both abundance and location data for each target RNA in the sample. We performed the spatial addressing using microfluidic devices that we designed to define a rectangular array of assay sites. Assay products were collected and sequenced in bulk using an Illumina next-generation DNA sequencer. We used the system to map the expression of 134 unique target RNA sequences from 69 genes across an array of 256 positions on formalin-fixed paraffin-embedded (FFPE) pancreas and liver samples. We constructed virtual images of the FFPE sections based on the data, revealing local regions of high gene expression that were well correlated with immunofluorescence data. For example, we mapped regions of high insulin mRNA abundance to islets of Langerhans in pancreas by correlation with insulin protein abundance determined by immunofluorescence. The technology is flexible. We have used the devices to analyze both individual sections of tissue and tissue microarrays. The method currently achieves an effective pixel resolution of ~ 50 x 50 microns by using disposable microfluidic devices instead of complex imaging instrumentation. Although the resolution of our devices is currently lower than that of optical imaging, our approach has the advantage of enabling multiplexed assays to be carried out at many sites in parallel. We are currently developing methods to assay proteins as well as RNAs using this novel technology, and we envision our assays enabling high-dimensional analysis of gene expression in both normal and disease samples.

1683S

Development of a novel methodology for RNA-microbiome enrichment. *L. Ettwiller, E. Yigit, E. Dimalanta, I. Schildkraut.* New England Biolabs, Ipswich, MA, USA.

The analysis of the metagenome of human microbiomes has highlighted the fundamental role of the microbiome in human health. Though the metagenomics field has just recently emerged, there is a paradigm shift from the understanding of the microbiome composition in terms of taxonomical classification towards a functional understanding of the microbiome (the metatranscriptome). Such shift is motivated by the recent findings that functional analysis of the metatranscriptome better addresses the underlying complex molecular interplay which affects the emergence of different phenotypes and diseases. This paradigm shift means that future microbiome studies will be complemented with a metatranscriptome analysis. Human microbial metatranscriptome sequencing poses significant challenges. Specifically, prokaryote non-ribosomal RNA typically constitutes a very small fraction of the total RNA, where ribosomal RNA (rRNA) and contaminating host RNA are the overwhelming majority of transcripts (up to 99%). Despite the urgent need to avoid sequencing rRNA, the existing technologies to deplete all rRNA using subtractive hybridization are far from perfect. Here we describe a novel technology that allows specific capture of the active and informative transcriptome of all species of bacteria and archaea while removing both the host RNA and rRNA in a single step. Unlike subtractive hybridization where the rRNA sequence must be known a priori to generate RNA probes, our strategy is universal. All prokaryotic and archaeal mRNAs and small RNAs will be captured. On the other hand, processed transcripts such as mature ribosomal and transfer RNAs (rRNA, tRNA) from all organisms and mRNA from eukaryotes will be discarded. As a proof of principle for metatranscriptome analysis, we demonstrated that mRNA from E. coli could be enriched 10 fold from a mixture of total human RNA and total E. coli RNA while maintaining the relative representation of each prokaryotic transcript.

1684M

Extremely low-coverage whole genome sequencing in South Asians captures population genomics information. *N. Rustagi[1], A. Zhuo[2], S. Wang[2,3], N. Ramesh[1], W.S. Watkins[4], D. Muzny[1], R.A Gibbs[1], L.B. Jorde[4], F. Yu[1], J. Xing[2,3].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Rutgers, the State University of New Jersey, Department of Genetics, Piscataway, NJ,; 3) 3Rutgers, the State University of New Jersey, Human Genetics Institute of New Jersey, Piscataway, NJ; 4) 4 University of Utah School of Medicine, Eccles Institute of Human Genetics, Salt Lake City, UT.

The cost of whole genome sequencing (WGS) has been reduced considerably in recent years due to advances in next-generation sequencing technology (NGS), but many aspects of large scale cohort WGS studies are still daunting. Challenges include the requirements for quick turn-around times and high variant calling accuracy. We have focused on sensitive variant site discovery, robust association study replication and comprehensive population genomics characterization. The ability to identify common variants (minor allele frequency (MAF) > 5.0%) is a central aim of such studies (Li Y. et al., 2011).

To develop a cost-effective path to variant discovery, we evaluated extremely low-coverage WGS (~1x) on 185 Indian individuals, followed by imputation using BEAGLE with reference panels from the HapMap project and the 1000 Genomes project. We applied two variant discovery pipelines, SNPTools and GATK, and generated a consensus call set. Comparisons with previous SNP microarray and sequencing results show the consensus set has an over 90% sensitivity for identifying variants with a MAF> 5%, with a false discovery rate (FDR) <15%. Imputation provides further improvement in sensitivity for both common and rare variants. Restricting standard population genetics analysis to shared regions with high overall coverage, shows a high level of consistency with past results. This data set will enable in-depth understanding of genomic diversity and detailed inference of demographic history in South Asia.

We further characterized expected sensitivity and accuracy of SNP calling in extremely low- coverage sequencing studies using simulations. We down-sampled from the 1000 Genomes data to reflect coverage of 0.25x and 0.75x and found ~ 500 samples with 0.75x coverage are sufficient to recover > 60% common SNPs (MAF > 5%) with an FDR of ~3%. For the dataset with ~ 0.25x coverage, > 40% SNPs are recovered from ~ 500 samples for MAF> 20% and FDR <3%. Only 350 samples with 0.75x coverage will yield 80% true SNPs, with MAF >= 20%. Overall, we demonstrate that extremely low-coverage WGS with imputation can be a useful study design for variant discovery with a dramatically reduced cost, even in populations without available reference data.

1685T

Advancing Clinical Diagnostics Using Whole Exome Sequencing. *D. Muzny[1], C. Buhay[1], J. Hu[1], M. Wang[1], Y. Han[1], H. Dinh[1], C. Kovar[1], H. Doddapaneni[1], M. Bellair[1], M. Bainbridge[1], Y. Yang[2], C. Eng[2], J. Lupski[2], A. Beaudet[2], E. Boerwinkle[1, 3], R. Gibbs[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston TX 77030; 3) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.

Advancement of whole exome sequencing (WES) in the clinical diagnostic arena has required methods for rapid, robust delivery of high-quality sequencing data covering clinically relevant genes. We previously described a DNA capture reagent (VCRome 2.1), developed by the Baylor College of Medicine (BCM-HGSC) in collaboration with Roche/NimbleGen. Our CLIA/CAP certified clinical Whole Genome Laboratory (WGL) has now analyzed more than 3,650 cases of childhood disease with VCRome 2.1. We have also improved these methods by development of an innovative exome 'polishing' technique and fast workflow for quick WES library preparation and DNA capture. WES capture target designs have focused on the ~1,800 loci listed in the GeneTests series. We generated a novel 'spike-in' design comprising 220 Kb of additional oligonucleotide probes for use in conjunction with VCRome design for improved sequence coverage of inadequately (<20X) covered exonic regions in the GeneTests list. The addition of the 'spike in' converted >400 GenTests genes from partially covered to fully covered. Analysis has shown that the design effectively captures high GC regions (79-90% GC) but does not address low coverage regions in repetitive DNA. An expanded design comprising 3,671 additional unique gene targets derived from GeneTests, OMIM, selected cancer genes and WGL positive cases, is now under evaluation. This new design is anticipated to improve gene coverage even with relatively low total sequence input, and will therefore be compatible with analysis of standard research samples, where cost constraints the total sequence depth. Development of the quick WES protocols has allowed us to complete the entire NimbleGen-based whole exome sequencing process within 5 days including library generation through analysis. Both methods have been implemented in the WGL where they can impact applications in prenatal, neonatal intensive care and other critical settings as well as for undiagnosed genetic disease in clinical and research samples.

1686S

Next generation sequencing in a diagnostic laboratory: Pros and cons of enrichment technologies. *B.P. Dworniczak[1], S. Fleige-Menzen[1], P. Pennekamp[2].* 1) Institut fuer Humangenetik, Universitaetsklinikum Muenster, Germany; 2) Klinik fuer Kinder- und Jugendmedizin -Allgemeine Paediatrie-Universitaetsklinikum Muenster, Germany.

Molecular diagnosis of complex human genetic diseases is still challenging because in most cases multiple genes harboring putative deleterious mutations have to be analyzed. So far in most diagnostic laboratories Sanger sequencing is still used as the golden standard but capillary sequencing is excessive time-consuming and expensive at least for the screening of multiple genes. However, recently there was a shift away from Sanger sequencing after introduction of high-throughput sequencing methods, which are often collectively referred to as "next-generation" sequencing (NGS) which have facilitated substantial increases in sequencing content while dramatically decreasing the cost per base. But because these technologies are originally introduced especially for large sequencing projects it is difficult to scale down this technology for screening disease causing genes in a diagnostic laboratory with its specific needs and requirements. To fill this gap table top NGS Systems have been introduced by Life Sciences (Ion Torrent PGM and Proton) by Illumine (MiSeq) and by Roche (GS Junior). While Illumina and Roche launched sequencer adapted to established technology Ion Torrent introduced a sequencing device using a sequencing technology based on the detection of hydrogen ions that are released during the replication of DNA. To validate this technology in respect to usability, software requirements and accuracy we tested several gene panels comprising between 3 and 420 genes covering between 100 and 16000 exons. Regions of interest were enriched in different ways: single PCR; multiplex-PCR (AmpliSeq: Ion Torrent; GeneRead NGS System: Qiagen) or HaloPlex custom designed kits (Agilent). Our validation showed that technically multiplex PCR seems to be the superior technology for target enrichment because of its easy workflow, but that severe problems can emerge for proper detection of sequence variants if the user is not aware of its limitations. Especially the formation of sequencing blocks of the target region prevents the detection of complex variants at the ends of these blocks facilitating false negative claiming. This serious problem might impair the overall use of PCR amplification for target enrichment at least as long as software solutions cannot handle the problem.

1687M

Unique Haplotype structure determination in human genome using Single Molecule, Real-Time (SMRT) sequencing of targeted full-length fosmids. *K. Eng[1], R. Hall[1], L. Hon[1], C. Pyo[2], D. Geraghty[2], S. Ranade[1].* 1) Pacific Biosciences, Menlo Park, CA; 2) Fred Hutchinson Cancer Research Center, Seattle, WA.

Determination of unique individual haplotypes is an essential first step toward understanding how identical genotypes having different phases lead to different biological interpretations of function, phenotype, and disease. Genome-wide methods for identifying individual genetic variation have been limited in their ability to acquire phased, extended, and complete genomic sequences that are long enough to assemble haplotypes with high confidence. We explore a recombineering approach for isolation and sequencing of a tiling of targeted fosmids to capture interesting regions from human genome. Each individual fosmid contains large genomic fragments (~35 kb) that are sequenced with long-read SMRT® technology to generate contiguous long reads. These long reads can be easily de novo assembled for targeted haplotype resolution within an individual's genomes. The P5-C3 chemistry for SMRT sequencing generated contiguous, full-length fosmid sequences of 30 to 40 kb in a single read, allowing assembly of resolved haplotypes with minimal data processing. The phase preserved in fosmid clones spaned at least two heterozygous variant loci, providing the essential detail of precise haplotype structures. We show complete assembly of haplotypes for various targeted loci , including the complex haplotypes of the KIR locus (~150 to 200 kb) and conserved extended haplotypes (CEHs) of the MHC region. This method is easily applicable to other regions of the human genome, as well as other genomes.

1688T

Examination of the Performance of Whole Genome Amplified DNA across Multiple Capture Methodologies and Sequencing Platforms. *B. Hicks[1,2], C. Dagnall[1,2], M. Malasky[1,2], H. Lee[1,2], R. Eggebeen[1,2], J. Mitchell[1,2], W. Luo[1,2], X. Zhang[1,2], C. Chung[1,2], M. Yeager[1,2], K. Jones[1,2].* 1) Cancer Genomics Research Laboratory, Leidos Biomedical Research, DCEG, NCI, Bethesda, MD; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), Rockville, MD, USA.

While input requirements for targeted, whole exome and whole genome sequencing have been driven down considerably by the advent of new and improved technologies, often whole genome amplified DNA (wgaDNA) remains the only source of critical and informative subjects. Additionally, the advent of single cell technologies demands use of wgaDNA in downstream analysis. We performed whole exome and targeted sequencing on a series of matched genomic DNA (gDNA) and wgaDNA samples, utilizing different capture techniques, including the hybridization based Roche NimbleGen SeqCap EZ Human Exome v3.0, the amplification based ThermoFisher Ion AmpliSeq Exome, and two custom Ion AmpliSeq panels (47 amplicons and 856 amplicons). The captured libraries were sequenced on the Illumina HiSeq 2500, the ThermoFisher Ion Proton, and the Ion Torrent PGM, respectively. Data filtering, alignment and variant calling was done per internal standard protocols. We reviewed overall coverage statistics, as well as positional coverage variance in areas of high GC content, centromeric and telomeric regions. Sensitivity and specificity of the different approaches was determined via use of the NIST reference genotype data from the well characterized NA12878 (Genome in a Bottle). Assessment of coverage statistics (mean coverage for Illumina platform and uniformity for Ion platforms) show sequencing coverage is comparable between gDNA and matched wgaDNA samples. Duplication rates for the NimbleGen/Illumina samples were slightly (0.5%) higher for the wgaDNA source, but well within acceptable limits. All approaches showed some bias towards gDNA in coverage of high GC content areas and telomeric regions, with the amplification based capture methods being most susceptible to differential coverage in these specific areas. SNV concordance for the NimbleGen/HiSeq gDNA and wgaDNA as compared to the high confidence genotypes from GiaB was 99.7% and 99.8% respectively, with indel concordance rates of 98.3% and 98.2%. SNV concordance for the AmpliSeq Exome/Proton gDNA and wgaDNA was 98.9% and 99.1%, with indel concordance rates of 90.7% and 90.9%. False negative rates were higher when utilizing AmpliSeq Exome, but only a slight increase (1.8% for SNVs and 5.5% for indels) when comparing wgaDNA to gDNA. These results show an ability to utilize wgaDNA for all sequencing platforms, with some caution on use of amplification based capture techniques.

1689S

Improved Exclusion Amplification Chemistry Supports Sequencing TruSeq PCR-free Libraries for Human Whole Genome Sequencing on Illumina's HiSeq X Ten System. *A.C. Kwasniewska[1], P. McInerney[2], C. Bevis-Mott[1], J. Boutell[1], S. Hunter[2], M. Tsan[2], M. Fabani[2], A. Silbergleit[2], M. Chen[2], H. You[2], M. Niziolek[2], E. Guzman[2], E. Kanterakis[1], M. Pham[2], E. Bomati[2], M. He[2], R. Shen[2], S. Tousi[2], J. Whitacre[1].* 1) Illumina Cambridge Ltd, Chesterford Research Park, Essex, CB10 1XL, UK; 2) Illumina, Inc. - Worldwide Headquarters, 5200 Illumina Way, San Diego, CA 92122 USA.

The HiSeq X Ten was designed to unlock the potential of population sequencing, delivering the world's first $1000 human genome at 30x coverage. By using flow cells with billions of patterned nanowells at fixed locations, we have significantly increased the density of clusters that could be achieved by random clustering. To support clustering on patterned flow cells we developed a new cluster chemistry called Exclusion Amplification (ExAmp). Here we describe an improved ExAmp v2 formulation that supports simultaneous seeding and amplification of TruSeq PCR-free libraries, the gold-standard for genome sequencing due to more uniform coverage of regions that are typically underrepresented in amplified libraries (high GC-rich regions, promoters and repetitive regions). This new formulation continues to support amplification of TruSeq Nano PCR libraries and also offers a significant improvement in genomic coverage uniformity. We present data showing improved coverage uniformity for challenging regions of the genome whilst maintaining high quality sequence at genes and exons. Furthermore, by utilizing the Platinum genome data sets from the CEPH pedigree, we have validated our data and found excellent sensitivity and specificity of SNP calling with the improved ExAmp clustering.

1690M

Lower Cost, Higher Throughput Library Preparation with the Echo liquid handler® and the NuGEN Ovation® Single Cell RNA-Seq System. *J.D. Lesnick[1], J.D. Heath[2].* 1) Labcyte Inc, Sunnyvale, CA. USA; 2) NuGEN Technologies, San Carlos, CA. USA.

The rapid evolution of next-generation sequencing (NGS) technologies is accelerating our knowledge of gene expression, regulation and pathway complexities in mammalian cells. Transcriptome analysis with NGS offers increased transcript coverage to enable the detection of rare transcripts, novel alternative splice isoforms and the measurement of transcript abundance. However, traditional library preparation methods for NGS are often not amenable to transcriptome analysis. Traditional methods carry a requirement for a large amount of total RNA to yield sufficient mRNA to analyze which can be unobtainable or cost prohibitive in most experiments. The NuGEN Ovation Single Cell RNA-seq System is a highly sensitive and complete library preparation procedure for whole-transcriptome sequencing that requires total RNA from samples as small as a single cell or 10 picograms. In this study, the Ovation Single Cell RNA-Seq System was validated by examining the resulting library complexity, reproducibility, and evenness of transcript coverage. Furthermore, a majority of the sample and reagent transfers were automated using the Echo acoustic liquid handling technology. Echo liquid handlers transfer a wide range of fluids without contact of tips or recalibration between fluid types. The industry leading accuracy and precision of Echo liquid handlers at microliter and nanoliter volumes in combination with the NuGEN Ovation Single Cell RNA-Seq System increases library preparation throughput while reducing the costs to enable a broader application of transcriptome analysis with NGS.

1691T

Low to mid-throughput automation of hybridization based capture technologies using Apollo 324 NGS Library Prep System. *M. Srinivasan, R. Acob, J. Berger, L. Chan, S. Silveria, S. Wei, J. Dunne, S. Nasarabadi.* WaferGen Biosystems, Fremont, CA.

Targeted resequencing of exonic regions (subsets to whole exome) using hybridization-based capture has been used as a key discovery tool for understanding many biological problems including disease mechanisms of Mendelian diseases and cancer. In addition, discoveries from next generation sequencing (NGS) efforts are increasingly being used to design targeted panels for clinical research and diagnoses. However, as targeted sequencing moves to the clinical setting, very few automation solutions exist that minimize process variability and reagent costs for low and mid-tier labs. Furthermore, labs are asked to perform multiple tests with limited amount of material. To address these needs, we have tested our the Apollo 324 NGS Library Prep System to enrich for exomes from NGS libraries prepared from low input material. The Apollo 324 is a flexible automation system that has been shown previously to process up to 8 or 32 samples in a single run. The number of samples to run in a given day is matched to the amount of reagents/sample to eliminate reagent waste. Genomic DNA and cDNA library preparation kits for the Apollo 324 robot can process samples for whole genome sequencing, RNA-seq and Chip-Seq experiments. Now, we are extending the utility of automation to include the ability to perform targeted capture. Commercially available target enrichment kits (SureSelect, Nimblegen) were processed manually according to manufacturer's instructions and compared with the automated system at different input levels of gDNA from HapMap and tumor samples. Data will be presented from exome-enriched samples from manual and automated workflows, sequenced using a HiSeq system and compared for their ability to identify variants.

1692S

Sequencing Beyond the Read Length Officially Supported on HiSeq 2500: the Error Profile and Remedy. *W. Wang[1], D. Storton[1], J. Wiggins[1], L. Guo[1], J. Peng[2], P. Andolfatto[1,2].* 1) Institute for Integrative Genomics, Princeton University, Princeton , NJ; 2) Dept of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544.

Background: The throughput of Next-Generation Sequencers has increased dramatically since their advent. However, the read length remains relatively short among sequencers which are able to generate large number of reads, for example, the 150nt maximum officially supported by Illumina on HiSeq 2500. Such read length is generally sufficient for alignment to the genome, but longer read length, even to some small extent, is needed for de novo assembly of genomes and transcriptomes. Methods: The observed error rate of HiSeq 2500 Rapid Run is generally below 1% at the very end of 150nt reads, therefore further extending the read length might yield slightly decreased but still acceptable read quality as long as the sequencing reaction is optimized accordingly. Pair-End Rapid Runs of 215+230nt and 215+215nt reads were performed on HiSeq 2500 with slight modification of the standard sequencing procedure. Small amount of PhiX control library was spiked-in to each lane to estimate the actual sequencing error profile in addition to the Illumina Quality Score. Size-selected libraries of narrow size range were used to generate clusters of homogeneous size and the clustering density was reduced from the standard value to help improve the signal-to-noise ratio and read quality. Result: By controlling the raw cluster density around 600K/mm2, the overall error rate of 215nt read pairs could be kept below 1%, while the per cycle error rate at the very end of these reads remained below 2%. The Illumina Q30 bases of each lane could be maintained over 80%. The bias in the base call composition became more noticeable beyond the official read length of 150nt, especially when the cluster density is relatively high. Very low cluster density could lead to reduction in the base composition bias, although not much improvement in the base call accuracy. In conclusion, we have optimized the sequencing condition to extend the HiSeq 2500 read length from 150nt by 43% to 215nt. At the expense of reduced cluster density, the quality of the extended reads is mostly maintained.

1693M

Genome-wide transcriptome enrichment sequencing for research and clinical applications. *H. Doddapaneni[1], J. Hu[1], H. Chao[1], X. Liu[1], S. White[1], K. Walker[1], C.J. Buhay[1], M. Wang[1], M. Bellair[1], L. Wang[1], K.R. Covington[1], A. Roy[2,3], P. Sumazin[3], S.E. Plon[1,3,4], D.W. Parsons[1,3,4], D.A. Wheeler[1], E.A. Boerwinkle[1,5], D.M. Muzny[1], R.A. Gibbs[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX; 3) Department of Pediatrics, Baylor College of Medicine and Texas Children's Hospital, Houston, TX; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.

Transcriptome sequencing (RNA-Seq) together with whole exome sequencing (WES) offers an integrated informative dataset for clinical diagnosis and characterization of human transcriptome, especially for characterizing cancer etiology and outcomes. For example, detection of pathogenomic gene fusion events is important to accurately characterize childhood sarcomas where morphologic characteristics or WES is not sufficient. However, current RNA-Seq methods rely heavily on very high quality RNA samples, which are often not available in routine clinical setting. Towards the goal of incorporating RNA capture methods in a clinical setting, we have developed an RNA-Seq capture protocol combining the BCM-HGSC strand-specific RNA-Seq protocol with our CLIA certified WES protocol. Inclusion of a capture step enables RNA-Seq to become RNA quality independent as it provides probes to capture the targeted transcripts. Libraries were prepared using Universal Human Reference RNA and co-captured (4-plex) using VCRome 2.1 capture reagents and sequenced on HiSeq 2500. On average, 66 million reads were generated per sample and these data were compared to standard RNA-Seq data generated for the same sample. Capture libraries had a significantly higher mapping rate (95.1% vs 77.5%), higher exonic rate (93.4% vs 77%) and lower intergenic rate (0.2% vs 4.2%) compared to RNA-Seq libraries. Correlation of expression values (in FPKM) between the replicate capture libraries was excellent (R^2 = 0.9882). Capture libraries detected 18,903 protein coding genes compared to 18,611 genes found in RNA-Seq libraries. There were 756 genes unique to capture libraries (FPKM-0.22) and 465 genes unique to RNA-Seq libraries (FPKM- 2.8). The low FPKM values observed for the novel genes found in capture libraries suggest 'enrichment' of low expressed transcripts due to presence of their probes. Further, use of a genome-wide capture probe such as VCRome allows for capture and validation of novel fusion-transcripts. We are validating the protocol on samples with 13 known fusion transcripts found in hepatoblastoma and pediatric sarcoma samples that will then serve as test controls. Concurrently, we are also testing RNA of different types (RIN numbers, input amounts) and from FFPE. We anticipate that our RNA-Seq capture protocol will provide robustness and through high through-put process automation, will result in fast turnaround times to have practical utility in both research and clinical settings.

1694T

Whole Genome Sequencing on DNA extracted from Saliva: a systematic evaluation of SNV, CNV and Structural Variant Calling. *S. Germer[1], A.-K. Emde[1], A. Abyankar[1], B. Cornes[1], D.M. Oschwald[1], T. Hu-Seliger[2], P. Yurttas Beim[2], R.B. Darnell[1].* 1) New York Genome Center, New York, NY 10013; 2) Celmatix, 1 Little West 12th St. New York, NY 10014.

As whole genome sequencing (WGS) becomes increasingly affordable, DNA repositories are being established for a variety of research purposes. These repositories often rely on blood specimens, which presents several practical hurdles. The ability to source DNA from saliva would reduce costs and increase flexibility for many researchers. Given sample procurement is a common bottleneck in genomic studies, this development would have significant implications. Previously, it has been reported that saliva yields DNA of sufficient quality to reliably make single nucleotide variant (SNV) calls in WGS datasets. Here we address whether larger variants, including copy number and structural variants (CNVs and SVs) can also be reliably detected from saliva-extracted DNA. We evaluated WGS data (~30x coverage, Illumina HiSeq) generated from matched blood and saliva (DNA Genotek, Oragene) samples selected to represent a spectrum of bacterial contamination levels (avg=10%). Saliva DNA showed both a reduced alignment rate to the human genome (0.74-0.93) compared to blood sample DNA (>0.95) and a corresponding reduction in mean genome-wide coverage. A pre-sequencing qPCR assay predicted bacterial contamination rates that correlated with the alignment rates for the human genome (r=-0.94). SNV calls from both sets of samples (GATK best practices pipeline) showed similar rates of discordant calls between replicate saliva and blood samples (~0.02% at 99% VQSR tranche). Similarly, DNA from saliva and blood displayed very similar Mendelian inheritance errors (0.01%-0.05%). To investigate the impact on SV calls, we employed a dual analysis: a) paired saliva-blood somatic SV calls using CREST (SVs) and NBIC-seq (CNVs), and b) replicate analysis of saliva-blood SV calls from GenomeSTRiP. Across matched saliva-blood pairs we did not identify any somatic SVs that could be confirmed by replicates or any somatic CNVs. GenomeSTRiP calls exhibited greater variability, but the variability between blood and saliva was comparable to the variability between replicates and largely due to differences in genome coverage. Therefore, the discordant calls likely represent lack of sensitivity/specificity in the calls made by GenomeSTRiP. Our study suggests that DNA from saliva can yield comparable WGS results to DNA from blood across a range of variant sizes and types, especially when coupled to pre-sequencing estimation of bacterial contamination to allow for comparable mean genome coverage.

1695S

Precise Quantification of Bias in Whole-Genome Amplification Using Droplet Digital™ PCR. *N. Heredia, J. Berman, W. Yang, S. Hodges, E. Hefner.* Bio-Rad Laboratories Inc., Digital Biology Center, 5731 W Las Positas Blvd Pleasanton, CA 94588.

Whole-genome amplification (WGA) is commonly used upstream of next-generation sequencing and microarray protocols when available genetic material is severely limiting, such as in single cell work. However, WGA can suffer from sequence-specific bias, allelic drop out, and non-uniform coverage, hampering data interpretation. To quantitatively assess bias introduced by WGA, we used Droplet Digital™ PCR (ddPCR™), targeting multiple sites across the genome. Limited samples were prepared using standard WGA protocols. The use of droplets to improve WGA output will be discussed.

1696M

ThruPLEX-FD as high sensitivity library prep tool for whole exome and target panel sequencing. *J.P. Jerome, E. Jan, J. Wibbenmeyer, J. Langmore, K. Shazand.* Rubicon Genomics Inc., 4355 Varsity Dr., Suite E, - Ann Arbor, MI., USA.

Whole exome sequencing (WES) and target panel sequencing (TPS) are currently the most popular applications in the NGS community. Although ThruPLEX-FD has already been validated for WES and TPS applications in the laboratories of current customers, Rubicon wants to make its internal protocols available publically for the convenience of all its customers. Here, we describe experimental results answering the following questions: -Can ThruPLEX-FD be easily used for enrichment with Agilent XT2, Roche SeqCap EZ for WES and TPS? (protocol establishment) -How do ThruPLEX-FD enrichment metrics compare with the standard Roche (Kapa) and Agilent libraries? -What are the advantages of using ThruPLEX-FD? Library synthesis: Three inputs were used to build ThruPLEX-FD libraries with Covaris-sheared 200bp average human gDNA, respectively 500pg, 10ng and 50ng. Recommended low input libraries for Roche (10ng) and Agilent XT2 (100ng) were prepared with the same gDNA and enriched according to the vendor protocols. Capture: ThruPLEX-FD libraries were pooled with the Agilent library for hybridization (different barcode systems), however hybridized separately from the Kapa library (identical barcode system). Vendor-specified reagents and protocols were used except for addition of ThruPLEX-FD-compatible blocking oligonucleotides IDT xGen® Universal Blocking Oligo - TS-p5 and xGen® Universal Blocking Oligo - TS-p7 (6nt), spiked into ThruPLEX-FD hybridizations at the concentrations suggested by the vendor. Sequencing: Pools of the enriched libraries were sequenced on Illumina MiSeq, and the reads from each library were down-sampled to 3.9M read-pairs before removal of duplicates. Detailed data will be shown at the meeting. The following conclusions were drawn: -ThruPLEX-FD can be easily enriched using standard Agilent and Roche protocols and reagents by simply spiking in IDT universal blocking oligos. -ThruPLEX-FD was as highly enriched as Kapa and better than Agilent libraries. For comparable inputs, duplicate rates are often lower than other tools, enabling higher coverage. -Finally, the low input results open the possibility of efficient enrichment for the clinical studies such as CTC, plasma, FFPE or LCM applications. These data show that ThruPLEX can be efficiently used in any enrichment approach with metrics that equal or exceed other existing tools.

1697T

Identify enhancer elements at genome-wide scale using MIT-seq. *X. Wu, L. Su, D. Wells.* Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Frederick, MD.

Enhancer elements are important for regulation of gene expressions. However enhancers are not easy to identify and they can be far away from genes. Retrovirus integrates into the host genome as part of their life cycle and leaves a permanent tag at the location of integration. Different retroviruses use different mechanisms for target selection. Recently, we have demonstrated that murine leukemia virus (MLV) integration site clusters correlate highly with enhancer elements in the human genome. MLV retroviral vector has been widely used for gene delivery and is highly efficient to transduce a broad range of cells. Using linker-mediated PCR, we can easily identify millions of MLV integration tag by sequencing (MIT-seq). The MIT-seq tags are analyzed similarly as ChIP-seq tags to identify peaks and clusters, making this a new approach to identify enhancer elements in a variety of cells. (Funded by FNLCR Contract HHSN261200800001E).

1698S

Contiguity Preserving Transposition Sequencing (CPT-seq): Haplotype-resolved sequencing and assembly. *J. Fisher[1], C. Turk[1], L. Christiansen[1], F. Zhang[1], E. Kostem[1], S. Amini[1], M. Ronaghi[1], A. Adey[2], J. Shendure[2], K. Gunderson[1], F. Steemers[1].* 1) Advanced Research Group, Illumina, Inc, San Diego, CA, 92122, USA; 2) Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA.

Haplotype-resolved genome sequencing enables the accurate interpretation of medically relevant genetic variation, deep inferences regarding population history, and the non-invasive prediction of fetal genomes. We describe a novel library preparation and haplotyping approach that enables haplotype assembly of short sequencing reads from as little as 50 ng gDNA using 2-4 lanes of Hiseq over the standard VCF. The method employs a Nextera Tn5 transposition approach that preserves the contiguity and ordering information of the target DNA sequence while inserting common primer, and optionally index sequences at 100-200 bp intervals. Subsequently, the contiguously linked-libraries are processed through the assay while maintaining this contiguity information. Finally, the individual short libraries of each DNA molecule are resolved through sequencing. The method effectively generates long-strobed reads that are on average ~30-50kb in length with ~5-15% coverage. Phased strobed reads can be as long as the physical length of input DNA. Additionally, we have greatly streamlined the haplotyping protocol. The total assay time is <3 h with ~ 30 min hands-on-time and the workflow can be integrated with robotic or microfluidic processing. As a proof-of-concept, we apply this method to phase over 95% of heterozygous variants from a HapMap trio into long, accurate haplotype blocks with N50 block sizes ranging from 1.4-2.3 Mb. We evaluated phasing yield and accuracy of the current method by plotting the probability that heterozygous SNP pairs are on the same phasing block as a function of distance between them. For all pairs that are on the same phasing block, the chance that a pair is phased correctly, again as a function of distance is plotted. From these results it can be concluded that phasing yield is in Mb-scale (>80% SNPs sharing phasing block) and accuracy of 99.9% and 99% extends to 100kb and 200kb pairwise SNP distances, respectively. We demonstrate utility of this imputation-free approach by accurate and complete phasing of de novo and compound heterozygous SNPs. We anticipate the scalable, rapid, and cost-effective workflow could enable haplotype resolution to become routine in human genome sequencing.

1699M

A Method for Selectively Enriching Microbial DNA from Contaminating Vertebrate Host DNA. *E. Yigit[1], G. R. Feehery[1], F. J Stewart[1], S. O. Oyola[2], Y. Wei Lim[3], B. W. Langhorst[1], V. T. Schmidt[4,5], E. T. Dimalanta[1], L. A. Amaral-Zettler[4,6], T. Davis[1], M. A. Quail[2], S. Pradhan[1].* 1) New England Biolabs, Ipswich, MA; 2) Wellcome Trust Sanger Institute, Cambridge, UK; 3) San Diego State University, San Diego, CA, USA; 4) The Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA, USA; 5) Department of Ecology and Evolutionary Biology, Brown University, Providence, RI, USA; 6) Department of Geological Sciences, Brown University, Providence, RI, USA.

Recent discoveries have implicated the human microbiome as playing a role in certain physical conditions and disease states, and these advances have opened up the potential for development of microbiome-based diagnostic and therapeutic tools. The majority of microbiome DNA studies to date have employed 16S analysis, but these provide very little information regarding function. In contrast, sequencing of the total DNA of a microbiome sample provides a broader range of information including genes, variants, polymorphisms, and putative functional information. However, many samples, including those derived from vertebrate skin, bodily cavities, and body fluids, contain both host and microbial DNA. Since a single human cell contains approximately 1,000 times more DNA than a single bacterial cell, even low-level human cell contamination can substantially complicate the analysis of a sample. In some cases, as low as 1% of sequencing reads may pertain to the microbes of interest and a large percentage of sequencing reads must be discarded, making such experiments impractical. To address this issue, we developed a method to enrich for microbial DNA using methyl-CpG binding domain (MBD) to separate methylated host DNA from microbial DNA. Importantly, microbial diversity and relative abundance is maintained after enrichment. This simple magnetic bead-based method was used to remove human or fish host DNA from bacterial and protistan DNA. We describe the enrichment of DNA samples from human saliva, human blood, a mock malaria-infected blood sample, human cystic fibrosis sputum, and a black molly fish, followed by next generation sequencing on multiple platforms. Sequence reads aligning to host genomes were reduced approximately 50-fold, while the percentage of sequence reads corresponding to microbial sequences increased approximately 10-fold. This new method for microbiome sequence analysis holds promise for use with a variety of sample types, enabling enrichment while accurately reflecting the diversity of the original sample.

1700T

Woman endometrium biopsy immediate single-cell analysis. *K. Krjutskov[1,2], S. Katayama[1], M. Saare[2], K. Samuel[2], D. Lubenets[3], S. Linnarsson[4], J. Kere[1], A. Salumets[2].* 1) BioNut, Karolinska Institutet, Stockholm, Sweden; 2) CCRMB, Tiigi 61b, Tartu, Estonia; 3) UT IMCB, Riia 23, Tartu, Estonia; 4) Med Biochem and Biophysics, Karolinska Institutet, Stockholm, Sweden.

The recent progress in single-cell analysis has unlocked clinical interests of women's health- and infertility research like endometrial pathologies, implantation failure and etiology of endometriosis. Although woman endometrium is extensively studied, the most essential players - individual endometrial cells, have gained little attention. The objective of my study is to take deeper view on endometrial functioning at individual cellular level by providing gene expression based phenotyping and defining the functionally active major cell populations and their molecular interactions. This goal will be achieved by using clinical endometrial biopsies and combining the most advanced cell- and molecular-biology techniques, like scrutinized biopsy handling from collection to treatment, multiple-color FACS analysis with single-cell sorting option, and highly multiplex single-cell transcriptome analysis by RNA-sequencing. Voluntary healthy fertile women donate the endometrial biopsies during the proliferative and mid-secretory, e.g. receptive phase of the natural menstrual cycle. The samples are cryopreserved in proper media and the thawed tissues are enzymatically disaggregated with the cell suspensions containing the epithelial- (CD9+), stromal- (CD13+), and the rest repertoire of living cells. During the sample treatment, minimum manipulation times at +4°C are used to retain the natural gene expression profile. Cells are labelled by combination of surface marker antibodies for single-cell FACS sorting directly into 96-well plate lysis buffer and are converted to 48-plex Illumina libraries using Single-cell Tagged Reverse Transcription (STRT) protocol. After cell sorting and cDNA synthesis, qPCR-based QC assay is used to evaluate each well/cell's sorting accuracy and cDNA yield before pooling and further library preparation. FACSArial single-cell sorting accuracy and cDNA turn-around success rate is 90-100%. This multidisciplinary study have faced us to different challenges like (i) proper biopsy cryopreservation and treatment procedure to prepare stable single-cell solution in minimum time, (ii) single-cell QC after FACS and cDNA synthesis, and (iii) STRT adaptation for differentiated low mRNA cells analysis. Here mentioned improvements enable us to study endometrium single-cells directly from biopsy and brings single-cell genomics closer to clinical needs.

1701S

High Accuracy Variant Detection using HaloPlex with Molecular Barcodes. *h. johansson, b. skarpas, e. agne, l. forsmark, p. eriksson, m. isaksson.* Agilent Technologies, santa clara, CA., Select a Country.

HaloPlex is a next generation PCR target enrichment method that enables enrichment of thousands of targets in a single tube. The protocol utilizes specificity gained form restriction enzyme recognition, hybridization and DNA ligation to capture molecules originating from the target region to be sequenced. The target region is fully customizable and may consist of either a continuous or up to several thousand discrete regions. Amplicon based methods for multiplex target enrichment are, in general, convenient methods for capturing a wide range target region sizes. In contrast to hybridization capture methods where random shearing is deployed, it is not possible for Haloplex and other amplicon based techniques to use read start point of paired end reads to identify duplicate reads. Duplicate read information can be useful for improving base calling accuracy and to monitor sampling to determine the degree of confidence to assign calls at different presumed allelic fractions. For somatic variants which are generally present in lower than 50% allelic fraction it is even more advantageous to know how many molecules have been sampled in a particular region. The main steps of the Haloplex protocol are restriction digestion, target hybridization and DNA ligation of motifs needed for sequencing on Ion Torrent or Illumina Platforms. To enable identification of duplicate reads in sequencing libraries prepared with Haloplex, we have added a molecular barcode to the introduced primer cassettes. The molecular barcode consists of ten degenerative bases allowing for over one million unique sequences to be present for tagging of molecules. Using information derived from the molecular barcode sequences we demonstrate observation of variants down to 5% allelic fraction in multiple molecules tagged with different molecular barcodes. The new protocol has, besides the introduction of molecular barcodes, been optimized in few additional aspects. Due to improved reagent formulations and streamlining of workflow, complete target enrichment can now be completed in less than 5 hours. Using 50 ng input we demonstrate >85% specificity and above 90% of target regions being covered at >10% of average depth.

1702M

Immune sequencing protocol for complete B-cell and T-cell repertoire sequencing. *F.J. Stewart[1], E.T. Dimalanta[1], A.W. Briggs[2], T. Gilbert[2], C. Clouser[2], W. Donahue[2], G. Yaari[2], L. Apone[1], S. Russello[1], T.B. Davis[1], F. Vigneault[2].* 1) New England Biolabs, Inc., Ipswich, MA; 2) AbVitro, Inc., Boston, MA.

Immune sequencing, which allows for the study of complex immunological diseases by sequencing B-cell antibodies and T-cell receptors, is gaining in popularity due to recent throughput and read length improvements in next-generation sequencing (NGS) technologies. However, the structural and sequence complexities of antibody genes have made reliable targeting approaches challenging. We have developed and optimized a method for accurate sequencing of the immune gene repertoires of B-cells and T-cells. In contrast to previous studies, our method generates full-length sequences of B-cell antibody and T-cell receptor genes. This allows for exhaustive somatic mutation profiling across complete V, D and J segments, full isotype information analysis (IgM, IgD, IgG, IgA and IgE), and the possibility for synthesis and expression of complete antibody chains for downstream immunological assays. By introduction of a unique barcode ID into every captured mRNA molecule, all PCR copies of each mRNA fragment can be collapsed into a single consensus sequence, making the assay extremely accurate by resolving PCR bias and sequencing errors, as well as allowing quantitative digital molecule counting. The assay can work with as low as sub-nanogram levels of input total RNA and is the first method that allows targeted amplification of all possible antibody heavy and light chains (IGH+IGK+IGL) and T-cell receptors (TCRA+TCRB) in a single reaction simultaneously.

1703T

Optimized DNA extraction and repair improves the yield and quality of sequencing libraries derived from FFPE samples. *L. Chen[1], E. Rudd[2], G. Durin[2], T. Evans[1], L. Ettwiller[1].* 1) New England Biolabs, Inc. Ipswich, MA 01938; 2) Covaris, Inc. Woburn, MA 01801.

Cancer biopsy samples are routinely formalin fixed and embedded in paraffin in order to preserve the morphological features of suspected tumor samples. While these formalin fixed, paraffin embedded (FFPE) samples are valuable for archiving histological data, the formalin treatment is detrimental to the quality of nucleic acids. Indeed, DNA isolated from archived tissue is often of poor quality and damaged due to fragmentation, oxidation, deamination, and protein-DNA crosslinks. Thus, it remains challenging to 1) extract DNA from FFPE samples and 2) obtain high quality DNA for NGS library preparation. Here we successfully coupled Covaris® truXTRAC™ to extract FFPE DNA with PreCR® -B Repair Mix to repair DNA damage. Covaris truXTRAC use highly controlled acoustic energy to efficiently remove paraffin from FFPE cores, sections, and slides enabling efficient tissue rehydration, tissue digestion, crosslink reversal, and nucleic acid release. The PreCR®-B Repair Mix is an enzyme cocktail formulated to repair damaged template DNA and blunt ends in one step. PreCR® is active on a broad range of DNA damage, including modified bases, nicks and gaps, and a variety of blocking moieties at the 3'end of DNA. We found that DNA repair pre-treatment of extracted FFPE DNA samples consistently increased library yield up to 2-fold in the four samples tested. Furthermore, downstream data analysis of the sequencing reads demonstrated a remarkable reduction of sequence miscalls. In conclusion, the incorporation of PreCR®-B in library preparation workflows remarkably improves the quantity and quality of NGS libraries from FFPE samples resulting in more robust library preparation and more reliable base calling.

1704S

Targeted enrichment of forensically relevant STRs for improved human DNA profiling. *M.R. Nandineni, S.R. Gadipally, A. Sarkar.* Lab of DNA Fingerprinting Services, Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, Telangana, India.

Human DNA profiling is employed worldwide for unambiguous identification of individuals in mass fatality incidents, criminal cases and also to establish parentage by genotyping repeat length polymorphisms in 16-18 highly polymorphic short tandem repeats (STRs) or microsatellites. DNA typing involves a multiplex PCR-based assay to generate STR profiles from even trace amounts of available biological source in a reasonably short duration. In spite of the availability of various commercial STR kits for DNA typing, there is a need for development of newer technologies to overcome limitations posed by challenging and recalcitrant forensic samples like those containing excess non-human DNA and PCR inhibitors, which often leads to allele/locus dropouts, peak imbalances and loss of signal during STR profiling. In the present study, a sequence-specific DNA enrichment strategy has been developed and evaluated to selectively 'capture' and enrich the forensically relevant human STR regions from highly compromised samples. The strategy involved simultaneous enrichment of the 18 widely used STR loci (which includes the 13 core CODIS loci) with the aid of two sequence-specific biotinylated oligonucleotide probes targeting each locus. The enriched fragments were captured using streptavidin-coated magnetic beads, eluted and subsequently subjected to multiplex amplification using commercially available STR kits followed by fragment analysis. Experiments conducted with simulated forensic samples revealed that prior enrichment of STR regions greatly facilitated the generation of complete and/or improved DNA profiles even in the presence of 5- fold excess of non-human DNA contaminants and 3-4 fold of excess PCR inhibitors (than those normally tolerated by thermostable DNA polymerases incorporated into the commercial STR kits). The pre-PCR enrichment technique provides definite improvement over the existing methodologies for DNA (STR) profiling from challenging forensic samples and holds promise for increasing the success rate of DNA testing based human identification from skeletal remains in missing persons and mass disaster victim identification programmes.

1705M

Spatially resolved single cell miRNA expression analysis in tissue sections. *M. Asp[1], P. Ståhl[2], F. Salmén[1], J. Hällman[1], A. Lundmark[3], S. Vickovic[1], S. Giacomello[1], E. Berglund[1], J. Fernandez Navarro[2], J. Sjöstrand[2], E. Sjölund[1], J. Frisén[2], J. Lundeberg[1].* 1) Department of Gene Technology, Royal Institute of Technology, SciLifeLab, Stockholm, Sweden; 2) Department of Cell and Molecular Biology, Karolinska Institute, Stockholm, Sweden; 3) Department of Dental Medicine, Division of Periodontology, Karolinska Institute, Stockholm, Sweden.

In the course of the last decade, miRNA has been shown to play a key role in a range of carcinogenic processes. Due to its high level of interspecies conservation and its known importance in cell differentiation, proliferation and apoptosis, they have been natural targets of study for researchers investigating cancer formation. Many miRNAs have already been stated as important signatures in a variety of cancers because of their abnormal expression pattern. This gives us an enormous potential of using this knowledge in clinical diagnostic practice. Tumors are frequently heterogeneous with multiple subclonal populations and it is often difficult to distinguish the exact location of cells in a tissue that express tumor related miRNAs. Standard miRNA expression analysis of a tumor biopsy produces an average expression pattern of the entire tissue, which might mask abnormal miRNA expression patterns. We are developing a method by which miRNA expression on the single cell level and within the context of an entire tissue section is possible. By using microarrays covered with miRNA-specific surface probes, we are able to capture the entire miRNA population of every cell within a single tissue section. Clusters of surface probes are individually barcoded, which gives every cell a unique positional tag. This allows the miRNA material captured on the microarray surface to be analyzed in parallel retaining information of spatial position. The final miRNA libraries are then analyzed by Massively Parallel DNA Sequencing, and the obtained miRNA data can hereby be superimposed back onto the histology image using an in house developed software.

1706T

Microbial detection using Affymetrix' Axiom® Genotyping Solution. *M. Shapero[1], S. Gardner[2], M. Mittmann[1], K. McLoughlin[2], C. Jaing[2], L. Bellon[1], T. Slezak[2].* 1) Affymetrix, Inc., 3420 Central Expressway, Santa Clara, CA 95051; 2) Lawrence Livermore National Laboratories, Livermore, CA 94551.

The human body contains a diverse set of microbes including eukaryotes, archaea, bacteria, and viruses. These microbes play a key role in both human health and disease. As sequences of microbial genomes continue to expand, microarrays are well-positioned to capitalize on this evolving information content. Here we present an overview of Axiom® Genotyping Solution with regard to the design and testing of a microbial detection array (MDA) to evaluate differentiation of several model organisms at the species and strain level.

A pilot array was designed leveraging organisms where samples and full genome sequences are available for both the primary target species and relevant near-neighbors. Multiple probe designs that iterate on probe length and tolerance to position-specific mismatches were evaluated for six families, which include Bacillaceae, Burkholderiaceae, Enterobacteriaceae, Francisellaceae, Thermotogaceae, and Paramyxoviridae. Using pure and mixed samples of defined origin as the hybridization targets, probes were scored using a composite likelihood maximization algorithm for their ability to detect perfect match sequences as well as identify novel, previously uncharacterized species/strains within a known family. Probe- and system-level performance characteristics will be presented with regard to sensitivity and specificity.

In summary, new Axiom® applications in broad microbial detection further extend the platform's capabilities. The scalable sample throughput enabled by 24-, 96-, and 384-array layouts coupled with laboratory automation will allow processing of tens to thousands of samples per week with minimal manual intervention and is consistent with the need for cost-effective examination of the microbiome in the context of bioterrorism, environmental monitoring, food safety, and human and animal health.

1707S

Design of a biobanking genotype array optimised for Chinese populations. *R.G. Walters[1], I.Y. Millwood[1], Y. Lu[2], H. Lin[3], J. Brodsky[2], M.A. Ansari[4], Y. Zhan[2], W.W. Kretzschmar[5], N. Cai[5], R. Bowden[5], J. Flint[5], P. Donnelly[5], T. Webster[2], Q. Li[3], J. Schmidt[2], L. Li[6,7], R. Peto[1], R. Collins[1], Z. Chen[1].* 1) Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), University of Oxford, UK; 2) Affymetrix, Inc., Santa Clara CA, USA; 3) BGI-Shenzhen, Shenzhen, China; 4) Department of Statistics, University of Oxford, UK; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 6) School of Public Health, Peking University Health Science Center, Beijing, China; 7) Chinese Academy of Medical Sciences, Beijing, China.

The China Kadoorie Biobank is a prospective study of 515,000 adults recruited from 10 geographically diverse areas of China. Preliminary genotyping of 82,494 subjects (after QC), conducted using a panel of 384 SNPs selected on the basis of previous genome-wide or candidate association with major clinical endpoints (e.g. stroke, diabetes, coronary artery disease) and intermediate phenotypes (adiposity, blood pressure, cholesterol), revealed substantial population stratification. Common SNPs varied in minor allele frequency (MAF) between different recruitment centres by as much as 4-fold — 311 out of 335 SNPs with overall MAF>0.05 displayed MAF differences between regions that were significant after Bonferroni correction. Even though data were available for only 216 unlinked common SNPs, principal component analysis nevertheless revealed a strong North-South axis. The scale and pattern of the differences across China strongly suggest that they are due to substantial admixture between northern and southern populations with distinct ancestries.

In part to address the challenges that admixed population structure presents for the investigation of genetic associations with disease and intermediate phenotypes, we have designed a new Affymetrix Axiom® genotyping array optimised specifically for Chinese populations, using data from diverse sources: 1000 genomes data from CHB/CHS subjects; high coverage sequence data for 156 of the same subjects and exome data from 1,746 subjects mainly from southern China, both from BGI; allele frequencies from 1,800 Taiwan Biobank subjects; and allele frequencies from the CONVERGE consortium, derived from sequence data for 9,000 subjects recruited at 59 hospitals across China. Together, these provided improved MAF estimates for known SNPs and indels, identified novel variation not previously reported in Chinese, and enabled efficient selection of array content to maximise genome-wide coverage. The array was designed following a strategy similar to that for the UK Biobank array, and includes: 142,454 variants selected on the grounds of known or suspected functional effects, including all previous GWAS hits; 81,522 variants not present in 1000 genomes CHB/CHS data; and a GWAS grid of 479,385 additional variants. With imputation, the array provides coverage of 93% of 1000 genomes variants with MAF>0.05, and 87% of variants with MAF>0.01.

1708M

High Performance Micro RNA Enrichment using Solid Phase Reverse Immobilization Magnetic Bead Technology. *B.N. Lee.* Beckman Coulter Life Sciences, Brea, CA.

Most extracted RNA samples contain ribosomal RNA and messenger RNA with only a low percentage of microRNA (miRNA). In order to measure miRNA more precisely and achieve better quality miRNA sequencing data, it is desirable to remove the larger unwanted ribosomal RNA as well as other species of RNAs. This abstract describes the use of Beckman Coulter's SPRIselect reagent kit for miRNA enrichment. The Solid Phase Reverse Immobilization (SPRI) magnetic bead method is a nucleic acid purification procedure that does not require centrifugation or filtration steps. This simple, rapid, automation friendly and produces high quality samples of miRNA for use in downstream applications Enrichment of miRNA from total RNA can be achieved in three binding steps. Step 1: Removal of large ribosomal RNA fragments. Step 2: Removal of intermediate RNA fragments. Step 3: Capturing of miRNAs and small RNAs. The micro-particles with bound miRNA are thoroughly washed with molecular biology-grade ethanol. In the elution step, the purified miRNA is easily recovered from the micro-particles using nuclease free water, which provides maximum flexibility for downstream applications. The poster presents that SPRIselect is a superior miRNA enrichment method. 1. It provides flexibility to separate different fragment sizes. 2. It shows higher miRNA binding efficiency. 3. It enables users to keep all fractions of RNA species during miRNA enrichment. Beckman Coulter, the stylized logo, SPRI, SPRIselect are registered trademarks of Beckman Coulter, Inc. All other trademarks are the property of their respective owners.

1709T

Standardizing High-Throughput Sequencing of Extracellular RNA from Human Plasma. *Y.E. Wang[1], K.M. Danielson[2], R. Rubio[1], S. Das[2].* 1) Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, MA; 2) Cardiovascular Institute, Beth Isreal Deaconess Medical Center, Boston, MA.

Background. Extracellular vesicles have been shown to regulate intercellular signaling by transmitting RNA materials such as mRNA, microRNA and snRNA. This phenomenon implies extracellular RNA (exRNA) may partially reflect cellular content within the human body and may show disease specific variation. As such, exRNA profiles have great potential as disease biomarker. While high-throughput sequencing technology offers a potentially sensitive means to characterize and quantify exRNA, there is a lack of a understanding of the efficacy and reliability of commercial kits for extracellular RNA sequencing. In this study, we aim to optimize protocols for sequencing extracellular RNA from human samples. **Methods.** Blood plasma samples were obtained from one healthy individual at a single time point. The samples were frozen and thawed to mimic clinical sample condition. The samples were treated as followed: a) RNA purification by Exiqon (miRcury) column; b) proteinase K treatment preceding Exiqon purification; c) addition of carrier glycoblue during proteinase K treatment/Exiqon purification. Sequencing libraries were generated using random hexamer based, SMART technology based, and smRNA specific library preparation kits. The resulting libraries were sequenced on a HiSeq 2000 on a Single Read 50 bp flowcell. The sequence data was mapped against the transcriptome using BWA and analyzed using miRDeep2 algorithm to quantify miRNA. **Results.** Approximately 10 million raw reads were obtained for majority of the samples but with variable mapping rates across library preparation protocols. The variation is less across different sample treatment process within a protocol. The NEB protocol resulted in the highest number of miRNA species identified in all three conditions. Proteinase K and glycoblue treatment appear to enhance the number of identifiable miRNA species. We then compared the sets of miRNA species derived from different library preparation and found large differences between them with some overlaps. This suggests miRNA species detection is confounded by the sequencing protocol used. **Conclusion.**In this study, sample treated with proteinase K treatment prior to vesicle lysis and prepared by smRNA specific protocol resulted in the best outcome. However, with sample preparation and sequencing protocol exerting strong impact on extracellular miRNA quantification and detection, additional studies is necessary to optimize the process.

1710S

RNA "SEQing" answers in the blood transcriptome: Benchmarking methods for globin message reduction. *N. Allaire, A. Day-Williams, S. Szaks, C. Sun, J. Carulli.* Department of Genetics and Genomics, Biogen Idec Inc., Cambridge, MA.

RNA sequencing (RNAseq) is gradually replacing micro-array technologies as the method of choice for transcriptional profiling studies. Greater resolution and lower processing costs are two important factors that are driving this trend; however technical limitations exist that have prevented more widespread adoption. RNA seq from whole blood represents a particular challenge in that ~ 50% of the transcriptome is encoding by globin. This issue of globin overabundance reduces detection of informative transcripts and increases processing and storage costs. Recently, several methods have been developed to address the problem of excessive globin; however no comprehensive comparison has been reported. To this end, we undertook a systematic and rigorous evaluation of three promising approaches; Insert dependent adapter cleavage (InDAC), globin removal prior to sequencing, and ultra-deep sequencing and report our findings here. Both InDAC and Globin removal reduced globin reads by from ~50% to 2% compared to untreated controls. Off target effects were few and linearity was maintained in the samples that had globin removed prior to sequencing. InDAC resulted in more detectable off-target sequences and slight compression of dynamic range, however all three approaches were largely correlated when compared in fold change space.

1711M

Very Low Input RNA-Seq is Enabled by Digital Microfluidics. *T.M. Hill, L.C. Watson, S.M. Gross, I. Khrebtukova, F. Schlesinger, S. Pathak, T. Singer, G.P. Schroth.* Illumina 5200 Illumina Way San Diego, Ca, 92122.

Recent advances in NGS library construction utilizing a digital microfluidics platform has enabled us to significantly lower the sample input requirements into TruSeq Stranded mRNA preparation workflow. We find that the reduced reaction volume of our digital microfluidics platform has greatly improved the kinetics of each enzymatic step. In addition to improved kinetics, the enclosed automated fluidics of this system facilitates improved sample recovery and yield of each step in the process. As a result of these efficiencies, it is possible to prep high quality RNA-Seq libraries even when starting with only nanogram levels of total RNA. We will show detailed sequencing results generated from libraries made with inputs of 100, 10, and 1 ng, or less, of total RNA using this new technology. Even at these levels, we observe a large dynamic range of gene expression, and see results that are typically observed with 100 ng of total RNA utilizing the standard bench protocol. Furthermore, the improved sensitivity of the system is demonstrated by detecting equivalent number of genes in 10 ng of RNA processed on the microfluidics platform versus 100 ng of RNA processed with the standard bench protocol. The digital microfluidics platform enables researchers to access sample types with limited RNA recovery such as RNA extracted from small numbers of isolated cells or fine needle aspirates (FNA) resulting in a greater understanding of gene expression at the cellular level. Finally, this microfluidics technology is very easy-to-use and automated.

1712T

Analysis of PCR duplicates and Library Diversity in RNA-Seq studies using very low input and degraded samples. *S. Pathak, I. Khrebtukova, S. Gross, F. Schlesinger, T. Hill, L. Watson, R. Kelley, T. Singer, G. Schroth.* Illumina, Inc., San Diego, CA.

In DNA sequencing, duplicates or reads that map to the same position are discarded but in RNA sequencing (RNA-Seq), these reads can represent highly expressed genes. The issue of duplicates in RNA-Seq is even more complicated in low input or degraded samples. Higher percentages of duplicates in very low input and degraded samples are routinely observed in RNA-Seq using standard bioinformatics tools such as picard but the source of duplicates is commonly misunderstood. Under normal assay conditions and with recommended input levels, three different RNA-seq assays give different apparent numbers of duplicates on the same UHRR and Brain samples: <20% duplicates for targeted RNA seq, <10% duplicates for mRNA-Seq and <6% duplicates for Total RNA-seq. These differences are not necessarily due to PCR artifacts but occur because of the differences in complexity between the coding regions, the mRNA, and the total RNA of a cell. When we measure true PCR duplicates using a molecular bar coding approach, it becomes clear that there are in fact much lower levels (much less than 1%) of potential PCR duplicates in standard RNA-Seq preps. However, we find that when reducing DNA input amounts for any of these three assays to 1ng or less, we observe dramatic increases in percentage of duplicates. This value then becomes an important metric for overall efficacy of the experiment.

1713S

SureSelect Clinical Exome Panels for NGS Research Applications. *E. Lin, M. Corioni, F. Useche, M. Issakson, D. Lande, B. Novak, A. Vadapalli, C. LeCoq, C. Pabon, M. Visitacion, D. Roberts.* Genomics, Agilent Technologies, Santa Clara, CA.

Protein coding genes constitute approximately 1% of the human genome but harbor most of the disease-associated variants. By focusing on only the protein-coding regions of the human genome, scientists are now able to more efficiently detect these variants. Targeted re-sequencing enables highly sensitive and comprehensive detection of variants and provides insights into the biology behind a given phenotype. With recent major advances in this technology, combined with a better understanding of biological pathways, NGS is now being considered for use in clinical research. To efficiently and comprehensively catalogue variants within a sample, we developed a new set of SureSelect target enrichment research panels optimized for analysis of disease-associated regions. 1) Our latest Clinical Research Exome targets the full human exome for broad research applications. It is specifically designed to achieve excellent coverage of disease-associated content, especially those relevant to constitutional disease research. It provides deep and comprehensive coverage of genomic content derived from highly curated databases including CCDS, RefSeq, GENCODE, Vega, MirBase, UCSC known genes, Human Gene Mutation Database (HGMD®), ClinVar and Online Mendelian Inheritance in Man (OMIM™); 2) The Focused Clinical Research Exome narrows the capture regions to only include ~5000 targets with strong association to disease and annotated within HGMD®, OMIM™ and ClinVar; 3) The Inherited Disease Panel further narrows the focus to cover only a set of ~2700 genes within OMIM™ that are highly relevant targets for Mendelian disease research. These three exome options allow researchers to maximize efficiency depending on their application space and sequencing platform. Here we demonstrate capture efficiency for these exome designs and describe sequencing coverage statistics as well as sensitivity and concordance for SNP determination. Additionally we show low-input SureSelect workflows that enable greatly reduced hybridization and turn-around times. Further, we provide data analysis and visualization using Agilent's SureCall software.

1714M

Flexible Content TaqMan® Pathway Panels. *M. Laig[1], J. Clarke[1], K. Lee[1], M. Augustine[2], E. Shelton[1], K. Varma[1], P. Farahani[1].* 1) Thermo Fisher Scientific, South San Francisco, CA; 2) Ceino Technologies Inc., Menlo Park, CA.

Identifying patterns of gene expression and SNP alleles related to a pathway or disease can help identify specific genes of interest. Fixed content panels are a good first approach, but there are frequently additional genes pertinent to the pathway. To this end, we are developing customizable TaqMan® assay panels as flexible research tools for pathway and disease related genes and markers. The panels contain assays for a suggested list of target genes and SNP markers that are customizable to omit and add assays of interest. The researcher can choose between three platforms: TaqMan® OpenArray®, TaqMan® Array Card and TaqMan® Array Plate. The initial panels are gene expression assay sets covering 15 popular research areas in human, mouse and rat: Alzheimer's disease, Cardiotoxicity, Hepatotoxicity, Nephrotoxicity, Neurotoxicity, Inflammation, Diabetes, Cell cycle, Apoptosis, Hypertension, Angiogenesis, Glycosylation, Parkinson's disease, Cardiovascular disease and p53 Signaling. Each panel contains up to 92 markers as a core assay set, and some panels contain additional assays in the "Off Your Array" list that may be used to substitute core assays. Additional assays of interest may be added, pre-designed or custom assays. The repertoire of panels will be expanded to include additional gene expression panels and SNP genotyping panels. Flexible Content TaqMan® Pathway Panels enable a better selection of assays for the immediate research interest.

1715T

Design and implementation of a transplantation-targeted whole genome genotyping array. *A. Shaked[1], Y.R. Li[2,3], F.W. Asselbergs[4], S. Bala[5], K. Karczewski[6], K. Birdwell[7], W. Oetting[8], P. de Bakker[9], P. Jacobson[8], D. MacArthur[6], Y. Lu[11], G. Lord[10], A.K. Israni[8], B.J. Keating[1,2].* 1) Dept of Transplantation,University of Pennsylvania, Philadelphia, PA, USA; 2) The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA;; 3) Medical Scientist Training Program; Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA;; 4) Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, Utrecht, Netherlands; 5) Yale Medical School; 6) Harvard Medical School and Broad Institute, Massachusetts General Hospital, MA, USA; 7) School of Medicine, Vanderbilt University; 8) College of Pharmacy, University of Minnesota; 9) Epidemiology and Bioinformatics at UMC Utrecht; 10) Dept of Nephrology, Transplantation and Internal Medicine, Guy's and St. Thomas' Hospital and King's College London, UK; 11) Affymetrix, Santa Clara, California.

More than 559,000 solid organ transplantations (kidney, liver, heart and lung) have been performed in the United States since 1988, with nearly 30,000 occurring annually in the US. Despite significant improvements in transplant outcomes, chronic organ rejection affects up to 60 percent of patients representing a major cause of morbidity and mortality following allogeneic transplantation. Whole genome sequencing projects have identified millions of common and rare genetic polymorphisms across human populations, which are thought to contribute to genetically-associated risk factors for transplant and immunosuppression-related adverse event, as well as serve as a source of genetic disparity between donors and recipients that can lead to both acute and chronic organ rejection. Array-based genotyping technologies permit significant flexibility in choosing the scope and density of SNPs to examine for disease or phenotype-specific genome-wide association studies. We report and discuss the design, implementation, and preliminary results from thousands of samples typed on a state-of-art transplantation-focused SNP array utilizing an Affymetrix custom genotyping platform composing of nearly 800,000 markers with coverage of a number of genomic features of interest in particular to the transplant community: 1. Exonic and Loss of Function (LoF) variants (274K) from >40K human exomes; 2. Imputation panel (463K) with all available EUR ancestry data from the 1000 genome project and ~90K African ancestry panels SNPs to boost coverage and improve fine-mapping. We added significant content (55K) for UTRs, pharmacogenomic markers (7.5K) from PharmaADME.org, PharmaGKB, and the candidate studies in the literature related to immunosuppression therapy response, and copy number variant probes (16.4K) from the Wellcome Trust and UKbiobank (n=2.4K) and analysis of 68K GWAS samples from CHOP; 5. Expression QTL markers (17.4 K SNPs) from NCBI/NIH GTEx eQTL database; 6. Dense coverage across the MHC and KIR regions; and 7. Cardiovascular & metabolic disease related ontent for analysis of new onset of diabetes after tx (NODAT) & Cardiac allograft vasculopathy (CAV) by collation of SNPs from >600 PubMed manuscripts. This genotyping array offers rich information for the transplant community utilizing state-of-the-art array design technology and data from large-scale resequencing projects and publicly available functional and gene expression dataset.

1716S

Performance of seven mutation pathogenicity prediction methods in the classification of missense variants of the CYP1B1 gene. *G. Chavarria-Soley.* Biology, University of Costa Rica, San José, Costa Rica, San José, Costa Rica.

Nonsynonymous single nucleotide polymorphisms (SNPs) in the coding regions of genes can lead to aminoacid changes and potentially affect protein function and, therefore, susceptibility to disease. Several computational methods have been developed for the classification of SNPs according to their predicted effect on protein function and resulting pathogenic potential. In this study, we evaluated the performance of seven commonly used pathogenicity prediction methods available on the Internet (SIFT, nsSNPAnalyzer, Panther, pMut, PolyPhen, PhD-SNP, and SNAP). In order to test them, nonsynonymous SNPs in the CYP1B1 gene- which codes for the cytochrome P450 1B1 enzyme-were selected. A total of 129 missense variants in CYP1B1 were identified in the literature, from which 87 could be classified as pathogenic or neutral according to criteria such as segregation with disease phenotype, effect on function, among others. The algorithms showed significant variation in the assignment of the variants to three categories (non-neutral, neutral, no prediction), with a low 37% prediction rate for Panther. Pairwise concordance between methods in the classification of variants as pathogenic or neutral varied between 37% and 94%. The accuracy in the prediction of the pathogenicity of the variants was higher than 68% with all methods except pMut (47%). The highest false positive and false negative rates were found for SIFT and pMut, respectively. Taking into account the rate of prediction, accuracy of prediction, false positive, and false negative rates, the method with the overall best performance in the present study was nsSNPanalyzer, closely followed by SIFT, Polyphen and SNAP.

1717M

A Comprehensive IT System to Support GTEx Biospecimen Collection Operations. *P. Guan[1], C. Shive[2], L. Qi[2], D. Tabor[2], P. Hariharan[2], S. Wu[2], K. Um[2], V. Santhanam[2], P. Kigonya[2], J. McLean[2], J. Vaught[1], H.M. Moore[1] on behalf of GTEx consortium.* 1) National Cancer Institute, Bethesda, MD; 2) Leidos Biomedical Research Inc., Frederick, MD.

The NIH Common Fund's Genotype-Tissue Expression (GTEx) project aims to study gene expression and regulation across multiple human tissues (30+ tissue types) from approximately 900 healthy normal donors. It is expected to provide valuable insights into gene regulation and its tissue specificity, to identify correlation between genetic variations and variations in gene expression levels as expression quantitative trait loci (eQTLs), and to help understand inherited susceptibility to diseases. To meet the challenge of GTEx requirements for collecting and tracking high quality biospecimen samples, a custom-built software system, the Comprehensive Data Resource (CDR) was developed to support sample collection workflow, clinical data entry, case management, and review and curation of study data. CDR is built with a combination of technologies from Grails, Oracle, Groovy, jQuery, and Apache Solr. CDR provides secure data access based on pre-defined user roles and privileges. Personally Identifiable Information (PII) and Protected Health Information (PHI) are restricted to a limited data set (LDS) and to authorized users through dynamic content redaction. Intuitive graphic user interfaces for Biospecimen Source Sites streamline data entry workflow by following detailed SOPs for sample collection and processing. Automated data check and business rule validation confirm data integrity and SOP adherence simultaneously. Web service API's allow a Pathology Resource Center to access digital imaging data housed remotely at a Comprehensive Biospecimen Resource (CBR). API's connect to the CBR's LIMS system for real-time sample inventory data. De-identified data is provided to the Laboratory Data Analysis and Coordinating Center (LDACC) at the Broad Institute through a private API before the final release into dbGaP. CDR's analytics and reporting module supports data analysis and aggregation, report generation and real-time operational data snapshots. CDR is a distributed web-based bioinformatics system that has supported GTEx biospecimen operations from an earlier pilot phase to the current scale-up stage. CDR manages and maintains multi-dimensional data models around each donor case (average 500+ data elements per case). As an efficient case management tool capable of connecting to various remote informatics systems, CDR could be adapted by the broader research community to standardize and streamline biobanking operations.

1718T

Clinical phenotype-based gene prioritization using semantic similarity and the Human Phenotype Ontology. *A. Masino[1], E. Dechene[2], M. Dulik[2,3], A. Wilkens[2], N. Spinner[3,4,5], I. Krantz[2,5], J. Pennington[1], P. White[6,7].* 1) Center for Biomedical Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 4) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 5) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 6) Department of Pediatrics, Cincinnati Children's Hospital and Medical Center, Cincinnati, OH; 7) Department of Biomedical Informatics, University of Cincinnati, College of Medicine.

Exome sequencing is a promising method for diagnosing patients with a complex phenotype. However, variant interpretation relative to patient phenotype can be challenging in some scenarios, particularly clinical assessment of rare complex phenotypes. Each patient's sequence reveals many possibly pathogenic variants that must be individually assessed. Variant impact to gene function alone is insufficient to determine clinical relevance. Clear association between a gene harboring a damaging variant and patient phenotype is also required for accurate clinical diagnosis. We implemented an algorithm that can assist this process by ranking a given set of genes relative to patient phenotype. The model orders genes by the semantic similarity between Human Phenotype Ontology terms associated with a gene and those describing the patient. Model validation was performed for 33 Mendelian diseases with 100 simulated baseline patients per disease. Each patient was assigned phenotypic features with probability based on published feature penetrance. To measure robustness, we modeled likely clinical conditions by adding noise, i.e. terms unrelated to the disease, and imprecision, i.e. terms less specific than the actual disease terms, to each patient. For optimal and noise conditions, the disease gene median rank was 1 of 2488. We also provisionally examined a clinical cohort of subjects with hearing impairment. The disease gene median rank was 22 of 2488. However, considering the patient's exome data and filtering non-exomic and common variants, improved median causative gene rank to 3 of 125.5. We have demonstrated that the addition of phenotypic information can substantially improve the ability to prioritize genes, and by extension variants, for clinical characterization. With continued research, we expect that automated gene prioritization based on phenotypic descriptors may increase accuracy and decrease effort for predicting genomic variant significance.

1719S

The Orphanet Rare Diseases Ontology (ORDO) : a reference tool integrating clinical and genetic data. *A.M. Rath[1], A. Olry[1], C. Gonthier[1], L. Chanas[1], H. Parkinson[2], J. Malone[2], D. Vasant[2], M. Hanauer[1], B. Urbéro[1], S. Aymé[1].* 1) US14, INSERM, Paris, Paris, France; 2) EMBL-EBI, Hinxton, UK.

The growing complexity of genetic knowledge, the scattered phenotype data in clinical databases, as well as the multiplicity of medical terminologies in use impose the need for a reference tool in which all these data are integrated in a normalised fashion in order to render them accessible for health information systems and for research. Since 1997, Orphanet maintains a multilingual database of rare diseases (RD), based on the available literature and on expert advice. This data is manually curated, comprised of a nosology (classification of RD), relationships (genes-diseases, epidemiological data, orphan drugs) and cross-references with other terminologies (MeSH, SNOMED CT, UMLS), databases (OMIM) or classifications (ICD10) in use. Genes are cross-referenced with other scientific databases (HGNC, OMIM, UniProt, Genatlas, Reactome, ensembl, IUPHAR). These data are already freely available for download via the OrphaData platform (www.orphadata.org). However, it is necessary to make data available for researchers in a machine-readable format, ready to be integrated to any technical environment (obo and owl formats). In order to achieve this, Orphanet set up a collaboration with the European Bioinformatics Institute (EBI, Hinxton, UK) in order to produce the Ontology of Rare Diseases, which provides a robust and consistent modeling of data and their semantic relationships, as well as interoperability standards with other scientific resources in use in both research and in public health. The Orphanet Rare Diseases Ontology is available on BioPortal (http://bioportal.bioontology.org/ontologies/ORDO) and OrphaData and is updated monthly.

1720M

Comprehensive Transcriptome Analysis Reveals that Nonsense-Mediated mRNA Decay Is Not Globally Suppressed in Lung Adenocarcinomas. *L. Hu, P. Zhang.* State Key Lab of MedicalGenomics, SIBS, Shanghai, China.

The nonsense-mediated mRNA decay (NMD) pathway ensures the rapid degradation of mRNAs containing premature translation termination codons (PTC). Prior research has shown that the tumor microenvironment, which exhibits hypoxia and amino acid deprivation, might inhibit NMD activity. Using next-generation RNA-seq data of cancerous tissue, we show that, in spite of several important differences in the expression of NMD factors and NMD targets, such as down-regulation of SMG6, cancer cells probably manipulate the NMD mechanism for their own benefits rather than globally inhibiting the process. Cancer cells strategically prefer different degradation targets that might confer a growth advantage in the tumor microenvironment. We also found that genes containing NMD targets tend to have high expression levels and low expression noise. Thus, we believe that cancer efficiently uses the NMD mechanism.

1721T

Combined use of mutant loxP sites, JT15 and JTZ17, is a useful approach for sophisticated genome engineering. *K.C. Chen[1], P.L. Chen[1,2,3,4].* 1) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 4) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan.

Cre-mediated site-specific integrative recombination system is a useful tool for genome engineering. Integrative recombination with a pair of mutant loxP sites is a developed and successfully strategy to increase the diversity and possible combinations. Using several mutant loxP sites combination system for manipulation of multiple genes on knock-out or knock-in experiments can improve complex genetic studies. In 2003, Piedrahita *et al.* (*Genesis*, 2003, 36:162-167) reported a couple of mutant loxP combination, JT15 and JTZ17, which yielded the most efficient integration in E. coli. However, their recombination efficiency with wild-type loxP or other mutant loxP sites remains unclear. In order to address this issue, we constructed a series of neomycin selection cassette flanked with the wild-type or mutant loxP sites to test the frequency of site-specific integration. Our results showed that the frequency of targeted recombination between JT15 and JTZ17 was 74%. We also observed that the recombination between JT15 and the JT15:JTZ17 mutant loxP sequence was very low. There was no recombination between JTZ17 and the JT15:JTZ17 mutant loxP sequence. Most importantly, JT15, JTZ17 and JT15:JTZ17 could not recombine with wild-type loxP sites. In conclusion, we suggest that combined use of JT15 and JTZ17 is a useful approach in sophisticated genome engineering strategies.

1722S

Identification of rare causal variants in sequence-based studies: methods and applications to VPS13B, a gene involved in Cohen syndrome and autism. *I. Ionita-Laza[1], M. Capanu[2], S. De Rubeis[3], K. McCallum[1], J. Buxbaum[2].* 1) Biostatistics, Columbia University, New York, NY; 2) Memorial Sloan-Kettering Cancer Center; 3) Icahn School of Medicine at Mount Sinai.

Pinpointing the small number of causal variants among the abundant naturally occurring genetic variation is a difficult challenge, but a crucial one for understanding precise molecular mechanisms of disease and follow-up functional studies. We propose and investigate two complementary statistical approaches for identification of rare causal variants in sequencing studies: a backward elimination procedure based on groupwise association tests, and a hierarchical approach that can integrate sequencing data with diverse functional and evolutionary annotations for individual variants. Using simulations, we show that incorporation of multiple bioinformatic predictors of deleteriousness, such as PolyPhen-2, SIFT and GERP++ scores, can improve the power to discover truly causal variants. As proof of principle, we apply the proposed methods to VPS13B, a gene mutated in the rare neurodevelopmental disorder called Cohen syndrome, and recently reported with recessive variants in autism. We identify a small set of promising candidates for causal variants, including two loss-of-function variants and a rare, homozygous probably-damaging variant that could contribute to autism risk.

1723M

Effect of Haplotype Estimation in Exact Tests for Association. *L. Ehwerhemuepha, A. Smith, C. Rakovski.* Chapman University, Orange, CA.

We compared the loss of power of three exact tests for association (Multinomial, Fisher's and Barnard's Exact Tests) due to haplotype frequency estimation in unphased data. Two methods for estimating haplotype frequencies were used: An EM algorithm implemented to determine weighted haplotype assignments of all subjects based on the haplotype pairs compatible with the unphased genotypes; and a Bayesian haplotype reconstruction method based on Bayesian Inference. The EM algorithm allows for population-level haplotype frequency estimation while the Bayesian counterpart allows for individual-level haplotype estimation from which the population haplotype frequency can be calculated under the assumption that each individual's estimated haplotype frequency is its true haplotype frequency. We performed extensive simulations (at several alpha levels) to assess the type I error rates under various null hypothesis and the power of the test under various alternative hypothesis.

1724T

Bootstrap Tests of Association For NextGen Sequence Data That Allow for Systematic Differences in Read Depth between Cases and Controls. *G.A. Satten[1], Y. Hu[2], H.R. Johnston[2], P. Liao[2], Y. Jiang[3], A.S. Allen[3,4].* 1) Centers for Disease Control and Prevention, Atlanta, GA; 2) Dept of Biostatistics and Bioinformatics, Emory University, Atlanta, GA; 3) Dept of Biostatistics and Bioinformatics, Duke University, Durham, NC; 4) Duke Clinical Research Institute, Duke University, Durham, NC.

Background: The quality of genotype calling for next-generation sequence data depends on read depth. Loci with high coverage can typically be called reliably, while those with low coverage may be difficult to call. In a case-control study, if data from case participants is sequenced to a greater depth than data from controls, the difference in genotype quality can introduce a systematic bias. This can easily occur when historical controls (e.g., data from The 1000 Genomes Project) are used. This imbalance may also occur by design, to reduce genotyping costs among controls. For trios, bias can arise even when the coverage is the same in parents and offspring since errors in parental genotype calls are considered non-transmissions while errors in offspring genotype calls are detected as non-Mendelian transmissions.

Methods: We develop likelihood-based methods for analyzing data from case-control and trio studies that directly uses data on reads without first making intermediate genotype calls. When the location of polymorphic loci is known, we show these likelihood approaches have appropriate size and good power compared with methods that use called genotypes. When the locations of polymorphic loci are not known in advance, we develop screening methods to screen out loci that are estimated to be monomorphic, based on read data alone. We use a bootstrap approach to estimate which of the loci that screen in are truly polymorphic. Using these estimates, we then construct bootstrap tests for association that properly account for screening and preserve size. We further show that restricting to loci with estimated allele frequency $\geq 1/2N$, so that the expected number of alleles seen is greater than one, increases the power of our approach by excluding loci that have negligible effect.

Results: We illustrate our approach using data from the UK10K project. We use data from 784 cases from the Severe Childhood Onset Obesity Project, and are exome sequenced at 60x. Data for 1702 controls are from the Avon Longitudinal Study of Parents and Children and the TwinsUK study (only one twin used), and are whole genome sequenced at 6x coverage.

1725S

Binary Trait Analysis in Sequencing Studies under Trait-Dependent Sampling. *Z.Z. Tang[1], D.Y. Lin[2].* 1) Vanderbilt University, Nashville, TN; 2) University of North Carolina, Chapel Hill, NC.

Trait-dependent sampling design has been adopted in many sequencing studies to enrich causal signal and increase power of the association tests. In the National Heart, Lung, and Blood Institute Exome Sequencing Project (NHLBI ESP), subjects with the highest or lowest values of body mass index, low-density lipoprotein, or blood pressure were selected for whole-exome sequencing. However, such study design produces challenges for genetic association analysis. We provide a valid and efficient maximum likelihood framework for analyzing binary traits under the trait-dependent sampling. Under this framework, we produce the commonly used gene-based association tests including burden tests, variable-threshold tests and variance-component tests. We compare our methods with the naive methods, namely the standard logistic regression methods. We demonstrate through extensive simulations that our methods preserve type I errors, whereas naive methods can yield severely inflated type I errors. For a particular trait of interest, our approach properly combines the association results from all studies with measurements of that trait. This meta-analysis is substantially more powerful than the analysis of any single study. By contrast, meta-analysis of naive methods can be less powerful than the analysis of a single study. The usefulness of the proposed methods is further illustrated with data from NHLBI ESP.

1726M

Functional regression for genetic association studies. *O. Vsevolozhskaya[1], D. Zaykin[2], L. Qing[1].* 1) Epidemiology & Biostatistics, Michigan State University, East Lansing, MI; 2) Biostatistics Branch, NIEHS, Research Triangle Park, NC.

We propose a general framework to perform gene/region based analysis of sequencing data by regressing a functional response on one or multiple scalar predictors. Next generation sequencing technologies make it possible to uncover genetic information from millions of variants. Since the observed sequenced variants are very close in their genetic positions, we can consider them to be realizations of random continuous functions. Therefore, instead of analyzing multiple individual genetic variants per subject, we can estimate the underlying continuous function and treat it as a functional response in a regression model. Smoothing splines are used to fit these functional responses by maximizing the penalized likelihood. Covariates can also be incorporated in the analysis to control for confounding, including qualitative and quantitative predictors. By utilizing a connection between penalized spline regression and linear mixed models, we are able to fit our model using standard linear mixed models statistical packages. To illustrate our approach, we conduct simulation studies and apply our proposed methodology to sequencing data from the Dallas Heart Study.

1727T

A Generalized Similarity U test with application to multiple-trait sequencing association study. *C. Wei[1,2], Q. Lu[2].* 1) Department of Biostatistics and Epidemiology, University of North Texas, Fort Worth, TX; 2) Department of Epidmiology and Biostatistics, Michigan State University, East Lansing, MI.

Sequencing-based studies are emerging as a major tool for genetic association studies of complex diseases. It also poses great challenge to the traditional statistical methods (e.g., the single variant analysis) due to the high-dimensionality of the data and low frequency of the genetic variants. Joint test has been shown to be more suitable for sequencing studies, by jointly testing multiple variants to increase the power and reduce the dimensionality. Meanwhile, there are growing needs for statistical methods are distribution free and can handle multiple phenotypes. In this paper, we proposed a generalized similarity U test, referred to as GSU. GSU rst summarizes the genetic information and multiple traits into the genetic similarity and trait similarity, and then combines the two similarities in the framework of weighted U statistic. We derived the asymptotic distribution of GSU under null hypothesis so as to eciently calculate the signicance level. We also studied the asymptotic behavior of GSU under alternative and provide sample size and power calculation for study design. To evaluate the performance of GSU, we conducted extensive simulation studies and compare it with the existing methods. Through simulation, we found GSU had advantage over existing methods in terms of power comparison and robustness to trait distribution. Moreover, GSU is computationally more ecient than the existing methods. Finally, we applied GSU to the multiple traits analysis of Dallas Heart Study and identied joint association of 4 genes with 5 metabolic related traits.

† Trainee Award Finalist

1728S

Utilizing Private Variants in Large Genome-Wide Association Studies: Issues, Techniques, Experiences. *U. Bodenhofer, S. Hochreiter.* Institute of Bioinformatics, Johannes Kepler University, Linz, Austria.

High-throughput sequencing technologies have facilitated the identification of large numbers of single-nucleotide variations (SNVs), many of which have already been proven to be associated with diseases or other complex traits. Several large sequencing studies, such as, the 1000 Genomes Project, the UK10K project, or the NHLBI-Exome Sequencing Project, have consistently reported a large proportion of private SNVs, that is, variants that are unique to a family or even a single individual. The role that private SNVs play in diseases and other traits is currently poorly understood — which is largely due to the fact that it is statistically very challenging to consider private SNVs in association testing. While it is generally impossible to use single-marker tests for private SNVs, burden tests are potentially able to deal with private SNVs, but only if the number of private SNVs occurring in a region is correlated with the trait under consideration. Moreover, burden tests have a disadvantage if deleterious and protective SNVs occur together in the same region. Non-burden tests like the popular SNP-set (Sequence) Kernel Association Test (SKAT) are typically utilizing correlations between SNVs — a strategy that is not applicable to private SNVs either, since singular events are generally uncorrelated. We propose the Position-Dependent Kernel Association Test (PODKAT), which is designed for detecting associations of very rare and private SNVs with the trait under consideration even if the burden scores are not correlated with the trait. PODKAT assumes that, the closer two SNVs are on the genome, the more likely they have similar effects on the trait under consideration. This assumption is fulfilled as long as deleterious, neutral, and protective variants are grouped sufficiently well along the genome. This contribution focuses on the use of PODKAT for large whole-genome studies. On the one hand, we will discuss issues related to data handling, computational complexity, and statistical significance. On the other hand, we will present results obtained for UK10K whole-genome cohorts that unveil the potential of considering private and very rare SNVs in genome-wide association studies.

1729M

A non-threshold region-specific method for detecting rare variants. *D.P. Chen[1], A.R. Hsieh[2], C.S.J. Fann[1].* 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 2) Graduate Institute of Biostatistics, China Medical University, Taichung, Taiwan.

Rare variants have a proven role in some complex diseases. Many statistical methods proposed for the detection of rare variants associated with diseases have some limitations, such as the threshold of rare variants, and the direction of effects. Accordingly, we developed a region-specific method that do not use the threshold for defining rare variants and take the directions of effects into account. Our method also considers the linkage disequilibrium (LD) within the region, and can handle common and rare variants simultaneously. Our region-specific method used the concept of weighting variants according to their minor allele frequencies and odds ratios (OR) to combine effects of common and rare variants on disease occurrence into a single score, and provided a test statistic in assessing the significance of the score. To evaluate the performance of our method, we simulated extensively under different effect sizes according to Basu and Pan (2011). We found that the power of our method increased as the effect sizes increased. The type I error of our method was controlled well in spite of the simultaneous variations. Moreover, we compared our proposed method to several currently available methods, including kernel-based adaptive cluster (KBAC) and Sequence Kernel Association Test (SKAT). We found our method can generate comparable or better power in simulations. Results from our method showed a 15% increase in power comparing with SKAT (61% vs 47%) under small OR and lower LD, and 36% increase in power comparing with KBAC (98% vs 62%) while variants have different directions. However, our method performs well in 2-direction setting, but moderate in independent-variant scenario. We conclude that our proposed method can be used as a complementary tool with others to assist the dissection of the etiology of complex diseases.

1730T

Evaluating the calibration and power of three gene-based association tests for the X chromosome. *C. Ma, M. Boehnke, S. Lee.* Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

While genome-wide association studies (GWAS) have identified thousands of trait-associated genetic variants, the proportion of findings on the X chromosome lags behind those on the autosomal chromosomes. Existing X chromosome analysis methods focus on single marker association analysis. For analysis of rare variants (minor allele frequency < 0.5%), gene-based tests where multiple markers are analyzed jointly as a unit can be more powerful than single marker tests. To date, there are no gene-based tests designed to analyze the X chromosome. Using simulated case-control and quantitative trait (QT) studies, we evaluate the calibration and power of three gene-based tests for the X chromosome: burden, Sequence Kernel Association Test (SKAT), and optimal unified SKAT (SKAT-O). Specifically, we evaluate the impact of different ratios of males and females in cases and controls, and different coding of males alleles with X-inactivation (coding minor alleles as X=2) and without (X=1).

For case-control studies, all three tests are well-calibrated or slightly anti-conservative for all scenarios evaluated. As previously shown, power of the three tests depends on the underlying genetic architecture of the genomic region analyzed; burden is most powerful for multiple causal variants with the same direction of effect, SKAT is most powerful for causal variants with opposite directions of effect, and SKAT-O is generally powerful. For variants simulated assuming X-inactivation, coding male minor alleles as X=2 is slightly more powerful; for variants simulated assuming no X-inactivation, coding male minor alleles as X=1 is slightly more powerful. However, the power loss for misspecifying the generally unknown model is small. Different ratios of males and females in cases and controls have little effect on power. For QT studies, burden and SKAT are well-calibrated, while SKAT-O can be slightly anti-conservative across all scenarios. Power comparisons between tests for QTs are very similar to those for binary traits. We demonstrate that these three gene-based tests are well-calibrated and powerful for both binary and quantitative trait data, and can be directly applied to analyze rare variants on the X chromosome.

1731S

Exploiting correlation of genetic effects in rare variant association studies. *M.A. Rivas[1], M. Pirinen[2], L. Moutsianas[1], C. Spencer[1], K. Banasik[1], D. van Heel[3], K. Hunt[3], P. Soininen[4], A.J. Kangas[4], M. Ala-Korpela[4,5,6], M.J. Daly[7,8], F. Karpe[9], P. Donnelly[1,10], M.I. McCarthy[1,9].* 1) Wellcome Trust Centre for Human Genetics Research, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK; 2) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 3) Barts and The London School of Medicine and Dentistry, Queen Mary University of London; 4) Computational Medicine, Institute of Health Sciences, University of Oulu, FI; 5) NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, FI; 6) Computational Medicine, School of Social and Community Medicine and the Medical Research Council, Integrative Epidemiology Unit, University of Bristol, UK; 7) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, US; 8) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 9) Oxford Centre for Diabetes, Endocrinology, and Metabolism, University of Oxford, UK; 10) Department of Statistics, University of Oxford, Oxford, UK.

High-dimensional data present exciting opportunities in enhancing our understanding of the relationship between genetic variation and multiple diseases or health related quantitative traits. A first step towards utilizing high-dimensional phenotypes in genetic studies is to understand how their genetic components are related. A second step is to apply this knowledge in association studies.

Here we present novel statistical approaches that leverage estimates of the correlation of genetic effects for rare variant association studies that enable association testing (via extension of the C-alpha framework) and effects estimation in cross-disorder and cross-phenotype study designs. The intuition is simple: 1) for some groups of genetic variants, such as protein truncating variants, annotation suggests high correlation of effects; we want to exploit this by building models which incorporate information on the correlation between the effects of these variants on a single phenotype; 2) for any causal genetic variant that is pleiotropic, incorporation of the estimate of correlation of genetic effects obtained from genome-wide genotype data may improve power to detect association.

Simulations suggest that our approach improves power to detect association compared to standard univariate approaches. First, we consider the study of rare variants and multiple continuous traits. We apply the tests to coding rare variant data and plasma NMR metabolite levels in 4,522 healthy individuals from the population based Oxford Biobank study and identify novel genic signals that were not identified with univariate aggregate methods. As an example, we identify multivariate association ($p=1\times10^{-6}$) of protein-altering variants in *PAH* (phenylalanine hydrolase) to the amino-acid trait profile (univariate $\min(p) =4.5\times10^{-4}$). Second, we consider the study of rare variants, multiple diseases, and shared controls. We apply the tests to coding sequence data from 25 GWAS risk genes in 41,911 UK residents of white European origin, comprising 24,892 subjects with six autoimmune disease phenotypes and 17,019 controls and identify new association of protein truncating variants at *TNFAIP3*. Our framework is computationally efficient, making the analysis of large rare variant association study designs practical.

1732M

Integrated statistical model of genetic variation reveals new insights into the genetics of autism. *X. He[1], K. Roeder[1,2], B. Devlin[3], MJ. Daly[4,5], JD. Buxbaum[6,7,8,9], Autism Sequencing Consortium.* 1) Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA; 2) Department of Statistics, Carnegie Mellon University, Pittsburgh, PA; 3) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA; 4) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 5) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 6) Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY; 7) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY; 8) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 9) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY.

Analysis of de novo mutations has been effective in identifying susceptibility genes of developmental disorders such as autism. Recurrent de novo loss-of-function (LoF) mutations on the same gene generally provides strong evidence for the gene's involvement in risk. Recently we developed a statistical model, named TADA, that extends this approach by combining information on de novo LoF mutations, de novo missense mutations and inherited variants for the same gene (He et al, PLoS Genetics, 2013). In the current work, we improve the TADA model by refining the weights mapping onto different types of genetic variation. Under the new model, de novo LoF mutations contribute most to a gene's evidence, followed by de novo missense mutations, transmitted LoF mutations and transmitted missense mutations. We achieve this weighting strategy by using different prior distributions for the relative risks of different types of genetic variation. Using parameters derived from whole exome sequencing (WES) data of autism, we show through simulations that TADA is almost twice more powerful than using only de novo LoF mutations. Using TADA we analyzed a large WES dataset of autism, consisting of more than 2,000 parent-child trios and 7000 case-control samples. The data supported an estimate of more than 1,000 genes involved in risk for autism. We next used TADA to predict 33 high-confidence autism risk genes (FDR < 0.1) and 107 genes at a relaxed threshold (FDR < 0.3). We developed two strategies for estimating the relative risks of mutations, one based on counting genes with single de novo mutations vs. genes with multiple mutations and the other based on the different frequencies of mutations in affected male vs. female subjects. Both lead to an estimate of relative risk of about 20 for de novo LoF mutations in high-confidence genes and lower estimates for other classes of variants. A large fraction of the 107 genes are under strong evolutionary constraint, defined by a shortage of non-synonymous mutations relative to the rate of synonymous mutations. Other enrichment analyses identify important biological processes involved in autism (De Rubeis for the Autism Sequencing Consortium). Finally, a newly developed statistical method named DAWN was used to analyze the gene networks underlying autism. DAWN identified several subnetworks that are enriched with genes of high TADA scores and gene pairs of high co-expression in brain transcriptome data.

1733T

Meta-Analysis of rare variants association studies with multiple correlated traits. *X. Wang, X. Zhao, C. Huang.* P.O. Box 413, Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI.

For many diseases such as asthma, attention deficit hyperactivity disorder, or hypertension, genetic association studies are often conducted to test multiple correlated traits which are routinely measured and thought to be more proximal to the biological etiology of the clinical disorder. Testing multiple correlated traits can identify quantitative trait loci (QTL) shared between correlated traits in addition to trait-specific QTL. Meta-analysis as a cost-efficient powerful tool for combining distinct genome-wide association studies (GWAS) has been widely used to detect common variants for single trait in GWAS. However, existing meta-analysis methods may not be optimal or even not be applicable for detecting rare variants in sequencing data with multiple correlated traits due to allelic heterogeneity, extreme rarity of individual variants, and correlation among traits. We propose a general statistical framework for meta-analysis of gene-based rare variants association studies with multiple correlated traits. Due to low minor frequencies of rare variants, it is often unstable or not feasible to estimate regression coefficients of rare variants in a multi-marker regression model. Thus, we developed a score test to test the effect of an optimally weighted combination of variants in a gene in each study for a single trait, and perform meta-analysis for each single trait with Lee's meta-analysis method and obtain the p value for each gene for all the studies. Then, we use the Brown's method to combine p values for the multiple correlated traits and obtain the p value of the meta-analysis with the multiple correlated traits. There are three key advantages of the proposed method: 1) through aggregating score statistics, it circumvents the estimation of the regression coefficients of rare variants; 2) it is computationally efficient even for whole-genome analysis since p values can be calculated analytically; 3) it can decrease the false-positive rate by incorporating the extra information provided by the cross trait covariance. To investigate the performance of the proposed method, we conducted extensive simulation studies for type I error rates and power comparison. The multiple correlated traits meta-analysis method maintained appropriate type I error rates in all the simulation settings. Further simulation studies are ongoing to compare power of the proposed method with the single trait meta-analysis method proposed by Lee et al.

1734S

Haplotype length regression for identifying rare disease-predisposing variants. *S.P. Sajuthi, C.D. Langefeld.* Biostatistical Sciences and Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC.

Van der meulen and te Meerman (1997) proposed using haplotype sharing statistics (HSS) to identify disease-predisposing loci. The two hypotheses that underlie the HSS paradigm: 1) affected individuals will share not only the same mutation but also the surrounding haplotypes, 2) the length of the haplotype sharing will be affected by the age of the disease variant, with newer variants residing on longer haplotypes than older variants. The second hypothesis is motivated by fewer generations will result in fewer recombination and mutation events in the sample that contains the newer variant. Building upon Van der meulen and Meerman (1997), Allen and Satten (2007) and Lange and Boehnke (2004), we have developed a novel statistical method, the haplotype length regression (HLR) that leverages these two important concepts to identify genomic regions that contain more recent disease-predisposing variants. Formally, our method uses a generalized linear model framework to test for association between haplotype sharing length and disease trait. To investigate the performance of our HLR for rare variants, we simulated a diploid population using the SimuPOP software (Peng 2005). Generation 0 was composed of 993 independent individuals from the HapMap Phase 3 dataset. This initial population underwent random mating for 500 generations until it reached 50,000 individuals; simulation assumes a mutation rate of 10-8/bp/generation and recombination rate of 10-8/bp/generation. After 500 generations, the disease allele is introduced and then the population continues random mating for another 150, 300, 450, 600, 750 and 900 generations. A range of penetrance functions for a disease allele frequence of 0.005 were employed to simulate the disease trait given the disease locus genotype. We compared our approach with other haplotype sharing statistics and SKAT-O (Lee 2012). For a single risk variant within the window of analysis, the HLR and haplotype regression methods had greater power than the HSS or SKAT-O. Although not as large a difference, HLR continue to have greater power than the other methods when three rare risk variants within the analysis window were simulated. The HLR approach is appropriate for discrete and continuous traits. We are expanding the HLR method to family data. In summary, leveraging evolutionary characteristics of rare variants via haplotype sharing length methods can increase our power to detect rare variants.

1735M

A fast and powerful test of independent assortment with implications for the analysis of 'big data'. *V. Hager[1], W.C.L. Stewart[1,2,3].* 1) Nationwide Children's Hospital, 700 Children's Drive Columbus, Ohio 43205; 2) The Ohio State University Department of Pediatrics, 700 Children's Drive Columbus, Ohio 43205; 3) The Ohio State University Department of Statistics, 1958 Neil Avenue, 404 Cockins Hall Columbus, OH 43210.

Quantitative methods for the analysis of genetic sequences on related individuals continue to place a high premium on computational and statistical efficiency. As such, the most commonly used methods tend to incorporate as much information as possible as quickly as possible. However, much less attention is paid to the perhaps more important issue of statistical significance. For example, although most modern workstations can easily handle the computations associated with a large-scale, genome-wide linkage scan, the p-value for such a scan is either accurate but time consuming, or fast but inaccurate. To address this issue, we developed a test of independent assortment (TIA) that permits the accurate and rapid computation of p-values for large-scale genome-wide linkage scans. Furthermore, we show that our proposed test (1) has a limiting distribution under the null hypothesis of no linkage; (2) yields an approximate 300x speed-up over existing methods like MERLIN1; and, (3) has the same statistical power as the widely used Kong & Cox lod2. Overall, these attractive features should facilitate the analysis of 'big data' genetics and genomics through the meta-analysis of linkage, association, and gene expression data.

1736T

Generation of sequence-based data for pedigrees-segregating Mendelian or Complex traits. *B. Li, G.T. Wang, S.M. Leal.* Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

There is great interest in analyzing next generation sequence data that have been generated for pedigrees. However, unlike rare variant association methods for population-based data there are limited rare variant methods which have been developed to analyze pedigree data. One of the limitations in developing methods to analyze pedigree data is the availability of software to generate realistic pedigree sequence data with the variants generated either conditionally or unconditionally on qualitative or quantitative phenotypes. Rare variant pedigree simulated data is necessary to evaluate type I and II errors and also to compare the power of different methods. Therefore, we have developed RarePedSim (rare-variant pedigree-based simulator), a program to simulate pedigree-based gene-level genotype and phenotype data for complex and Mendelian trait rare variant studies given any user-specified pedigree structures. For complex traits rare variant association studies, RarePedSim can generate genotypes of founders based on site-specific variant information and non-founders according to principles of segregation. Site-specific variant information including allele frequencies, positions and functionalities can be derived from simulated sequences under realistic demographic and genetic models or obtained from real sequence data. RarePedSim can generate phenotypes based on a wide range of genetic etiology for both qualitative and quantitative complex traits, including logistic model of odds ratios, population attributable risk model and linear mean-shift model. For Mendelian disease rare variant analysis RarePedSim generates genotypes for individuals within the family given their phenotypes and mode of inheritance, e.g. dominant and recessive. Allelic heterogeneity between and within families and locus heterogeneity can also be simulated. Currently, RarePedSim is the only available program that can generate both genotypes and phenotypes for gene-based rare variants regardless of pedigree structure. The data generated by RarePedSim are in standard Linkage file format, ready to be used for a variety of purposes including evaluating type I error and power, for association methods including mixed model tests and parametric and non-parametric linkage analysis.

1737S

GenLib: an R package for the analysis of genealogical data. *M.-H. Roy-Gagnon[1,2], H. Gauvin[2,3], J.-F. Lefebvre[2], C. Moreau[2], E.-M. Lavoie[4], D. Labuda[2,5], H. Vézina[4].* 1) Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, ON, Canada; 2) CHU Sainte-Justine Research Center, Montreal, QC, Canada; 3) Department of Social and Preventive Medicine, Université de Montréal, Montreal, QC, Canada; 4) BALSAC Project, Université du Québec à Chicoutimi, Chicoutimi, QC, Canada; 5) Department of Pediatrics, Université de Montréal, Montreal, QC, Canada.

Founder populations play an important role in the study of genetic diseases. Their advantages often include access to detailed genealogical records. These genealogical data provide unique information for researchers in evolutionary and population genetics, demography and genetic epidemiology. However, analyzing large genealogical datasets require specialized methods and software. The GenLib software was originally developed to study the large genealogies of the French Canadian population of Quebec, Canada. These genealogies are accessible through the BALSAC database, which contains over 3 million records covering the whole province of Quebec over four centuries. Using this resource, extended pedigrees of up to 17 generations can be constructed from a sample of present-day individuals. We have implemented GenLib as a package in the R environment for statistical computing and graphics, thus allowing optimal flexibility for users. GenLib includes functions to manage genealogical data, for example extracting part of a genealogy or selecting specific individuals. Functions to describe genealogies using relevant summary measures, such as genealogical completeness and generational depth, are also available. In addition, GenLib can compute measures of relatedness (kinship and inbreeding) and the genetic contribution of founders. Finally, functions for gene-dropping simulations are also available in GenLib. We illustrate the use of GenLib with a sample of 140 individuals from regional populations of Quebec previously described in Roy-Gagnon et al., 2011. Ascending genealogies were reconstructed for these individuals using BALSAC, yielding a large pedigree of 41,523 individuals. With GenLib, we provide a more detailed description of these genealogical data in terms of completeness, genetic contribution of founders, relatedness and inbreeding, further illustrating the regional differences reported in these data. We also present gene-dropping simulations based on the whole genealogy to estimate the probability of sharing, identical by descent, chromosomal segments of varying lengths introduced at varying frequencies by different number of founders. In conclusion, the R package GenLib provides a user friendly and flexible environment to analyze genealogical data, allowing a more efficient and easier integration of different types of data and analytical methods and making it ideal for further developments.

1738M

Genotype calling and phasing in sequence data from complex families. *L. CHANG[1], B. Li[2], S. Vrieze[3], M. McGue[4], W. Lacono[4], D. Weeks[1,5], G. Abecasis[6], G. Tseng[1], W. Chen[1,5,7].* 1) Biostatistics, University of Pittsburgh, PITTSBURGH, PA., USA; 2) Department of Molecular Physiology & Biophysics, and Neurology, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Psychology & Neuroscience, Institute for Behavioral Genetics, University of Colorado Boulder, CO; 4) Department of Psychology, University of Minnesota, MN; 5) Department of Human Genetics, University of Pittsburgh, PA; 6) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 7) Division of Pulmonary Medicine, Allergy and Immunology, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA.

As sequencing technologies can help researchers detect common and rare variants across the human genome in many individuals, it is known that jointly calling genotypes across multiple individuals based on linkage disequilibrium (LD) can facilitate analysis of low to modest coverage sequence data. However, genotype-calling methods for family-based sequence data, particularly for complex families beyond parent-offspring trios, are still lacking. In this study, firstly, we proposed an algorithm that considers both linkage disequilibrium (LD) patterns and familial transmission in nuclear and multi-generational families while retaining the computational efficiency. Secondly, we extended our method to incorporate external reference panels (e.g., 1,000 Genome Project) to analyze family-based sequence data with a small sample size. In simulation studies, we show that modeling multiple offspring can dramatically increases genotype calling accuracy and reduce phasing and Mendelian errors, especially for data sets with low to modest coverage. In addition, we show that using external panels can greatly facilitate genotype calling of sequencing studies with a small number of individuals. We applied our method to a whole genome sequencing study of ~1,800 individuals at ~10X coverage from the Minnesota Center for Twin and Family Research. We show that our methods significantly outperform existing ones that ignore family constraints or LD information. In summary, we present a statistical and computational method for genotype calling and phasing in sequence data from complex families and anticipate that our method will be useful for many ongoing family-based sequencing projects.

1739T

Assessing mitochondrial DNA variation and copy number in lymphocytes of 2,077 Sardinians using tailored sequencing analysis tools. *J. Ding[1], C. Sidore[2,3], T.J. Butler[1], M.K. Wing[2], O. Meirelles[1], Y. Qian[1], F. Busonero[2,3], R. Nagaraja[1], F. Cucca[3], G.R. Abecasis[2], D. Schlessinger[1].* 1) Laboratory of Genetics, National Institute on Aging, Baltimore, MD; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 3) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy.

To quantitate mitochondrial DNA (mtDNA) variation, genotype calling and analytic programs developed for nuclear DNA must be modified, because each cell has 100-10,000 mtDNA copies that can vary at any site (i.e., heteroplasmy). We report an algorithm adapted to identify mtDNA variants. It incorporates the sequencing error rates of the reads at each base in a likelihood calculation, and allows for different allele fractions at a variant site across all individuals. In addition, the algorithm is adapted to the circular mitochondrial genome, a key difference from the linear chromosomes assumed by most read mapping algorithms.

In the sample case of sequence from lymphocytes of 2,077 SardiNIA Project participants, we find that both homoplasmic and heteroplasmic mtDNA variants show 5-fold higher transition/transversion ratios than variants in nuclear DNA. As expected, mothers and their children share essentially all homoplasmies but a lesser proportion of heteroplasmies. Overall, heteroplasmy increases with age, but on average only ~1 heteroplasmy reaches the 4% level between ages 20 and 80. The total extent of accumulation of variants, however, remains hard to assess because heteroplasmy levels could be similar to or even lower than sequencing error rates; but resolving power can be increased in several ways, including deep sequencing of mtDNA.

We have also made a sequence-based estimate of mtDNA copy number based on the observed ratio of sequence coverage between mtDNA and autosomal DNA. The average in the cohort studied was 112 copies ±25 per cell, a number in agreement with estimates from qPCR assays. The computational methods thus facilitate comprehensive mtDNA analysis from whole-genome sequencing data. The algorithms can also be applied to quantitate sequence variation in other instances of high DNA copy number, such as cancer.

1740S

Improving effective sample size using extrapolated log p-values. *B. Engelhardt[1,2], X. Guo[2], S. Mukherjee[2], J. Tung[3].* 1) Biostatistics & Bioinformatics, Duke University, Durham, NC; 2) Department of Statistical Science, Duke University, Durham, NC; 3) Evolutionary Anthropology, Duke University, Durham, NC.

While the per-sample cost of genomic data collection has plummeted in recent years, population-based genomic studies, counterintuitively, remain extremely expensive. This expense arises because attempts to maximize power have relied almost exclusively on increases in sample size. An alternative, more cost-effective approach is to improve the statistical power associated with a given sample through development of more sensitive statistics. This approach is particularly attractive in the context of association mapping, or identifying genetic variants that are associated with complex traits. In particular, a large number of genetic variants may have an effect on a given complex trait, and effect sizes across these associated variants are thought to have a double exponential distribution (i.e., a few large magnitude effects and many effect sizes close to zero). However, association mapping, both in genome-wide association studies (GWAS) and quantitative trait locus-mapping (QTL) studies of molecular phenotypes, are restricted in their ability to detect associated variants with low effect size because of the limited number of samples in each study and the large number of associations tested. We present a statistical approach to extrapolate association significance (extrapolated log p-values, or ELPs) in a given study using subsampling methods. We validate this approach by quantifying gains in statistical power using simulated trait values and real genotype information from 4,646 individuals in the Wellcome Trust Case Control Consortium (WTCCC) data set. Our simulations show substantial gains in statistical power to detect low effect associations for a given sample size using ELP. We quantify this increase in effective sample size for association mapping in three diverse data sets: case-control mapping of Crohn's disease, and quantitative trait mapping of gene expression levels (eQTL mapping) in HapMap phase 3 individuals and the Genotype-Tissue Expression Project. Despite the simplicity of ELP, the theoretical justification for this approach is elegant in the context of classical statistics. ELP enables the detection of small effects in studies where the sample size has prevented these associations from reaching statistical significance, making association studies possible for populations and phenotypes for which existing methods are ineffective due to insufficient power. Broadly, ELP maximizes the power to draw statistical signal from expensive samples.

1741M

Alternative peak calling methods on Hi-C data accommodating the whole spectrum of dispersion. *Z. Xu, Y. Li.* Department of Genetics and Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Advancement in chromatic conformation capture and next generation sequencing technologies are enabling genomewide investigation of dynamic chromatic interaction. To analyze Hi-C data, Ay et al (2014) developed a peak calling method (Fit-Hi-C) that assigns statistical confidence estimates to mid-range intra-chromosal contacts based on a varying-coefficient Binomial model with the success rate as a smoothing function of the distance between the midpoints of a fragment pair. Ay et al (2014)'s model assumes that Hi-C data are equal-dispersed given the distance between the midpoints of a fragment pair. We found that the dispersion (given the distance) for real Hi-C data within different domains, either before or after outlier/peak filtering, can be (1) under-dispersed, (2) equi-dispersed, (3) over-dispersed or (4) dispersed differently for different distances. To address the dispersion issue in real-hi-C data, we consider alternative peak calling methods: (1) varying-coefficient Poisson model, (2) varying-coefficient Negative Binomial model, (3) varying-coefficient COM-Poisson model and (4) nonparametric standardization method. We benchmark these methods together with Fit-Hi-C using the number of interactions between known enhancers and transcription start sites given the same number of total peaks called. We found that each method has its own advantages. For example, we found that varying-coefficient Negative Binomial model performs better than varying-coefficient Poisson model for the domains showing over-dispersion based on Jin et al (2013)'s Hi-C data. This motivates us to consider a combined two-step peak-calling method, where the level of dispersion is quantified in the first step to guide the method of choice in the second step.

1742T

Haplotype based fine mapping algorithms using meta-analysis summary results. *J. Zheng[1, 5], S. Rodriguez[1], J. White[2], C. Giambartolomei[2], D. Zabaneh[2], R. Morris[3], M. Kumari[2], J.P. Casas[2, 3], A. Hingorani[2, 4], T.R Gaunt[1, 5], I.N. Day[1], UCLEB Consortium.* 1) Bristol Genetic Epidemiology Laboratory, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; 2) University College London Genetics Institute, Department of Genetics, Environment and Evolution, London, United Kingdom; 3) 3.Department of Primary Care & Population Health, University College London, Royal Free Campus, London, United Kingdom; 4) Centre for Clinical Pharmacology, University College London, London, United Kingdom; 5) MRC Integrative Epidemiology Unit, School of Social and Community Medicine, Bristol, United Kingdom.

Meta-analysis of multiple genome-wide association studies (GWAS) is becoming a popular method to increase statistical power and reduce false positive findings. As a follow-up procedure to meta-analysis, detecting secondary association signals at loci are necessary to determining variants with true causality. However, individual-level data is often not pooled, making fine mapping of associated loci via conditional analysis time consuming and impractical. Here, we present two statistical methods, Sequential Sentinel SNP Regional Association Plot" (SSS-RAP) (Zheng et al., 2013) and E-M algorithm HAPlotype-based Regional Association analysis Program (HAPRAP) (Zheng et al., 2014, under review). Both methods enable fine mapping using group-level summary statistics and reference haplotype information. SSS-RAP detects SNPs with independent effects conditional on the top associated signal. HAPRAP extends multiple regression and conditional analysis to meta-analysis levels. We demonstrate by a performance comparison that SSS-RAP is as accurate as conditional analysis and ten model selection methods in individual-level. We compare HAPRAP with existing methods using simulated data, BWHHS cohort and 1000 Genomes. The results of HAPRAP are highly consistent with multiple regression results using individual-level data. Moreover, HAPRAP consistently outperforms GCTA joint and conditional analysis (Yang et al., 2012) across different linkage disequilibrium (LD) correlations (r^2) and reference panel sample sizes. The Web-based interfaces of SSS-RAP and HAPRAP are available online: http://apps.biocompute.org.uk/sssrap/sssrap.cgi and http://apps.biocompute.org.uk/haprap.

1743S

PedBLIMP: A Linear Predictor based Approach to Impute Genotypes in Pedigrees. *W. Chen, D. Schaid.* Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN., USA.

Recently, Wen and Stephens proposed a linear predictor, called BLIMP, that uses conditional multivariate normal moments to impute genotypes with accuracy similar to current state-of-the-art methods. One novelty is that it regularized the estimated covariance matrix based on a model from population genetics. We extended multivariate moments to impute genotypes in pedigrees. Our proposed method, PedBLIMP, utilizes both the linkage disequilibrium (LD) information estimated from external panel data and the pedigree structure or identity by descent (IBD) information. By stacking the genotype matrix of single nucleotide polymorphisms (SNPs) for all individuals in a pedigree into a vector G, we showed that the mean of G is essentially the same as the mean of genotypes for any individual which can be estimated from external panel data. The covariance of G is a Kronecker product of two matrices. The first is the covariance matrix between SNPs within the same individual and the second is the correlation matrix between individuals in the pedigree. The first covariance matrix can be estimated from external panel data similarly as in BLIMP. The second matrix can be estimated in two ways. For global correlation, it is twice the kinship matrix which can be estimated from the pedigree structure. When genotype is available for pedigree members, local correlation of a specific locus or region between two individuals can be estimated based on inferred IBD in each region. *gl_auto* from MORGAN package is used to sample inheritance vectors and IBD is estimated accordingly. With the estimated mean and covariance of G, the best linear predictor is applied to estimate the expected dosage of untyped markers. The proposed method was evaluated on a pedigree design where some individuals were genotyped with dense markers and the rest with sparse markers. We found that incorporating the pedigree/IBD information can improve imputation accuracy compared to BLIMP. Because rare variants usually have low LD with other SNPs, incorporating pedigree/IBD information largely improved imputation accuracy for rare variants. We also compared PedBLIMP with IMPUTE2 and GIGI. Results show that when sparse markers are in a certain density range, our method can outperform both IMPUTE2 and GIGI. Our developed software PedBLIMP may be useful for imputing small to medium sized pedigrees where some individuals are densely genotyped and other individuals are sparsely genotyped.

1744M

Detecting maternal-offspring gene interactions using linear mixed effect models: The Quantitative-MFG Test. *M.M. Creek[1], E. Sobel[2], J.S. Sinsheimer[1,2,3].* 1) Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA; 2) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA; 3) Department of Biomathematics, University of California, Los Angeles, Los Angeles, CA.

The genetic etiology of many complex diseases remains to be discovered. Maternal and offspring gene interactions or maternal-fetal genotype (MFG) incompatibility may be a contributor. MFG incompatibilities occur when the effects of maternal genes on the offspring's phenotype vary depending on the offspring's genotype and have been shown to be involved in complex diseases such as schizophrenia and autism. Existing methods to investigate the effects of MFG incompatibility on quantitative traits using retrospective likelihoods are limited to case-parent trios. Presently, there are no methods to investigate the effects of MFG incompatibility on quantitative traits that are suitable for both small and large families. We propose the Quantitative-MFG (QMFG) Test, a prospective approach using a linear mixed effects model, which takes into account familial correlations by partitioning the variance. This approach can handle varying pedigree sizes, general and specific scenarios of MFG incompatibility, and the inclusion of covariates. We validate the model under two familiar maternal-offspring gene interaction scenarios, RHD and NIMA incompatibility. We estimate type I error rates and power by simulation and investigate the effect of risk allele frequency on power. The type I error rates for both scenarios are approximately 0.05 (SE=0.003). When the expected variance explained by a specified effect size is 1% the power is over 0.8 for RHD incompatibility and is over 0.7 for NIMA incompatibility when testing for any genetic effect, a NIMA effect, or an offspring effect with sample sizes of 500 nuclear families. Simulations demonstrate that parameters and variances are well estimated and that this method works with arbitrary sets of families. Additionally, we find that treating the nuclear families within an extended pedigree as independent results in inflated false positive rates, supporting the need of a test for association of MFG incompatibility and quantitative traits that utilizes all available family data.

1745T

On the null distribution of Bayes factors. *Y. Guan.* Pediatrics and Molecular & Human Genetics, Baylor College of Medicine, Houston, TX.

The p-value is the blood-love of many theoreticians and practitioners alike. A Bayesian procedure often faces an inconvenient demand of producing p-values. When a data analysis involves multiple testing, such as in a genome-wide association study, the significant threshold of the p-value is exceedingly small, owing to Bonferroni correction, and it requires a prohibitively large number of permutations to produce a p-value. Thus, characterizing the null distribution of Bayes factors, which enables one to compute p-values without the need of permutation, is of practical importance. On the other hand, it is beneficial to understand how priors affect Bayes factors under the null, which may in turn shed light on prior specification. We investigated how priors affect the null distribution of Bayes factors for Bayesian linear regression (of p covariates) under the Normal-Inverse-Gamma prior specification. For a finite sample, the null distribution of $2 log(ext{Bayes factor})$ under the independent prior is approximately a mean-shifted linear combination of chi_1^2, whose distribution function can be evaluated using an existing polynomial algorithm, and hence the corresponding p-value of a Bayes factor can be computed efficiently, without the need of permutation. The independent prior favors null for a large sample size; Zellner's g-prior does not--asymptotically, $2 log{(ext{Bayes factor})} sim g/(1+g) chi^2_p - p log(1+g) + O(1)$ under the g-prior.

1746S

Mixed model with correction for case-control ascertainment increases power in multiple sclerosis association study. *T. Hayeck[1], N. Zaitlen[2], P. Loh[3], B. Vilhjalmsson[1], S. Samuela[1], A. Gusev[1], J. Yang[3], G. Chen[3], M. Goddard[4], P. Visscher[3], N. Patterson[5], A. Price[1].* 1) Harvard School of Public Health, Boston, MA; 2) University of California, San Francisco, CA; 3) University of Queensland, Brisbane, Australia; 4) University of Melbourne, Melbourne, Australia; 5) Broad Institute, Cambridge, MA.

We introduce a Liability Threshold Mixed Linear Model (LTMLM) association statistic for ascertained case-control studies that increases power vs. existing mixed model methods, with well-controlled false-positive rate. Recent work has shown that existing mixed model methods suffer a loss in power under case-control ascertainment (Yang et al. 2014 Nat Genet), but no solution has been proposed. Here, we solve this problem using a chi-square score statistic computed from posterior mean liabilities (PML) under the liability threshold model. Each individual's PML is conditional not only on that individual's case-control status, but also on every individual's case-control status and on the genetic relationship matrix obtained from the data. For example, disease cases with higher genetic relationships to other disease cases are assigned a larger PML than disease cases with lower genetic relationships to other disease cases. The PML are estimated using a multivariate Gibbs sampler, with the liability-scale phenotypic covariance matrix based on the genetic relationship matrix (GRM) and a heritability parameter estimated via Haseman-Elston regression on case-control phenotypes followed by transformation to liability scale (Lee et al. 2011 AJHG). Leave One Chromosome Out (LOCO) analysis was used to calculate the GRMs, avoiding the decrease in power that arises from including candidate SNPs in the GRM (Yang et al. 2014 Nat Genet).The Gibbs sampler does not iterate over SNPs, thus overall running time is comparable to existing mixed model methods. In simulations of unrelated individuals, the LTMLM statistic was correctly calibrated and achieved higher power than existing mixed model methods in all scenarios tested, with the magnitude of the improvement depending on sample size and severity of case-control ascertainment. In a WTCCC2 multiple sclerosis data set with greater than 10,000 samples, LTMLM was correctly calibrated and attained a 4.1% improvement (P equal 0.007) in chi-square statistics (vs. existing mixed model methods) at 75 known associated SNPs, consistent with simulations. Larger increases in power are expected at larger sample sizes. In conclusion, an increase in power over existing mixed model methods is available for ascertained case-control studies of diseases with low prevalence.

1747M

GARFIELD - GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction. *V. Iotchkova[1,2], M. Geihs[1], G. Ritchie[1,2], E. Birney[2], I. Dunham[2], N. Soranzo[1], UK10K Consortium Cohorts Group.* 1) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1HH, United Kingdom; 2) The EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, United Kingdom.

Genome-Wide Association Studies (GWAS) have been increasingly fruitful in discovering genotype-phenotype associations. The mechanisms underlying these associations, however, are still largely unknown as only a small fraction of these SNPs directly alter protein-coding genes. The interpretation of functional consequences of non-coding variants has been greatly enhanced by large-scale efforts to identify regulatory genomic regions (e.g ENCODE). However, robust methods are still lacking to systematically evaluate the contribution of these regions to genetic variation implicated in diseases or quantitative traits. Here we propose a novel approach that leverages GWAS findings with regulatory or functional annotations to find features relevant to a phenotype of interest. We perform greedy pruning of GWAS SNPs (LD r^2>0.1) and then annotate them based on functional information overlap. Next, we quantify Fold Enrichment (FE) at various GWAS significance cutoffs and assess them by permutation testing, while matching for minor allele frequency, distance to nearest transcription start site and number of LD proxies (r^2>0.8). Within this framework, we account for major sources of confounding that current methods do no offer. Using our method, we analysed publicly available GWA summary statistics on 16 disease and quantitative traits in 1005 annotations, including genic elements from GENCODE and DNaseI hypersensitive sites (DHS) for a range of cell types from ENCODE and Roadmap Epigenomics. Results showed coding exons as consistently highly enriched among traits, but also uncovered various trait specific enrichment patterns. For example, Height was significantly enriched in nearly all DHS cell types, consistent with its polygenic nature, while Crohn's disease exhibited enrichment predominantly in blood and intestines. We then systematically compared enrichment levels between associations from whole genome sequencing (WGS from the UK10K project) and sparser genotyping datasets (e.g. HapMap). When applied to 4 main lipid subfractions, the method found predominantly larger FE values as well as larger number of significant enrichments in the WGS data, highlighting a greater sensitivity of WGS data. Finally, we developed new software to facilitate the application of our method by the research community. Overall, it is expected that a more accurate classification of enrichment patters might lead to biological insights and help prioritise variants for follow-up studies.

1748T

Comparison of machine-learning methodologies to prioritize genetic variants based on functional data. *S.A. Gagliano[1,2,3], R. Ravji[1,2,3], M.R. Barnes[4], M.E. Weale[5], J. Knight[1,2,3], Schizophrenia Working Group of the Psychiatric Genomics Consortium.* 1) Centre for Addiction and Mental Health, Toronto, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, Canada; 3) Department of Psychiatry, University of Toronto, Toronto, Canada; 4) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom; 5) Department of Medical & Molecular Genetics, King's College London, Guy's Hospital, London, United Kingdom.

Incorporating a variety of functional characteristics as predictors, such as DNase I hypersensitive sites, transcription factor binding sites, and histone modifications, we developed a methodology using elastic net, a regularized regression, to prioritize genetic variants that are truly associated with complex disease (Gagliano et al. PLoS ONE 2014). We compared the results from our elastic net method to two other published methods: Ritchie et al. (Nat Methods 2014) and Kircher et al. (Nat Genetics 2014). These models use different methodologies (modified random forest and a support vector machine, respectively) and different predictors with regard to the quantity, type and coding of the functional characteristics. We tested which methodology performed the best for prioritizing sub-genome-wide-significant variants (5E-8<p<1E-6) by investigating which one assigned higher scores to the sub-genome-wide-significant variants from the first round of a schizophrenia genome-wide association study (GWAS) that come up as associated in the second round, and lower scores to those variants that do not come up as associated. We obtained prediction values of the sub-genome-wide-significant variants in the first round of the schizophrenia GWAS by the Psychiatric Genomics Consortium (PGC1, Schizophrenia GWAS Consortium, 2011) for the three methodologies. A quantile-quantile plot was created using the p-values from the larger second round of the schizophrenia GWAS (PGC2, submitted) stratified for each of the methodologies by the sub-genome-wide-significant variants from the first round belonging to the top and bottom quartiles with regard to their prediction value. Elastic net outperforms the other two methodologies when it comes to identifying truly associated genetic variants in the schizophrenia GWAS. We are also exploring the use of the different algorithms, for instance, elastic net compared to random forest, while holding constant the functional characteristics used in order to discriminate performance differences as being attributed to the methodology and/or the functional predictors.

1749S

A mixed model methodology to correct technical artifacts and enable meta-analysis of sequence based association studies. *C. Murphy[1], P. Syrris[2], P. Lambiase[3], D. Speed[1], V. Plagnol[1], UCL-exomes Consortium.* 1) Genetics Institute, University College London, London, United Kingdom; 2) Institute of Cardiovascular Science, University College London, United Kingdom; 3) Heart Hospital 16-18 Westmoreland Street London W1G 8PH.

High throughput DNA sequencing technologies, either whole exome (WES) or whole genome (WGS) sequencing are revolutionizing the diagnosis and novel gene discovery for rare disorders. As the field transitions from the early discovery for Mendelian and near Mendelian diseases to more complex and oligo-genic diseases, there is substantial benefit in being able to combine data across studies, performing the type of meta-analysis for cases and controls that have proven to be so successful for genome-wide association studies (GWAS). However, WGS and WES are substantially more affected by sequencing errors and technical artifacts than genome-wide genotyping arrays. As a consequence, meta-analysis of sequence based association studies are often dominated by spurious associations, which result in technical limitations. Here, we show that it is possible to take advantage of the type of mixed models developed initially to control for population structure in GWAS studies, and apply these ideas to control for technical artifacts. We developed computational improvements to optimize the estimation of the mixed model parameters to enable genome-wide testing in a realistic time (from days to minutes). These techniques are being implemented as part of the LDAK package (www.ldak.org). Using a dataset of 2,800 WES (UCL-exomes), which aggregates a diverse set of studies, we demonstrate that substantial reduction in the association statistic inflation can be achieved by applying these novel analytical techniques, both for single variant and gene based tests, while preserving the sensitivity of the test. We focus on several cardio-vascular diseases (ARVC and sudden cardiac death) to illustrate the ability of these novel methods to produce more interpretable results. Our methodology provides a general framework to facilitate current and future meta-analysis of sequence based association studies.

1750M

Sparse heterogenetic sequence association mapping with arbitrary population structure and cryptic relatedness. *H. Qin[1], W. Ouyang[1], X. Chen[2], M. Xiong[3], G. Gao[4], HS. Chen[5], YP. Wang[6], HW. Deng[1], X. Zhu[7].* 1) Department of Biostatistics and Bioinformatics, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 2) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA 23298, USA; 3) Human Genetics Center and Division of Biostatistics, The University of Texas School of Public Health, Houston, TX 77030, USA; 4) Department of Biostatistics, Virginia Commonwealth University School of Medicine, Richmond, VA 23298, USA; 5) Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA; 6) Department of Biomedical Engineering, Tulane University, New Orleans, LA 70118, USA; 7) Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH 44106,USA.

High-throughput DNA sequencing technologies have been generating datasets of massive rare and common genetic variants. In a typical sequence dataset, a short genomic region may contain sequences of many single-nucleotide polymorphisms (SNPs). For example, in the sequence data of Mexican pedigrees from GAW18, 68% of all the 11,933 non-overlapped 100-kb regions on odd chromosomes contain more than 500 SNPs. It is a fundamental challenge to identify a high-dimensional SNP set with sparse functional variants, where the set size m is close to or larger than the sample size n. Existent prominent association methods appear suitable to identify dense low dimension SNP sets, i.e., m/n<10%, and >20% of set wise SNPs are functional loci. Such methods are ineffective and insufficient to identify sparse high-dimensional SNP sets due to perceived limitations. First, they may be severely confounded by inherent population structure and cryptic relatedness when applied to high-dimensional sets of rare and/or common variants. Second, they suffer unnecessarily large degrees of freedom due to massive neutral variants. Third, they model mean genetic effects only and ignore inherent heterogeneity. In this report, we propose an efficient method for identifying sparse high- dimensional sets from sequence data with arbitrary population structure and cryptic relatedness. In this method, we first perform a heterogenetic feature selection to search for k (< n×10%) SNPs that best explain phenotypic data; we then link the selected variants with the confounders by a hetero heterogenetic mixed-effect model and perform likelihood ratio test for the genetic effects. In both feature screening and formal association test, we model population structure as fixed effect and model cryptic relatedness as a random effect in respective heterogenetic mixed-effect models. Under extensive simulations, the proposed method properly controlled type I error rates and appeared more powerful than existent prominent sequence association methods (e.g., SKAT, famSKAT and fastLMM). The proposed method is computationally efficient, integrates heterogeneity and diverse genetic effects, and simultaneously adjusts for arbitrary covariates and cryptic relatedness. Application to the deep sequence data of Mexican American pedigrees from GAW18 demonstrated practical utility of the proposed method.

1751T

Estimation of prognostic marker genes by public microarray data in patients with ovarian cancer epithelial. *S. Yang, J. Kim.* Biomedical Sciences Dept, Seoul National University College of Medicine 103 Daehakno, Jongnogu, Seoul 110-799, KOREA.

Lymphatic invasion is regarded as a predictor of aggressiveness of ovarian cancer. However, lymphatic invasion do not accurately represent 5 years survival. To diagnosis and treatment of ovarian cancer, we analyzed the differentially expressed genes between 5 years survival group and 5 years death group of lymphatic invasion in serous ovarian epithelial cancer with DNA microarray. Data from 63 ovarian cancer patients with lymphatic invasion and 35 ovarian cancer patients without lymphatic invasion from TCGA data were analyzed. Among these 98 patients, 16 patients were survived 5 years or more. DEGs identified by Bioconductor R package. Functional analysis of genes that were analyzed with DAVID web tool. We found 20 DEGs (P value<0.001) from 5 years survival 64 patients and 5 years death 8 patients without lymphatic invasion. Also we found 55 DEGs (P value<0.01) from 5 years survival 55 patients and 5 years death 8 patients with lymphatic invasion. Pathway analysis showed that the lymphatic invasion related DEGs were related with starch and sucrose metabolism pathways while lymphatic invasion related DEGs were related with renal cell carcinoma pathways. These DEGs for ovarian cancer results in high grade serous ovarian cancer patient survival, particularly in 5 years death and survival patients with and without lymphatic invasion. These findings may have implications for the diagnosis and treatment of the ovarian cancer.

1752S

Modeling Temporal Changes in Phenotypes in Pediatric Populations. *R. Hoffmann.* QHS, Pediatrics, Medical College of Wisconsin, Milwaukee, WI.

Statement of Purpose: During the first year of life, pathways develop and change in importance leading to the presentation of different phenotypes over time. Thus statistical genetic methods are needed which allow the phenotype-genotype model to change over time. This model developed here allows both regular and irregular sampling time points to be incorporated. **Methods:** A generalized linear mixed model (GLMM) was used to model the change in phenotype over time. This allows variability in the phenotype between subjects to be modeled, changes in the phenotype-genetic relationship over time within subjects to be modeled, the choice of different types of phenotypic relationships to be modeled (binary, categorical, normal and skewed-continuous) and testing of whether different models are appropriate for different variants at the same locus. In addition the functional form can be adjusted to the times at which the data was collected. In our case, 2 days, 7 days, 30 days, 90 days, 180 days and 360 days. We compared smooth parametric models for the relationships with non-parametric models which allow the values to be estimated at each sampling time point. Simulations were based on using an underlying fractional polynomial model over time so that sufficient differences in shape could be tested. **Results:** While the parametric model had more power if the model was known, in practical situations where the model is not known, the non-parametric model had both more power and more sensitivity to detect unknown relationships. The non-parametric GLMM also had a substantial advantage in correctly identifying models where the data was irregularly sampled at different time points and models with irregular behavior (peaks and troughs) over time. These relationships were poorly modeled by parametric models, even with sin-cosine terms (standard time series) because of the tendency to over-smooth the relationship over time.

1753M

A novel meta-analysis approach for genome-wide association studies with sex-specific effects. *E. Kang[1], J. Joo[2], N. Furlotte[1], E. Kostem[1], B. Han[3,4,5,6], E. Eskin[1,7].* 1) Department of Computer Science, UCLA, Los Angeles, CA; 2) Interdepartmental Program in Bioinformatics, UCLA, Los Angeles, CA; 3) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 4) Division of Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 5) Partners Center for Personalized Genetic Medicine, Harvard Medical School, Boston, MA; 6) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 7) Department of Human Genetics, UCLA, Los Angeles, CA.

Over the past several years, a significant amount of loci implicated in genome wide association studies show differences in effect sizes for males and females. The prevalence of sex-specific effects both motivates strategies for discovery of sex-specific effects as well as raises questions of how to analyze association studies consisting of both males and females. The traditional approach to discover sex-specific effects is to analyze each sex separately and the traditional approach to analyze association studies consisting of both males and females is to include sex as a covariate in the statistical model when performing association analysis. Unfortunately, these approaches may lead to a loss in power when sex-specific effects are present at the loci being tested. In this paper, we present a novel meta-analytic approach for the analysis of genome-wide association studies consisting of both males and females. In our approach, males and females are analyzed separately and the results are combined using a random effects meta-analysis approach allowing for differences in effect sizes between sexes. We show that by analyzing males and females separately, our method reduces the overall variance in each study leading to an increase in statistical power. Through simulations and application of our method to the Northern Finland Birth Cohort data, we show that our method has increased power over the traditional approaches for discovering associated loci with and without sex-specific effects while controlling for false positives.

1754T

Mapping of novel regulatory influences on genes encoding subunits of the L-type calcium channel, using digital measurement of allelic skew. *N. Kamitaki[1,2], S. Ghosh[2,3], K. Eggan[2,3,4], S.A. McCarroll[1,2].* 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA; 3) Harvard Stem Cell Institute, Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA; 4) Howard Hughes Medical Institute, Cambridge, MA.

Most of the SNP haplotypes implicated by genome-wide association studies (GWAS) do not include protein-coding variants, suggesting that the functional variant(s) has a regulatory function. One of the more proximal phenotypes to genotype is gene expression, but precise quantification of gene-regulatory effects in brain has both biological and technical challenges - in particular, genetic background and sample-to-sample variability from varying mixtures of different cell types in a complex tissue. Allele-specific expression can be used to control for environmental and trans-acting genetic influences, but techniques for measuring allele-specific expression often exhibit amplification bias or are insufficiently sensitive to detect modest, quantitative effects on gene expression. We developed a set of new molecular and statistical methods for making precise digital measurements of allele-specific expression and using such measurements to map gene-expression effects. These methods include: 1) A novel molecular method for measuring allele-specific expression with high precision, by digital counting of alleles in RNA. 2) A novel molecular method for inferring the chromosomal phase of candidate regulatory variants with transcribed sequence variants; this method is fast and scalable, and we have validated it at genomic distances up to 200 kb. 3) Companion statistical methods for combining the results of these assays with genome-variation data to map regulatory effects on gene expression. We applied these methods to *post mortem* brain samples from 5 brain regions in 105 individuals, focusing initially on genes encoding subunits of the L-type calcium channel, now implicated in genetic studies of schizophrenia and bipolar disorder. Using this approach, we were able to map novel regulatory effects that have not been identified in expression QTL studies of the same genes. For example, we find multiple LD-independent effects on the expression of *CACNA1C*; both effects replicated in an independent set of post mortem brains. Some regulatory effects were brain-region-specific while another showed a regulatory effect consistent in magnitude across brain regions. We have introduced several variants into embryonic stem cells using the Crispr/Cas9 system and are currently differentiating these into neurons to evaluate effects on expression.

1755S

A Candidate Pathway Approach Identifies Multiple Gene-Environment Interactions in Association with Colon Cancer Risk and Survival. *N. Sharaf Eldin[1], M.L. Slattery[2], Q. Liu[1], C. Franco-Villalobos[1], B.J. Caan[3], J.D. Potter[4,5,6], Y. Yasui[1].* 1) School of Public Health, University of Alberta, Edmonton, Alberta, Canada; 2) Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, Utah, USA; 3) Division of Research, Kaiser Permanente Medical Care Program, Oakland, California, USA; 4) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; 5) Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA; 6) Centre for Public Health Research, Massey University, Wellington, New Zealand.

Genetic association studies have traditionally focused on effects of individual genetic markers, usually single nucleotide polymorphisms (SNPs), on disease or phenotype. A comprehensive approach more suited to decipher causal relationships and underlying molecular mechanisms, however, considers multiple genes implicated in the same pathway and furthermore interacting with environmental exposures. This analysis aimed at assessing the interaction effects of the angiogenesis gene-pathway with three main lifestyle exposures (dietary protein intake, smoking, and alcohol consumption) on colon cancer risk and survival. Logic regression, which identifies Boolean combinations of predictor variables, was used to assess gene-pathway effects, followed by assessments of gene-environment interactions (GEIs) using logistic regression for risk and Cox's proportional-hazards models for survival. We analyzed data of 1,541 colon cancer cases and 1,934 controls from the Diet, Activity and Lifestyle as a Risk Factor for Colon Cancer Study conducted at three centers in the United States. The study selected a total of 257 SNPs in 34 genes in the angiogenesis candidate gene-pathway based on standard pathway maps and experimental evidence. We found five statistically significant GEIs for colon cancer risk and three GEIs for colon cancer survival with increasing levels of all three environmental exposures. For risk: FLT1 rs678714 and ≥ 20pack-years smoking (interaction odds ratio (ORINT) = 1.64, 95% confidence interval (CI) (1.11, 2.41)); FLT1 (rs2387632 OR rs9513070) and high animal protein intake (ORINT =1.69, 95% CI (1.03, 2.77)); KDR rs6838752 and heavy alcohol consumption (ORINT =1.53, 95% CI (1.10, 2.13)); BMP4 rs17563 and ≥ 20pack-years smoking (ORINT =1.60, 95% CI (1.10, 2.32)); and TLR2 rs3804099 and heavy alcohol consumption (ORINT =1.59, 95% CI (1.05, 2.38)). For survival: TNF rs1800630 and high animal protein intake (interaction hazard ratio HRINT =1.80, 95% CI (1.13, 2.86)); BMP1 (rs13257482 OR rs4075478) and ≥ 20pack-years smoking (HRINT =1.78, 95% CI (1.03, 3.09)); and BMPR2 rs12477602 and heavy alcohol consumption (HRINT =7.78, 95% CI (1.66, 39.03)). Our study shows that GEI effects on colon cancer risk and survival can be identified by adopting a comprehensive candidate pathway approach that emphasizes the biologic hypothesis in the selection of the pathway genes and environmental exposures and carries that logic through to the analysis.

1756M

Genome-wide association study in West Africans identifies SEMA4D as a susceptibility gene for Obesity. *G. Chen, A. Bentley, A. Doumatey, D. Shriner, J. Zhou, A. Adeyemo, C. Rotimi.* Center for Research on Genomics and Global Health, National Human Genome Research Institute, NIH, Bethesda, Maryland, USA.

The burden of obesity, a major risk factor for several cardio-metabolic disorders and some cancers, has reach epidemic proportion in several global populations. Thus, understanding the etiology of excess weight gain is a scientific and public health imperative. Although, more than 20 GWAS have been reported for obesity and related traits, none has been conducted in continental Africans. Here, we report on the first GWAS for Body Mass Index (BMI) using a total of 21.2 million SNPs in 1624 West Africans. We identified a novel genome-wide significant locus at 9q22.2 (SEMA4D, rs80068415 p = 2.10 × 10-8). SEMA4D has been reported to be crucially involved in the activation and differentiation of T cells and the promotion of Th1 (T helper 1) differentiation. Interestingly, Th1 polarization in adipose tissues has been observed in mice with diet induced obesity. We also identified five suggestive evidences of association in 2q31.1 (NOSTRIN p = 1.30 × 10-7), 5q32 (PPP2R2B, p = 1.26 × 10-7), 11p13 (SLC1A2, p = 7.56 × 10-8), 14q21.1 (LRFN5p = 9.16 × 10-8), and 15q15.1 (NUSAP1, p = 2.15 × 10-7). We used 3,000 bootstraps and 3,000 simulation studies to confirm identified associations. Of the reported 92 GWAS loci, 83 were available for assessment for possible transferability in this sample of West Africans. Notably, 56 of the 83 loci (67.5%) displayed trans-ethnic implications for obesity. In all, this first GWAS for obesity in Africans identified a novel gene (SEMA4D) for obesity and confirms several reported loci in European ancestry populations to be potentially important for obesity in Africans.

1757T

Effect Fine Mapping: a method to identify association-driving variants in large genomic datasets. *N.A. Patsopoulos[1, 2, 3], J. Replogle[1, 2, 3], A.H. Beecham[4], A. Brandes[3], J. Oksenberg[5], J.L. McCauley[4], S.E. Baranzini[5], C. Cotsapas[3, 6, 7], P. De Jager[1, 2, 3].* 1) Department of Neurology, Brigham & Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Broad Institute, Cambridge, MA; 4) John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, Florida, USA; 5) Department of Neurology, University of California, San Francisco, Sandler Neurosciences Center, San Francisco, California, USA; 6) Department of Neurology, Yale University School of Medicine, New Haven, Connecticut, USA; 7) Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA.

Although genome-wide association studies (GWASs) have been extremely successful in identifying a plethora of validated genetic associations, only a handful of casual variants have been identified to date. Even fine-mapping efforts with dedicated, high-density arrays designed to capture the majority of variation in a locus (e.g. ImmunoChip), or extensive re-sequencing studies have not been as successful in identifying such variants as expected. In a recent ImmunoChip analysis of multiple sclerosis individuals, state of the art methods could fine-map only 8 out of 68 confirmed loci to high resolution. More importantly, in only 5 of these 8 could we identify one variant accounting for more than 50% of the posterior probability to drive the association. Here we argue that, beyond uncertainty due to linkage disequilibrium, fine-mapping may be impaired by the presence of multiple independent effects in a locus or by true causal alleles being non-SNP variation. We address these issues in a method, effect fine mapping (EFM), based on information criterion theory that aims to identify the group of genetic variants that drive the statistical association. The method can easily be implemented in existing analytical models, under a frequentist or Bayesian framework, with or without permutations. We demonstrate how it can be used to capture the set of SNPs that explain the statistical signal of association (EFM set), and, in cases of one driving variant, how it can be equivalent of fine-mapping methods aimed to prioritize causal variants,. By applying EFM in large-scale genetic data we illustrate that most of the identified effects are mediated not by a single genetic variant but by many, which in some cases overlap more than one genes. Next, we illustrate how one can use the EFM set to prioritize genes and use this information in pathway, network, and enrichment analysis. Finally, we use the EFM set to estimate, via model averaging, one effect size and p-value for the identified effect that can be used to minimize winner's curse and assign summary statistics to associated loci or genes.

1758S

The study of epistasis and pleiotropic effects using multi-association for metabolic syndrome in Korean population-based cohort. *Y. Lee, S. Park, B. Han, B. Kim, J. Lee.* Korea National Institute of Health, Chungcheongbuk-do, South Korea.

Metabolic syndrome (MetS) is a kind of disease which is determined co-occurrence of three of five following medical conditions: abdominal obesity, elevated blood pressure, elevated fasting plasma glucose, high serum triglycerides, and low high-density cholesterol levels. It is well known that people with MetS become more dangerous to cardiovascular disease (CAD) or type 2 diabets mellitus (T2D). For several years, many genome-wide association studies (GWASs) have been conducted, but the genetic determinants of the MetS were still unknown. Recently, MFQLS tool which is likely to analyze the association between multi traits and multi genetic information was developed. In this study, we conducted multi-association (both multi traits-unit SNP and multi traits-multi SNPs) with MetS while adjusting age and sex. For the multi-association study, we divided this analysis into two step: genetic association study using the Korea Association Resource (KARE) cohort and replication study using the Health Examinee (HEXA) cohort. Also, this analysis was performed whether common genetic variants have pleiotropic effects accounting for some of the interaction among five metabolic phenotypes that define MetS and common genetic variants have the epistasis effects to illustrate the gene-gene interaction for MetS.

1759M

Analysis of pleiotropy at a fine genomic scale. *D.J. Balding, D. Speed.* Inst Genetics, Univ College London, London, United Kingdom.

Bivariate SNP-based heritability analysis is being widely applied to measure the overlap in genetic architecture between pairs of diseases. It is especially popular for psychiatric diseases, as better understanding of the concordance or discordance between different traits can inform nosology and improve prediction of disease progression. Given SNP-level data for a pair of diseases, the approach implemented in the software GCTA returns an estimate of ρ, the genome-wide average correlation between the SNP effect sizes for each trait: ρ>0 indicates that risk alleles for the first disease tend to also increase risk for the second, and vice-versa. A recent study considered five psychiatric disorders (schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorders (ASD) and attention deficit hyperactivity disorder (ADHD)), and found significant concordance (ρ>0) for five of the ten pairings; the strongest concordance was reported for schizophrenia and bipolar disorder, in line with previous knowledge of the overlap between these two conditions. A major limitation of this approach is that ρ represents concordance averaged across the entire genome, which means that similarity in one genomic region can be cancelled out by dissimilarity in another. Furthermore, a significantly non-zero estimate of ρ provides no indication of which genomic regions are concordant or discordant, knowledge of which which would assist in identifying shared genes and pathways. We propose a tool for measuring concordance at a fine genomic scale, at the level of individual genes or short genomic regions. Using computational optimization, our algorithm is extremely fast, able to analyse genome-wide data for many thousands of individuals in a matter of minutes. As proof of principle, our method identified the overlap between Rheumatoid Arthritis and Type 1 Diabetes, with substantial concordance in the major histocompatibility region. For two major subtypes of Inflammatory Bowel Disease, Crohn's Disease and Ulcerative Colitis, it identifies a gene where risk alleles for the former are protective for the latter. We re-examine the study of five psychiatric disorders, and find significant genomically-local concordance even where none exists genome-wide.

1760T

Development of efficient polygenic risk scores for personalized medicine: methodological concepts and examples. *K. Fischer[1], K. Läll[1], R. Mägi[1], A. Morris[2,1], A. Metspalu[1].* 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, UK.

Polygenic risk scores have great potential in predictive personalized medicine, even in cases where each of the individual variants identified by a genome-wide association study has a relatively weak effect on the disease of interest. There are, however, several methodological issues that need to be considered while constructing such scores. First, the genetic variants to be used in such scores, as well as their relative weights, are often chosen on the basis of a discovery study. This may lead to overestimation of the real effect of the score, due to selection bias towards variants that have an effect that is overestimated by chance ("winners curse"). Another potential source of biases is present when the outcome is binary and the discovery study (or a replication study) uses simple logistic regression to estimate the effects. Due to non-collapsibility of the odds ratios, such estimates are biased estimates for the corresponding multiple regression coefficients (and for the optimal weights while combining the SNP allele counts to a polygenic score). A third source of biases is created by evaluating the score in a cohort that was included in the initial discovery study - especially when genetic variants with relatively weak effects are included. We demonstrate by simulations, as well as by application to polygenic risk scores for Type II Diabetes (T2D) in the Estonian Biobank cohort, how such biases can be corrected for. We will also show that the efficiency of the predictive score can be increased by adding more than just the genome-wide significant variants to the score, but only up to certain threshold. Adding too many variants with no clear effect would lead to decreased predictive power. In addition, we discuss ways to estimate the predictive ability of a polygenic risk score in the context of other predictors, such as the Body Mass Index, as well as communicate its effect to a public, using the example of T2D risk prediction in the Estonian Biobank.

1761S

A Proper and Efficient Approach to Integrative Analysis of Sequencing and GWAS Data for Rare Variant Associations. *Y.J. Hu[1], Y. Li[2,3], P. Auer[4], D.Y. Lin[2].* 1) Biostatistics and Bioinformatics, Emory University, Atlanta, GA; 2) Biostatistics, University of North Carolina, Chapel Hill, NC; 3) Genetics, University of North Carolina, Chapel Hill, NC; 4) Joseph J. Zilber School of Public Health, University of Wisconsin, Milwaukee, WI.

In the large cohorts typically used for genome-wide association studies (GWAS), it is not economically feasible to sequence all cohort members. A cost-effective strategy is to sequence subjects with extreme values of quantitative traits or those with specific diseases. By imputing the sequencing data from the GWAS data for the cohort members who are not selected for sequencing, one can dramatically increase the number of subjects with information on rare variants. However, treating the imputed rare variants as observed quantities in downstream association analysis may inflate the type I error, especially when the sequenced subjects are not a random subset of the whole cohort. Although this problem can be alleviated by restricting the analysis to the variants that are accurately imputed, a large number of rare variants will be excluded with this strategy. In this article, we show how to properly account for the uncertainties in the imputation of rare variants. We consider all commonly used gene-level association tests, including the burden test, variable threshold (VT) test, and sequence-kernel association test (SKAT), all of which are based on the score statistic for assessing the effects of individual variants on the trait of interest. We show that the score statistic based on the observed genotypes for sequenced subjects and the imputed genotypes for non-sequenced subjects is unbiased. We construct a robust variance estimator that reflects the true variability of the score statistic regardless of the sampling scheme and imputation quality, such that the corresponding association tests always have correct type I error. Through extensive simulation studies, we demonstrate that the proposed tests are substantially more powerful than the use of accurately imputed variants only or the use of sequencing data alone. An application to the data from the Women's Health Initiative (WHI) is provided. The relevant software is freely available.

1762M

Genetic Studies of Functional Quantitative Trait with both GWAS and Next-Generation Sequencing Data. *D. Lee[1], C. Hanis[2], G. Bell[3], D. Aguilar[4], B. Cade[5], J. Below[3], M. Xiong[2].* 1) Biostatistics, University of Texas School of Public Health, Houston, TX; 2) Human Genetic Center, University of Texas School of Public Health, Houston, TX; 3) Department of Human Genetics, University of Chicago, Chicago, IL; 4) Cardiology, Baylor College of Medicine, Houston, TX; 5) Division of Sleep Medicine, Harvard medical school, Boston, MA.

Traditional quantitative genetics has primarily studied traits with cross section data. However, in real biologic world, many quantitative traits change over time. These quantitative traits are repeatedly measured as functions of time or complete curves. The traditional methods for such genetic studies summarize the time or space varying traits as means across the time or space interval and reduce these functional quantitative traits as scalar traits. Another challenge for genetic studies of quantitative traits is to study genetics with next-generation sequencing (NGS) data. NGS can generate millions of genetic variation data. As a consequence, these genetic variation data are so densely distributed across the genome that can be modeled as a function of genomic location. In summary, random functions appear as either responses or predictors, or both. Although the current widely used statistical methods for QTL analysis are based on multivariate analysis, this method often fails with functional data. To overcome the limitations of the traditional QTL analysis for functional quantitative traits and NGS data, we use a functional linear model (FLM) in which a trait curve is modeled as a response function, the genetic variation in a genomic region is modeled as a functional predictor, and the genetic effects are modeled as a function of both time and genomic position. By extensive simulations, we demonstrate that the FLM for genetic studies of the temporal quantitative traits has the correct type 1 error rates and much higher power to detect association than the current regression methods. The proposed method is applied to 833 individuals with 20,763 genes from Starr County health studies sleep data where oxygen saturation were measured for 22,670 seconds on average. We found 65 genes that were significantly associated with oxygen functional trait with P-values ranging from 2.4E-6 to 2.5E-21, including gene RELB, which are confirmed in the literature. Among them, 8 genes that were located in metabolic pathways and NF-kappa B signaling pathway. The excellent QQ plot of our results showed that we did not inflate the false positive rates. However, using mean of oxygen saturation over the whole sleep time period as a quantitative trait, we only identified two genes associated with oxygen saturation. Our results clearly demonstrate that the FLM with both functional response and predictors substantially outperform the traditional genetic models with scalar trait.

1763T

Genotype risk score may mislead physiological interpretation of quantitative trait associations. *N. Wang[1,2,3], Y. Shu[1,2,3], H. Allayee[1,3], A. Xiang[4], T. Buchanan[5,2], R. Watanabe[1,2,3].* 1) Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA; 2) Physiology and Biophysics, Keck School of Medicine of USC, Los Angeles, CA; 3) USC Diabetes & Obesity Research Institute, Keck School of Medicine of USC, Los Angeles, CA; 4) Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA; 5) Department of Medicine, Division of Endocrinology, Keck School of Medicine of USC, Los Angeles, CA.

The genotype risk score (GRS) has been used as a measure of total genetic risk and a predictor of complex disease risk. Better understanding of the physiology of complex disease can be achieved by studying disease-related quantitative traits (QTs). GRS has recently been increasingly used to explore association with disease-related QTs, although its validity has not been examined. We tested 38 type 2 diabetes (T2D) risk SNPs for association with 17 T2D-related QTs in the BetaGene study, which consists of Mexican American families of probands with or without previous gestational diabetes. We compared association by single-SNP, GRS, and stepwise forward selection (FS), to assess which method provides clear physiological interpretation. We tested each SNP for univariate association with Bonferroni correction for 646 tests (38 SNPs×17 QTs), which is overly conservative due to correlation among QTs. GRS was constructed by summing risk alleles without weighting and testing for association. FS picked SNPs with a threshold of $P \leq 0.1$, followed by omnibus likelihood ratio test for association. All tests were adjusted for age, sex and familial correlation. 727 subjects with complete data, to ensure fair comparison, were included. Univariate testing identified KCNQ1 rs2237892 (P=0.030) and MTNR1B rs10830963 (P=0.041) associated with disposition index (DI) and GCKR rs780094 with triglyceride (P=0.036). The proportion of variation explained by GRS for the 17 QTs ranged from 0-2.7% (median=0.14%). GRS was only associated with 3 QTs: DI ($P=1.6\times10^{-5}$; 2.5% variation explained), acute insulin response (AIR, $P=4.6\times10^{-5}$; 2.7% variation explained), and glucose effectiveness (P=0.022; 0.68% variation explained). FS showed significant association with 16 of 17 QTs, with 3-11 SNPs entering into final models. The proportion of variation explained ranged from 1.1-8.5% (median=3.8%). The two strongest associations were for AIR ($P=2.7\times10^{-7}$; 11 SNPs accounting for 8.5% of variation) and DI ($P=9.9\times10^{-9}$; 8 SNPs accounting for 7.6% of variation). In summary, GRS was only associated with 3 of 17 QTs, while FS was associated with 16. GRS does not identify which loci contribute to the association with a given QT. FS identifies the subset of SNPs contributing to the association, providing detailed information facilitating physiologic interpretation. We conclude GRS may lead to misleading physiologic interpretation of genotype-phenotype relationships due to lack of detailed information.

1764S

Development and application of a population based statistical framework addressing the n=1 problem in human genetics. *A.B. Wilfert[1], M. Lek[2,3], J.N. Constantino[4], W.A. Gahl[5,6,7], E. Flynn[5,7], E. Valkanas[5,7], D.G. MacArthur[2,3], D.F. Conrad[1].* 1) Genetics, Washington University in St. Louis, St. Louis, MO; 2) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Psychiatry, Washington University in St. Louis, St. Louis, MO; 5) NIH Undiagnosed Diseases Program, NIH, Bethesda, MD; 6) Office of the clinical director, NHGRI, NIH, Bethesda, MD; 7) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD.

The diagnosis of rare, idiopathic diseases is emerging as a primary application of medical genome sequencing, and has motivated high-profile research efforts such as the NIH's Undiagnosed Diseases Program (UDP). However, the application of standard tools from genetic epidemiology for many of these cases is frustrated by a combination of small sample sizes, genetic heterogeneity and gene by environment interactions. In response, we have developed an inference framework that uses massive control population sequencing datasets to measure the disease potential of a genotype or diplotype in a sample size as small as a single individual, what we refer to as the "n=1" problem. Our framework is built on 3 principles that are advances on historically applied pathogenicity prediction algorithms. (1) Like others, we integrate multiple prediction algorithms to estimate the probability that a variant damages protein function. (2) Most prediction methods that we incorporate are **generic** in that they use the same model to annotate variation in all genes and are naïve to ploidy. We believe the best predictors will be **gene-specific** and **genetic** in nature and have taken steps towards this by integrating information on gene physiology using model-based measures of haploinsufficiency for all genes in the genome. (3) We use simulation, mutation rate modeling, and exome sequencing data from over 15,000 individuals to create a null model of "normal" genetic variation and estimate the probability of sampling a genotype or diplotype of a given pathogenicity score from healthy individuals. Our models have unmatched ability to discriminate putative disease mutations in the Human Gene Mutation Database (HGMD) from variants identified by population re-sequencing (AUC of 91% and 97%, respectively). We simulate over 30 million cases of the n=1 problem by spiking HGMD variants into 1092 healthy exomes from the 1000 Genomes Project. For 30% of all HGMD variants our gene-specific and genetic model always ranks them into the top 10 most pathogenic variants in an exome. The results of our modeling and spike-in analyses replicate when replacing HGMD with ClinVar data. We discuss the weaknesses of our approach and our attempts to identify areas for improvement. Finally, we apply our framework to a number of real-life n=1 cases, including a pair of twins from a consanguineous marriage concordant for Autism and several hundred cases from the NIH's Undiagnosed Diseases Program.

1765M

Meta-analysis on polygenic effects. *J.H. Zhao, J.A. Luan, S.J. Sharp.* MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, United Kingdom.

Background: It has been shown recently that whole genome data can facilitate estimation of genetic contributions to a variety of traits via a mixed model framework (Yang et al. Nat Genet 2010, 42:565-9; Zhao & Luan. J Prob Stat 2012, doi 10.1155/2012.485174). Our aim was to use this approach to investigate the impact of meta-analysis using study specific summary statistics such as (the genomic) heritabilities and/or polygenic/residual variance components as composed to those estimates from individual level data. Methods: We analysed body mass index from population-based and family-based cohorts, EPIC-Norfolk, Fenland, and Framingham studies involving Affymetrix 500K and Affymetrix 6.0 SNPs, involving ~10,000 individuals in total. We obtained genomic relationship matrices, which were then used in mixed models as implemented in the computer program GCTA and alternative procedures in PLINK, R. We compared results from meta-analysis using summary statistics and from individual level data. We also conducted simulation experiments. Results: Although the participants in the three cohorts were all European descent, there was variability in their genomic heritability estimates, with point heritability estimates being ~20%, 30% and ~50%, respectively. In the case of Framingham data, although the familial and genomic relationship matrices did not necessarily take equal values numerically they yielded comparable estimates. When polygenic effects are involved, meta-analysis using either heritability or variance components generally agreed with those from individual level data but with larger standard errors, which were likely to be larger than those assuming independence between inter-cohort participants in other scenarios not involving polygenic effects. The availability of whole-genome data seamlessly allows for meta-analysis of population-based and family-based data under the polygenic model. We have made the relevant implementations available in the R/gap package. Conclusion: The work supports the utility of whole genome data in heritability estimation as well as availability of individual level data so as to account for correlation among all individuals. Further work will be to assess the uncertainty due to sampling of the genome as with other aspects of association analysis in the presence of polygenic effects.

1766T

Estimation of causal effects distribution from genome-wide association studies. *L. Zhang[1], Y.F. Pei[1], Y. Lin[2].* 1) School of public health, Soochow university, Suzhou, JiangSu, China; 2) University of Shanghai for Science and Technology, Shanghai, China.

Estimation of the distribution of causal SNP effects and their heritability from genome-wide association study has been a research interest as it could explain in part the mystery of missing heritability. In this study, we propose a novel statistical method for such estimation. Specifically, we study the full range of GWAS summary results and link observed p-values and unobserved effect sizes by (non-central) chi-square distributions. By modeling the observed full set of GWAS p-values into a multinomial event, we build likelihood function in terms of causal SNP effects. We present both parametric and non-parametric forms of maximal likelihood estimation. The simulation studies showed that the proposed method had the ability to accurately estimate the number of causal SNPs and their effect sizes. As a real application, we analyzed a publicly available GWAS summary dataset that was released by the GIANT consortium, for height trait. Our analyses showed that there was a total of over 6,000 SNPs that might be associated with height, and explained ~40% heritability, where the number was much larger than that was previously estimated. As a conclusion, our proposed method has the potential to estimate common SNP-based genetic basis of complex traits from large-scale GWAS meta-analyses results.

1767S

Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *B.J. Vilhjalmsson[1,4], J. Yang[2], H. Finucane[1,4], A. Gusev[1,4], S. Lindstrom[1], S. Ripke[3], G. Genovese[4], N. Patsopoulos[4,5,6], P.-R. Loh[1,4], G. Bhatia[1,4], R. Do[6], M. Pato[7], C. Pato[7], H.-H. Won[4], PGC. SCZ Working Group of the PGC[8], S. Kathiresan[4], E. Stahl[9], N. Zaitlen[10], B. Pasaniuc[11], P. De Jager[4,5,6], S. McCarroll[6], D. Chasman[5], M. Daly[3], B. Neale[3], S. Purcell[8], M. Goddard[12], P. Visscher[2], P. Kraft[1], N. Patterson[4], A.L. Price[1,4].* 1) Harvard School of Public Health, Boston, MA; 2) The University of Queensland, Brisbane, Queensland, Australia; 3) Massachusetts General Hospital, MA, USA; 4) Broad Institute, Cambridge, MA, USA; 5) Brigham & Women's Hospital, Boston, MA, USA; 6) Harvard Medical School, Boston, MA, USA; 7) Keck School of Medicine, USC, Los Angeles, CA, USA; 8) Schizophrenia Working Group of the Psychiatric Genomics Consortium; 9) Icahn School of Medicine at Mount Sinai, New York, NY USA; 10) UCSF, San Francisco, CA USA; 11) UCLA, CA Los Angeles USA; 12) The University of Melbourne, Parkville, Victoria, Australia.

Polygenic risk scores are a promising approach for predicting disease risk, with prediction accuracy expected to improve substantially as sample sizes increase (Chatterjee *et al.* 2013 Nat Genet; Dudbridge *et al.* 2013 PLoS Genet). The standard approach involves pruning variants in linkage disequilibrium (LD) and applying a P-value threshold to association statistics, but this discards information and may limit predictive accuracy. We developed a Bayesian polygenic risk score, LDpred (https://bitbucket.org/bjarni_vilhjalmsson/ldpred), that uses LD from a reference panel to estimate posterior mean causal effect sizes from GWAS summary statistics, producing optimal (best linear unbiased prediction) polygenic risk scores when model assumptions hold. These estimates have a closed-form solution under an infinitesimal prior where every marker is causal and effect sizes are drawn from a mean-zero normal distribution. Under a non-infinitesimal prior where only a fraction of markers are causal, we approximate them via Markov chain Monte Carlo (MCMC). In simulations using real genotypes, our LDpred method is more accurate than existing methods (particularly at large samples sizes), well-calibrated, and computationally efficient. We applied LDpred to seven WTCCC diseases and observed substantial improvements in prediction R^2 (Nagelkerke), e.g. for type-1 diabetes it improved from 30.4% to 37.8%. We also applied LDpred to two diseases for which we had GWAS summary statistics in large sample size and an independent validation dataset. The prediction R^2 (Nagelkerke) improved from 15.7% to18.3% for schizophrenia (SCZ) and 9.8% to 12.1% for multiple sclerosis. The relative improvement was larger when predicting in African-American SCZ validation samples (R^2 improved from 2.0% to 2.6%).

1768M

Mediation Analysis of Integrated Genetic and Genomic Data in the Presence of Missing Data. *R. Barfield, X. Lin.* Harvard University, Boston, MA.

It is of increasing interest to analyze integrated different types of genetic and genomic data. Mediation analysis provides a useful tool for analysis of integrated genetic and genomic data to understand disease causing mechanisms. In genetic and genomic studies, SNP data, such as GWAS or sequencing data, are often collected on all individuals enrolled in a study. The genomic data, such as gene expressions and DNA methylations are often collected in a subset of study subjects. When the two types of data are combined to perform an functional analysis, the individuals who only had the SNP data collected are typically ignored. These individuals however can still provide useful information to the analysis. We propose a mediation analysis method using all the data by leveraging the information from the individuals with only the SNP data. We show using all available data, we gained more efficient estimators of the direct effects of SNPs and the indirect effects of SNPs mediated through gene expressions/DNA methylations on a phenotype with varying levels of missingness. As the proportion of individuals with missing data increased, we gain an increase in the efficiency of our estimates. We also show power gain in detecting genetic and genomic associations using all available data. We applied our method to data from the MESA cohort using DNA methylation data and SNP data with obstructive sleep apnea as the outcome.

1769T

Discovering Disease Susceptibility Genes Using Predictors of the Transcriptome - PrediXcan. *H.K. Im, E.R. Gamazon, K. Shah, S. Mozaffari, H.E. Wheeler, J. Torres, L. Davis, N.J. Cox.* The University of Chicago, Chicago, IL.

Genome-wide association studies have found thousands of variants robustly associated with complex traits. However, for most of them the underlying biological mechanisms are not well understood. The enrichment of gene expression associated variants (eQTLs) among disease-associated variants indicates that a substantial component of trait variability is likely to be determined through the regulation of the transcriptome. We propose a method termed PrediXcan that directly tests this hypothesis by correlating genetically predicted gene expression levels with the phenotype. For this purpose, we developed prediction models of gene expression using data from multiple eQTL studies (GTEx pilot data, GEUVADIS/1KGenomes, Framingham). Among advantages of PrediXcan are that 1) actual gene expression levels are not needed since they are calculated directly from the genetic data, 2) mechanism of the effect is directly built into the test making biological interpretation straightforward, 3) unlike actual differential expression studies, disease status cannot affect in silico levels thus ruling out reverse causality, 4) it tells whether up or down regulation drives the risk, 5) multiple testing burden is greatly reduced. Using these models we computed the in-silico expression levels of 15200 genes in six disease studies from the Wellcome Trust Case Control Consortium (WTCCC). For example, in Rheumatoid Arthritis 23 HLA region genes were differentially expressed (genomewide significant threshold p<3.3e-6). Outside of the HLA region, RSBN1 (p=2.4e-6) and PSME1 (p=1.5e-7) were of interest. We found many positive control genes as well as many novel ones with replication efforts under way. In conclusion, we propose a powerful gene based association test that directly tests the hypothesis that genetic variants affect disease status through the regulation of genes. Application to the WTCCC data identified many novel genome-wide significant genes.

1770S

Statistics for genetic association in the presence of covariates - genome scanning considerations. *H. Lin[1], E. Feingold[1, 2], Y. Lin[1].* 1) Department of Biostatistics , University of Pittsburgh, Pittsburgh, PA; 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

A number of different statistics are available for genetic association analysis in the presence of covariates. In the context of a genome-wide association study, hundreds of thousands to millions of SNPs are tested, and whatever covariate model we specify is likely to be imperfect., In addition, the results of the study often focus on the list of SNPs ordered according to the statistics rather than on certain p-value cutoffs. Therefore, it is important to investigate the behavior of extreme values of the statistics rather than the behavior of the expected values. Gail et al. (2008) discussed this issue and proposed “detection probability” and “proportion positive” to measure the success (power) of a genomic study when ranked lists are the primary outcome. In theory, the ranked lists can be dominated by SNPs with misfit models rather than by true positive results. We are conducting a comprehensive comparative study to investigate the behavior of different association statistics that model covariates. We evaluate the statistics from the perspective of which statistics can provide robust ranked lists of “top hits.” These are not necessarily the same statistics that have the highest power in a conventional single-test context.

1771M

Addressing Potential Bias in Heritability and Coheritability Estimates within Ancestrally Homogeneous Populations. *J. Liu[1,2,3], T.J. Hoffmann[1,2], J.S. Witte[1,2,3,4].* 1) Department of Epidemiology & Biostatistics, UCSF, San Francisco, CA; 2) Institute for Human Genetics, UCSF, San Francisco, CA; 3) Diller Family Cancer Center , UCSF, San Francisco, CA; 4) Department of Urology, UCSF, San Francisco, CA.

A substantial heritability of complex traits can be explained by considering all SNPs across the genome simultaneously using linear mixed-effects models with an estimated genetic relationship matrix (GRM). This approach has been extended to assess the coheritability—or potential shared genetic basis—among different traits, including psychiatric disorders, inflammatory bowel disease, and major chronic obstructive pulmonary disease-related traits. However, the estimated GRM used in such (co)heritability analyses depends on sample-based allele frequency estimates across the genome. Any systematic differences in allele frequency estimates among the study sample can lead to a biased GRM and incorrect (co)heritability estimates. For example, when calculating the coheritability of a trait across different ancestral populations, differences in their allele frequencies can over- or under-estimate the relatedness between ancestrally similar individuals, resulting in more exclusions of samples from the analysis, and biased GRM and (co)heritability. As shown previously, one can address this issue with (co)heritability among ancestrally distinct or admixed populations by calculating the GRM by using individual-specific (or population-specific) SNP allele frequencies based on the individual ancestry. However, little if any consideration has been given to the potential impact of this issue on (co)heritability estimation in relatively homogeneous populations. It is increasingly common to combine samples from different studies with similar ancestral backgrounds for (co)heritability analysis. Here, samples from different studies may typed on different genotype platforms or have distinct cryptic relatedness. As with ancestrally distinct or admixed populations, ignoring these differences can bias the GRM and in turn biased (co)heritability estimates. We highlight this issue with two examples: 1) the bias in estimating lung cancer heritability within a European population comprised of samples from three studies; and 2) the bias in estimating the coheritability of lung-breast cancer within European population in which the traits are measured on different samples. In addition, we show that the allele frequency adjustment approach mentioned above is able to correct the bias efficiently. Our findings emphasize the need to control for the bias due to allele frequency differences between groups of samples even when they appear to be from a relatively homogeneous population.

1772T

A new prognostic model to predict renal outcome in autosomal dominant polycystic kidney disease (ADPKD). *E. Cornec-Le Gall[1-4], MP. Audrézet[2-4], M. Hourmant[5], MP. Morin[6], C. Charasse[7], E. Renaudineau[8], B. Wehbe[10], P. Jousset[11], MP. Guillodo[12], N. Terki[13], S. Benarbia[14], R. Perrichot[16], T. Sawadogo[9], S. Régnier-Le Coz[15], C. Savoiu[17], Y. le Meur[1-2], C. Ferec[2-4, 18], the Genkyst Study Group.* 1) Department of Nephrology, Centre Hospitalier Universitaire de Brest; 2) Université Occidentale de Bretagne; 3) INSERM U1078; 4) Department of Molecular Genetics, Epidemiologic Genetic and Histocompatibility, Centre Hospitalier Universitaire de Brest; 5) Department of Nephrology, Centre Hospitalier Universitaire de Nantes; 6) Department of Nephrology, Centre Hospitalier Universitaire de Rennes; 7) Department of Nephrology, Centre Hospitalier de Saint Brieuc; 8) Department of Nephrology, Centre Hospitalier de Saint Malo; 9) Department of Nephrology, Centre Hospitalier de Lorient; 10) Department of Nephrology, Centre Hospitalier de Quimper; 11) Department of Nephrology, Centre Hospitalier de Pontivy; 12) Centre de Dialyse, Association des urémiques de Bretagne, Brest; 13) Société Brestoise du Rein Artificiel, Brest; 14) Centre de Dialyse, Association des urémiques de Bretagne, Quimper; 15) Department of Nephrology, Centre Hospitalier de Saint Nazaire; 16) Department of Nephrology, Centre Hospitalier de Vannes; 17) Centre associatif de Dialyse, ECHO, Nantes; 18) Etablissement Français du Sang, Brest.

ADPKD is marked by a high clinical variability, especially in regards of age at end-stage renal disease (ESRD). As we are shifting to the era of targeted therapies in ADPKD, one needs to target the patients who are more likely to develop ESRD and thus to benefit from those therapies. In that setting, we conducted a cross-sectional study in a population of 1130 patients (525 males) from all the nephrology centers (n=17) of a single area, Brittany, France. We evaluated the effect of 10 clinical variables and of the molecular genetic data (mutation of PKD2, non truncating mutation (NTM) of PKD1 or truncating mutation (TM) of PKD1) using Kaplan-Meier curves and univariate followed by multivariate Cox regression analyses and built a score weighting each significant factor according to the HR obtained. Internal validation was assessed using non parametric bootstrapping with replacement. Discrimination of the model was evaluated using the c-statistic, which represents area under time-dependant ROC curves. Median age at inclusion was 54.3 yrs [5.45-94.3]. After multivariate analysis, 4 factors remained significantly associated with renal survival : Gender (HR=1.55, 0.95 CI=1.25-2.93), Hypertension onset before age 35 (HR=2.24, 0.95 CI= 1.76-2.86), first urological complication before age 35 (at least 1 amongst gross hematuria, flank pains related to cysts or cyst infection) (HR=2.1, 0.95 CI=1.62-2.73), and genetic status (PKD2 vs NTM of PKD1 HR=1.93, 0.95CI= 1.27-2.93, vs TM of PKD1 HR=4.88 0.95CI=3.39-7.02). We thus defined a score ranging from 0 to 9 (male gender:1 pt, hypertension onset before age 35: 2 pts, first urological complication before age 35: 2 pts, PKD2 mutation: 0 pt, NTM of PKD1: 2 pts and TM of PKD1: 4 pts). The accuracy of the PRO-PKD score applied to the sample was high, with a c-statistic of 0.863 at 65 yrs. To facilitate ease-of-use, we defined 3 risk categories: low risk, intermediate and High-Risk of early progression to ESRD with median ages at ESRD of respectively 73.6, 57.7 and 48.5 yrs (HR: Intermediate Risk= 3.99, High-Risk=13). Predicted probabilities of ESRD at age 60 were 16.5%, 57.5%, and 91.2%; for the Low, Intermediate, and High Risk Group, and accorded well with observed ESRD frequencies. The PRO-PKD score is a simple, accurate and easy-to-use prognostic score that might be of interest in the selection of the patients who should be included in the ADPKD clinical trials and be a valuable tool for personalized medicine in ADPKD.

1773S

A fast and accurate p-value imputation approach for genome-wide association study. *J. Kwan, M. Li, J. Deng, P. Sham.* HKU, Hong Kong, Hong Kong, Hong Kong.

Imputing individual-level genotypes (or genotype imputation) is now a standard procedure in genome-wide association studies (GWAS) to examine disease associations at untyped genetic markers. But the increasing GWAS sample and marker density of imputation reference panel means that more time is needed for imputing a single chromosome, while researchers are only interested in assessing the evidence of disease association at different markers in GWAS. Also, meta-analysis or expanded meta-analysis of publicly available GWAS summary statistics can allow more disease-associated loci to be discovered, but these data are usually provided for various marker sets and this makes such analyses difficult. Therefore, we develop a fast and efficient method that utilizes summary statistics only for imputation, called P-value Imputation (PI) at common variants. Its computational cost is linearly with the number of untyped markers and has little loss in accuracy compared with IMPUTE2, one of the leading methods in genotype imputation. In addition, based on the PI idea, we develop a metric to detect abnormal association at a marker using the association p-values at neighboring markers in LD and showed that it had a significantly greater power compared to LD-PAC, a method that quantifies the evidence of spurious associations based on likelihood ratio. All of our methods are implemented in a user-friendly software tool freely available at http://statgenpro.psychiatry.hku.hk/pi for meta-analysis conveniently.

1774M

Generalized likelihood ratios ensure statistically and clinically significant findings: application to genetic association with cystic fibrosis lung disease. *W. Li[1,2], L. Strug[1,2].* 1) Department of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; 2) Child Health Evaluative Sciences, Hospital for Sick Children, Toronto, ON, Canada.

P-values have been criticized for their use as a measure of statistical evidence. One reason is that it can lead to statistically significant but clinically irrelevant findings, especially with large sample sizes as we see in GWAS. The evidential paradigm uses the likelihood ratio evaluated at two simple hypotheses (LR) as a measure of statistical evidence, with k=8 or 32 as commonly used benchmarks for declaring strong evidence. The probability of observing a large LR favoring an incorrect hypothesis, M, is bounded for normal data and fixed-dimensional parametric models in large samples (Royall, 2000), suggesting the LR is a valid evidence function and does not lead one to incorrect conclusions often. However, the requirement of a simple alternative hypothesis can limit applicability in genomic applications such as GWAS, where the focus is on hypothesis testing rather than estimation, and there is no a-priori reason for choosing a specific alternative hypothesis value. Zhang (2009) and Bickel (2012) provided theoretical justification for extending the evidential framework to accommodate composite hypotheses through the use of a generalized likelihood ratio (GLR), where the evidence in support of one composite hypothesis $\theta_1 \in \theta$ over its composite complement, $\theta_2=\theta_1^c$ is given by the ratio of the two likelihoods maximized over θ_1 and θ_2, separately. We analytically derive the operational characteristics associated with the use of GLR. For normal data, we show that M has the same bound as shown in Royall (2000), and it appears to hold generally outside of the normal model. In a candidate gene study of Cystic Fibrosis (CF) lung disease (n=753), we compare the association evidence between the maximized likelihood ratio (LR_m) test p-value and GLR at 8 SNPs across 3 genes. We define θ_1=[-C, C] and $\theta_2=\theta_1^c$ with C > 0 chosen to define a null region, θ_1. The LR_m, its p-value and GLRs with C=0.001 and 0.05 are computed. As $C \to 0$, $GLR \to LR_m$, e.g. when LR_m=8.14, GLR=7.93 for C= 0.001. For C=0.05, corresponding to a 1 percent change in the lung phenotype, the GLR provides evidence that 2 SNPs are associated with CF lung disease at k=8, the same conclusion as from the multiple comparison adjusted LR_m test p-value. The GLR extends the evidential framework by accommodating composite hypotheses, making the paradigm more applicable to GWAS. GLRs capture clinically meaningful findings, avoiding one of the major limitations of p-values in studies with extremely large sample sizes.

1775T

Fine-mapping of additive and dominance effect SNPs using group-LASSO and Fractional Resample Model Averaging. *J.A. Sabourin[1,2], A.B. Nobel[3,4], W. Valdar[2].* 1) Genometrics Section, CSGB, NHGRI, NIH, Baltimore, MD; 2) Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Genome-wide association studies sometimes identify loci where both the number and identities of the underlying causal variants are ambiguous. In such cases, statistical methods that model effects of multiple SNPs simultaneously can help disentangle the observed patterns of association and provide information about how those SNPs could be prioritized for follow-up studies. Current multi-SNP methods, however, tend to assume that SNP effects are well captured by additive genetics. However, when genetic dominance is present, this assumption translates to reduced power and faulty prioritizations. We describe a statistical procedure for prioritizing SNPs at GWAS loci that efficiently models both additive and dominance effects. Our method, LLARRMA-dawg, combines a group LASSO procedure for sparse modeling of multiple SNP effects with a resampling procedure based on fractional observation weights. It estimates for each SNP the robustness of association with the phenotype both to sampling variation and to competing explanations from other SNPs. In producing a SNP prioritization that best identifies underlying true signals, we show that: 1) our method outperforms a single marker analysis; 2) when additive-only signals are present, our joint model for additive and dominance is equivalent to or only slightly less powerful than modeling additive-only effects; and 3) when dominance signals are present, even in combination with substantial additive effects, our joint model is more powerful than a model assuming additivity.

1776S

Allele Specific Expression Can Reduce Apparent Genotype/Phenotype Relations: A Simulation Study. *J.L. Dannemiller.* Psychology Dept, Rice University, Houston, TX.

Allele specific expression (ASE) can affect phenotypes. **Purpose:** To estimate the effects of unrecognized ASE on a gene association study in which it is incorrectly assumed that both alleles are equally expressed. **Methods:** Gillespie's (1976) Monte Carlo method of simulating transcription and translation was used to estimate the effects of unrecognized ASE on a standard effect size measure: the percentage of phenotypic variance explained by genotype (ExplVar). Representative rates were assumed for mRNA transcription and degradation and for protein translation and degradation. A biallelic gene (p = q = 0.5) in Hardy-Weinberg equilibrium was modeled. In each experiment, 200 cells were simulated for each of 600 subjects until the protein levels from the two alleles in each cell reached independent equilibria. One of the alleles produced a protein that was 25% more biologically active than the other protein. In no-ASE experiments, the rates of mRNA expression for both alleles were the same. In ASE experiments, the mRNA expression rate for the low activity allele was the same as it was in the no-ASE experiments; the rate for the other allele was randomly drawn from a lognormal PDF similar to a distribution of ASE ratios in Valle et al. (2008) for the gene *TGFBR1*. The mean ASE ratio in the theoretical distribution was approximately 2:1. I am also exploring MCMC to generate the ASE ratios with a sum of weighted Gaussian kernels approximating the Valle et al. target distribution. The final phenotype for each cell was the sum of the number of protein molecules at equilibrium times the activity level of each of the protein molecules (1.00 or 1.25). Zero-mean, constant variance Gaussian noise was added to the average final activity level for each of the 600 subjects across their 200 cells to reduce the average percentage of explained variance under no-ASE conditions to approximately 6% - a level indicating a polygenic phenotype. Genotype values of 0, 1, or 2 (number of higher-activity allele doses) were used in the computation of ExplVar. **Results:** Across 100 experiments of each type, unrecognized ASE approximately halved the mean percentage of ExplVar from 6% to 3%. There were two sources of additional within-genotype variance in the ASE experiments: 1) random variance in the ASE ratios, and 2) higher variance in final protein levels because of generally higher mRNA expression rates. Unrecognized ASE can reduce estimates of genotype/phenotype relations.

1777M

Parent of origin and recurrence risk bias: probabilistic modeling explains the broken symmetry of transmission genetics. *C. Shaw[1], I. Campbell[1], P. Olofsson[2].* 1) Dept Molecular & Human Gen, Baylor Col Medicine, Houston, TX; 2) Mathematics Department, Trinity University, San Antonio, TX.

Most de novo human mutations are observed to arise in fathers, and increasing paternal age strongly correlates with increased risk of paternally transmitted mutations. Strikingly, de novo mutations in X-linked recessive disease show strongly elevated familial recurrence rates, and in male offspring these mutations must be transmitted from mothers. We were intrigued by these contrasting patterns of mutation transmission and their consequences for human genetics. We developed a comprehensive stochastic process model of mutation in the human germ line, closely modeling the sexual dimorphisms between males and females and considering the effect of increasing paternal age. We extended our prior work on exact sampling formulas for multi-type Galton-Watson branching processes to develop new formulas for the mean and variance of the proportion of mutant gametes in each parent. Analysis of these models and new mathematical results for updating the mean and variance of the proportion of mutant gametes based on observation of an affected child led to unexpected predictions for recurrence risk that implicate parent of origin as a central variable in the analysis of recurrence risk. Consistent with empirical data, our model predicts that more transmitted mutations arise in fathers, and that this tendency increases with a father's age. Surprisingly, the variance in the proportion of mutant gametes is lower in fathers and decreases with age, so that knowledge of a transmitted mutation has small effect on updating the expected proportion of mutants and therefore recurrence in another offspring. Quite differently, for the female germ line, which arrests after clonal expansion in early development, variance in the proportion of mutants is much higher, and observation of a transmitted mutation in an affected child dramatically increases the risk of recurrence in another pregnancy. Review of the literature where parent of origin was determined in families with recurrent genetic disease shows that the predictions of our model are strongly consistent with data that support a maternal origin bias of recurrent de novo mutations. Our findings have important consequences for genetic counseling and for understanding the patterns of recurrence in transmission genetics.

1778T

What are genome-wide association studies detecting? Our experience predicting cystic fibrosis-related diabetes onset. *D. Soave[1,2], M. Miller[1], K. Keenan[3], W. Li[1,2], J. Gong[1], W. Ip[3], P. Durie[3,4], L. Sun[2,5], J. Rommens[6,7], L. Strug[1,2].* 1) Program in Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 3) Program in Physiology and Experimental Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Pediatrics, University of Toronto, Toronto, Ontario, Canada; 5) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 6) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

It is assumed that genome-wide association studies are identifying genetic determinants of the phenotype under study. However, we were reminded that this may not always be the case when a recent genome-wide association study for Cystic Fibrosis (CF)-related diabetes (CFRD) identified associated SNPs in *SLC26A9* (HR=1.34, at rs7512462). SLC26A9 is an apical epithelial cell transporter expressed in human pancreas, but is unlikely to have a direct role in the endocrine compartment as its protein payload is delivered to the bloodstream and surrounding cells without a specialized ductal system. We recently showed that rs7512462 accounts for >10% variability in early exocrine pancreatic damage in children with severe CF; and Mendelian randomization provided evidence that this early exocrine pancreatic damage causes later-onset of CFRD. These findings may explain how the SLC26A9 epithelial transporter contributes to CFRD. Despite only an indirect association between *SLC26A9* and CFRD risk, we hypothesized that *SLC26A9* may still assist with predicting CFRD onset. We developed and validated CFRD predictive models using 126 participants from the Canadian CF Gene Modifier Study on whom longitudinal measures of immunoreactive trypsinogen (IRT; the early exocrine pancreatic damage biomarker) and rs7512462 genotype were available. Cox proportional hazards models with 10-fold cross-validation were used to generate risk scores for each individual and these risk scores were used to assess the CFRD predictive ability for various ages using time-dependent receiver operating characteristic curves and the corresponding area under the curve (AUC). Using risk scores calculated from estimated IRT at birth and its rate of decline to predict CFRD by age 12, 16, 20 and 24y, our cross-validated AUC's were 65%, 64%, 64% and 78%, respectively. With the addition of rs7512462 genotype into the risk score calculation, the predictive accuracy improved significantly to 73%, 74%, 74% and 80%, respectively. Although we are cognizant that genome-wide significant SNPs require follow-up to identify the underlying causal variant that induced the association, less consideration is paid to identifying the underlying phenotype to which the variant is more proximally associated. Our findings suggest that once we better model the pancreatic phenotypic complexity, the addition of rs7512462 genotype can significantly improve predictive accuracy despite having only a weak and indirect effect on CFRD.

1779S

Test of Genotypic Association Allowing for Errors. *L. Zhou, A. Musolf, D. Londono, D. Gordon, T. Matise.* Genetics, Rutgers, the State University of New Jersey, Piscataway, NJ.

Genome-wide association studies (GWAS), have led to identification of an ever-increasing number of single nucleotide polymorphisms (SNPs) for further studies. However, the presence of genotype misclassification errors, differential or non-differential, among cases and controls may cause either an increase in the false positive rate or a decrease in the power of statistical tests. A commonly used statistic to test for association between SNPs in cases and controls is the chi-square test of independence. This statistic tests whether genotype frequencies (single or multi-locus) differ between case and control groups. This is commonly referred to as a test of association. Regions of the genome where frequencies significantly differ are regions that may harbor a disease locus or loci. Several researchers have documented that genotype misclassification errors may adversely affect the chi-square test. There are three main reasons for genotype errors. First, as a result of inaccurate genotyping technology, the sequencing procedure may introduce errors in the output. The second reason may be low quality of the sequencing samples. Last, if samples from cases and controls are sequenced at different times or by different methods, this approach may lead to different error rates between cases and controls (differential misclassification). Our overall objective is the development of a statistical test of association that uses NGS data and is robust to random sequencing error, both non-differential and differential. More concretely, we develop a chi-square test of independence that uses parameters such as the observed alternative variant reads at a given polymorphic site, the coverage per individual, the individual's phenotype, and error model parameters. This statistic is developed in a log-likelihood framework.

1780M

Statistical method for analyzing allele-specific expression across individuals for multiple statuses. *Y. Lee, X. Wen, M. Boehnke.* Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI.

It is well known that genetic variants in cis-regulatory regions can have large effects on the expression level of the target genes. Such effects can be detected by eQTL mapping and allele-specific expression (ASE) analysis. In this study, we are interested in examining the concordance between eQTL and ASE analysis, especially when the regulatory variants show state or tissue-specific effects. To this end, we develop a novel statistical method to perform ASE analysis using RNA-seq data across multiple individuals. One of the unique features of our method is that we require genuine ASE signals to be consistent across individuals, while accounting for potential heterogeneity among individuals. Also, using Bayesian model averaging, we effectively take phasing uncertainty into account. We apply this method to the RNA-seq data from GEUVADIS and Genotype-Tissue Expression (GTEx) project. In particular, we attempt to confirm strong (tissue-specific) eQTL signals using ASE analysis, and we also evaluate the consistency of effect size from eQTL mapping and ASE analysis.

1781T

A New Approach to finding Association with Complex, Longitudinal Phenotypes using Population Data. *A.M. Musolf[1], D. Londono[1], A.Q. Nato, Jr.[2], P. Vuistiner[3], J. Brandon[4], J.A. Herring[5], C.A. Wise[4,6], H. Zou[7], M. Jin[7,8], L. Yu[1,9], S.J. Finch[10], P. Bovet[11], M. Bochud[3], T.C. Matise[1], D. Gordon[1].* 1) Department of Genetics, Rutgers University, Piscataway, NJ; 2) Statistical Genetics Lab, Division of Medical Genetics, University of Washington, Seattle, WA 98195, USA; 3) Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; 4) Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, TX 75219, USA; 5) Department of Orthopedic Surgery, Texas Scottish Rite Hospital for Children, Dallas TX 75219, USA; 6) Department of Orthopedic Surgery, University of Texas Southwestern Medical Center, Dallas, TX 75219, USA; 7) Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; 8) ShanghaiBio China, Pudong, Shanghai 201203, China; 9) Center of Alcohol Studies, Rutgers University, Piscataway, NJ 08854, USA; 10) Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11790, USA; 11) Ministry of Health, P.O. Box 52, Mont Fleuri, Republic of Seychelles.

Previously, we detailed a new method for testing for association between longitudinal phenotypes and causal genotypes. The method uses growth mixture models to determine longitudinal trajectory curves. The Bayesian posterior probability (BPP) of belonging to a specific curve was then used as a quantitative phenotype in association analyses. To identify association for multiple SNPs, we did not perform association analyses on individual SNPs; instead the genome was sliced into blocks of 50 SNPs. A significance value was obtained on each block via the program TDT-HET. This method displayed greater than 80% empirical power in most simulations scenarios. This method exclusively used family-based data. Here, we extend the method to population-based data sets. The method maintains many ideas from the family-based method. However, the program SumStat is used to acquire significance levels on each block, instead of TDT-HET. Multiple scenarios are tested including four causal variants located within a single locus and eight causal variants spread between two loci on different chromosomes. Reduced models using environmental covariates were also considered. Our data set was highly stratified to ensure robustness in the presence of population stratification. To correct for population stratification, ancestry fractions from the program ADMIXTURE are regressed on the BPPs and the residuals are used as the phenotype for association analyses. Our method also utilized three distinct data sets, which represent a discovery data set and two confirmatory data sets. The final p-values of the association analyses were combined via Fisher's method and corrected for multiple testing by the false discovery rate (FDR). We report that our simulations: 1) maintain the proper type I error in the presence of population stratification and 2) have greater than 90% power for all simulations. We conclude that our method can detect multiple causal SNPs located in multiple loci in population data sets. We believe that is method will be use to researchers who are studying complex diseases that display longitudinal phenotypes. It allows for high detection of causal loci (and the causal variants within) for both population and family studies, even in the presence of confounding elements such as population stratification and environmental variables.

1782S

Multiple testing procedures for GWAS with high-dimensional phenotypes. *C.B. Peterson[1], M. Bogomolov[2], Y. Benjamini[3], C. Sabatti[1].* 1) Department of Health Research and Policy, Stanford University, Stanford, CA; 2) Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Technion City, Haifa, Israel; 3) Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel.

We are developing statistical methods to both improve power and reduce errors in identifying genetic variants that are relevant to multivariate phenotypes such as imaging features, actigraphy measures, or gene expression. Since both the predictor and response variables are high dimensional, this represents a massive multiple testing problem.

A standard approach to handle this multiplicity is to control the false discovery rate across the entire set of hypotheses representing the association of each SNP to each phenotype. Unfortunately, this method does not control the false discovery rate for SNPs or the average false discovery rate for selected SNPs. This is a serious drawback for gene mapping studies, where the identification of SNPs with functional effects is one of the primary goals. In addition, it fails to account for the genetic architecture of traits. In this type of study, most genetic variants will not affect any of the phenotypes of interest, but those with functional effects are likely to influence multiple traits. In applying error control to the pooled set of hypotheses across all SNPs, we are reducing power to detect the effects of functional SNPs, while increasing errors for SNPs that have no influence on any traits.

In this work, we employ a two-step selection procedure recently proposed by Benjamini and Bogomolov that allows us to both gain control of the appropriate error rates and improve power by taking advantage of the genetic structure of the problem. Here we explore the application of this strategy and its performance in the context of GWAS. Specifically, we define families of hypotheses by SNP, where each family consists of hypotheses on the association of a particular SNP to each of the phenotypes under study. In step (1), we perform selection of SNPs that affect one or more phenotypes under study, based on a summary statistic for each SNP family. In step (2), we apply an error controlling procedure within each selected family using an adjusted level to account for the selection bias incurred in step (1). We demonstrate the advantages of this method in terms of error control and power through simulation studies using both independent predictors and real genotypes obtained from the North Finland Birth Cohort (NFBC) study.

1783M

Making use of parental phenotypes in case-parent genetic studies. *M. Shi, D.M. Umbach, C.R. Weinberg.* Biostatistics Br, NIEHS, Res Triangle Park, NC.

In studies of case-parent triads, information may be available about occurrence of the condition in the parents. Typically parental phenotypes are ignored, but including that information in analyses may increase power to detect genetic association for autosomal variants. We propose a method of using parental phenotypes that assesses association independently of the usual case-parent-based association test. Our proposed method enjoys many advantages of case-parents designs: it is robust to population stratification and allows testing of maternally-mediated genetics effects as well as inherited-gene effects. The parental information can enable a natural internal replication for findings based on offspring and their parents, and it can be used to improve power. We develop composite tests that combine evidence from this parent-phenotype-based test with the traditional log-linear transmission-based test. A likelihood-based approach builds information from these two sources into a single coherent model: The model can impose equality of parental and offspring relative risks and permit testing of that equality. This model also allows us to use the expectation-maximization algorithm when some parental genotypes are missing. We evaluate the proposed method through non-centrality parameter calculations and simulation studies. We show that incorporation of parental phenotype data often improves power to detect a genetic contribution to the phenotype. As illustration, we apply our method to a nuclear-family-based study of young-onset breast cancer.

1784T

Efficient multiple imputation for missing phenotype using genome-wide DNA methylation data. *W. Guan[1], C. Wu[1], B. Wu[1], Y. Li[2,3], J. Pankow[4], E.W. Demerath[4], J. Bressler[5], M. Fornage[5], M.L. Grove[5], T. Mosley[6], C. Hicks[6], E. Boerwinkle[6].* 1) Division of Biostatistics, Univ of Minnesota, Minneapolis, MN; 2) Department of Genetics, University of North Carolina, Chapel Hill, NC 27516; 3) Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27516; 4) Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN 55455; 5) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX 77030; 6) Department Neurology and Department of Medicine, School of Medicine, University of Mississippi, Jackson, MS 39216.

DNA methylation is a widely studied epigenetic mechanism and alterations in methylation patterns may be involved in the development of common diseases. Unlike inherited changes in genetic sequence, variation in site-specific methylation varies by tissue, developmental stage, disease status, and may be impacted by aging and exposure to environmental factors such as diet or smoking. While these wide-range correlations pose analytical challenges, including reverse causality and confounding by non-genetic factors in epigenome-wide association studies (EWASs), it brings opportunities to infer missing phenotype values using rich methylation data. We propose a multiple imputation method using large-scale methylation data, such as that provided by the Illumina Infinium HumanMethylation450 (HM450) BeadChip. Based on established methylation-phenotype association at multiple CpG sites, plausible values for missing data in phenotype variables can be drawn multiple times. The imputed datasets can then be analyzed separately using standard complete-case analysis, with the point estimates and corresponding covariance matrix being combined subsequently for final inference. Through simulations, we show that the proposed method can make appropriate association inference while maintaining correct false positive rate. We also illustrate our method with data from the Atherosclerosis Risk in Communities (ARIC) study to carry out an EWAS between methylation levels and smoking status, in which missing cell type compositions and white blood cell counts are imputed.

1785S

Using local multiplicity to improve effect estimation from a hypothesis generating study. *W. Zou[1], H. Ouyang[2].* 1) InVentiv Clinical Health, LLC, 504 Carnegie Center, Princeton, NJ 08540 USA; 2) Lilly Corporate Center, Eli Lilly and Company, Indianapolis, Indiana 46285.

We propose a multiple estimation adjustment (mea) estimator as an alternative way to correct for the effect over-estimation in a selection of strong genetic associations from a hypothesis generating study (HGS). Mea uses a hierarchical Bayesian approach to model maximal likelihood estimation (MLE) of individual effects and to integrate local multiplicity by borrowing effect information from neighboring variations. Unlike many other adjustment methods that model certain types of selection bias specifically, mea attempts to improve effect estimation without stipulating a selection scheme. Nevertheless, between naive MLE, mea and conditional likelihood adjustment (CLA) method that models ascertainment selection bias, we compared their mean square errors (MSE) among top signals that pass a threshold selection, a ranking selection or their combination though simulation studies. We show that selection and data-generating parameters impact MSE of all methods, and no methods dominate the others under all settings for both null and causal effects. We observe that mea effectively reduces MSE from MLE on null effects with or without selection, and has a clearer advantage over CLA for effects that are far away from a selection threshold. We believe mea is a valuable complement to available methods and an actionable tool on its own when interpreting results from HGS and designing follow-up studies.

1786M

Confounded by Ancestry? Considerations for Ancestry Adjustments in Genetic Association Tests. *E.R. Martin[1], I. Tunc[1], Z. Liu[1], M.A. Schmidt[1], C.D. Bustamante[2], G.W. Beecham[1].* 1) Hussman Institute for Human Genomics, Dr. John T. Macdonald Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 2) Department of Genetics, Stanford University. Palo Alto, CA.

It is well known that ancestry differences can lead to confounding in genetic association tests. It has now become standard practice to adjust for principal components that capture global ancestry of samples, but this is often done without assessment of whether they are true confounders. Adjusting for covariates that are not true confounders can lead to loss of power; thus, judiciously assessing whether ancestry is a confounder prior to analysis is likely to improve power in association tests. In addition, there have been recent reports considering whether it is preferable to adjust for local or global ancestry, but the conclusions depend highly on the model for population structure and trait/disease model. Here, we review the definition of statistical confounding in general and the properties of covariate adjustment (with and without confounding) in relation to statistical power. We then examine theoretical expectations to illustrate when we might expect confounding and the consequence of ancestry adjustment in genetic association tests in the context of various genetic models for population structure and admixture. Theoretical calculations were verified and generalized to more complex situations using simulations. We show that in a single admixed population, when testing the trait locus itself there is no confounding by ancestry, and in this scenario, adjusting unnecessarily for ancestry can lead to a substantial decrease in power. In stratified admixed populations when the trait values or disease prevalence differ between strata, global ancestry is a confounder and must be adjusted for to obtain valid inference. Generally adjusting for global ancestry captures the effect and there is no additional benefit adjusting for local ancestry. The one scenario that we examined where we observed a local ancestry effect in addition to a global ancestry effect was testing a nearby marker locus rather than the trait locus itself. Since the true population dynamics and genetic model are usually unknown, we propose a variable-selection strategy for assessing whether ancestry adjustment is necessary. Our simulations show that this strategy maintains correct Type I error, has good power compared to the true model, and has greater power than arbitrarily adjusting for ancestry covariates. This data-informed strategy provides a viable alternative to ubiquitous adjustment that preserves power while protecting against spurious results.

1787T

Towards Estimation of the all-Phenotype by all-Phenotype Genetic Correlation Matrix. *B. Bulik-Sullivan[1,4], H. Finucane[3], A. Price[2,5], B. Neale[1].* 1) Medical and Population Genetics, Broad Institute, CAMBRIDGE, MA; 2) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 3) MIT Department of Mathematics, Cambridge, MA; 4) VU University Amsterdam, Amsterdam, NL; 5) Department of Epidemiology, Harvard School of Public Health, Boston, MA.

Quantification of the extent of shared genetic architecture between disease phenotypes is an issue of fundamental interest to statistical genetics. Such quantification could help refine nosology, explain co-morbidity, and increase power in association studies and precision in genetic risk prediction by leveraging GWAS data collected for related phenotypes Existing methods for estimating genetic correlation from genotype data, e.g., bivariate Haseman-Elston (HE) regression, and bivariate linear mixed models (LMM) require researchers to obtain access to large datasets with genotype and phenotype data for multiple traits; however, this is often infeasible because of privacy concerns that limit sharing of identifiable genetic data and medical records. For this reason, estimates of genetic correlation published in the literature to date have been limited to pairs of phenotypes studied within one consortium or lab. We present a bivariate extension of the recently proposed LD Score regression method (Bulik-Sullivan, et al, 2014, biorxiv) that allows estimation of genetic correlation from GWAS summary statistics. This method is supported by theory and extensive simulation, and performs robustly under various genetic architectures, even when one or both GWAS is confounded by cryptic relatedness or mild population stratification (similar to univariate LD Score regression). In addition, this method remains unbiased no matter how many samples are shared between the two GWAS, which is of great practical importance, since many GWAS consortia (e.g., the Psychiatric Genomics Consortium) use the same control set for many phenotypes. We validated the method on real data by replicating the genetic correlation estimates from the PGC Cross-Disorder Group Nature Genetics 2013 paper with similar standard error (as a representative example, the PGC CDG paper reports the genetic correlation between bipolar disorder and sczhiophrenia to be 0.68 (0.04); we get 0.64 (0.06) using summary statistics from the same samples). We present results from the application of this method to a large set of GWAS summary statistics, including many meta-analyses for which genotype data are not shared even within the consortium (e.g., BMI and diabetes). Finally, we discuss the mathematical interpretation of LD Score regression as a linear time (in the number of samples) approximation to existing methods (such as HE-regression and LMM) that require raw genotypes and quadratic or cubic time.

1788S

Effective genetic risk prediction using mixed models. *D. Golan, S. Rosset.* Statistics, Tel-Aviv University, Tel-Aviv, Israel.

Despite identifying thousands of genetic variants associated with dozens of high-impact diseases using genome-wide association studies, our ability to accurately predict the disease status using genetic data remains disappointingly low for many highly heritable diseases. One leading theory is that highly heritable diseases for which prediction is difficult (e.g., bipolar disorder, type-2 diabetes and hypertension) are typically driven by many common variants with small effects. The effects of most variants are too small to reach significance, but cumulatively these variants account for a considerable portion of the disease burden. This theory suggests that using only those variants which were found to be significantly associated with the disease is too conservative, and that a more permissive inclusion criterion should be used when selecting variants for the purpose of prediction. However, there is an obvious trade-off: when the inclusion criterion is stringent, a considerable part of the signal is left out, but when using a more permissive inclusion criterion, the number of estimated parameters (effect sizes), increases dramatically, resulting in overfitting problems and leading to poor performance. We propose a novel statistical approach for genetic risk score inference (GeRSI), based on mixed-effects models. Variants with strong associations are included in the model as fixed effects, just as in standard approaches. Additionally, variants with no significant association are not discarded, but instead are treated as random effects, thus circumventing the need to estimate their respective effects and considerably reducing the number of parameters in the model. By doing so, GeRSI takes advantage of all variants for better prediction, without losing accuracy due to estimating a huge number of parameters. We demonstrate the superiority of GeRSI to state-of-the-art methods in extensive simulation. When applying GeRSI to seven phenotypes from the WTCCC study, we confirm that the use of random effects is most beneficial for diseases that are known to be highly polygenic: hypertension (HT) and bipolar disorder (BD). For HT, there are no significant associations in the WTCCC data. The best existing model yields an AUC of 54%, while GeRSI improves it to 59%. For BD, using GeRSI improves the AUC from 55% to 62%. For individuals ranked at the top 10% of BD risk predictions, using GeRSI substantially increases the BD relative risk from 1.4 to 2.5.

1789M

Allele-specific DNase I hypersensitive sites exhibit H3K27ac enrichment in GM12878. *J.M. Peralta[1], M. Almeida[1], L.J. Abraham[2], E. Moses[2], J. Blangero[1].* 1) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) Centre for Genetic Origins of Health and Disease, University of Western Australia, WA.

Open chromatin accessibility is often assessed by the presence of DNase I hypersensitivity sites (DHSs). It correlates with cis-located elements that regulate the expression of a gene, like promoters and transcription binding sites. The acetylation of Lysine 27 of histone H3 (H3K27ac) is a modification strongly correlated with the active state of enhancers to promote transcription and gene expression. Non-coding variants that affect the binding of proteins to cis-regulatory elements can change the local sensitivity to DNase I digestion, leading to allele-specific DHSs that can serve as an indication of variants with a potentially functional effect on the regulation of neighboring genes. These allele-specific DHSs can be detected, in heterozygotes, as a significant deviation from the expectation of equal allele depth coverage at SNP sites within DHSs. We refer to the significance of this effect, in the -log10[p] scale and derived from a likelyhood ratio based test, as the functional potential (FP) of a SNP. Using mapped short sequencing reads (reads) from DNase-seq of five pooled GM12878 lymphoblastoid cell line replicates (R1-R5) for NA12878, released by the ENCODE Project, we examined the allele-specific chromatin accessibility at 2423308 known heterozygous SNPs sites, based on Illumina's Platinum WGS of NA12878. FP estimates were obtained from independent tests for 48237(1.99%) SNPs using the read depths from each of the two SNP alleles from GM12878's pooled DNase-seq reads. A stringent Bonferroni threshold (-log10[α]=6) was used to classify our FP estimates in two groups, FP$\geq$6 (n=161, 0.33%) and FP<6 (n=48076, 99.67%). Then we determined the FP overlap (n=7971) with ENCODE's GM12878 H3K27ac signal peak (n=56069) and found a highly significant enrichment (p=2.15$\times$10^-47) of FP$\geq$6 (n=108) within H3K27ac signals. Relaxation of the cutoff threshold for the classification of FP estimates by 2 -log10 units lead to an even better enrichment signal (p=6.76$\times$10^-50). These results support a role for allelic-specific DHSs in the regulation of gene expression and suggest that the FP of SNPs can be used to quicky classify potentially functional non-coding variants.

1790T

Genetic modifiers in TGFβ pathway affect disease severity in Duchenne Muscular Dystrophy. *J. Punetha[1,2], H. Gordish-Dressman[1,2], L. Bello[2], A. Kesari[2], M. Giri[2], E.P. Hoffman[1,2], Cooperative International Neuromuscular Research Group.* 1) Department of Integrative Systems Biology, The George Washington University School of Medicine, Washington DC, USA; 2) Center for Genetic Medicine Research, Children's National Medical Center, Washington DC, USA.

Duchenne muscular dystrophy (DMD [MIM 310200]), the most common monogenetic disease in boys, shows marked heterogeneity in disease onset and progression. This inter-patient phenotypic variability observed in DMD can be attributed to the presence of genetic modifiers. Genetic modifiers of DMD (*SPP1* rs28357094, *LTBP4* rs10880) have been associated with loss of ambulation and lower grip strength in patients. Both known DMD genetic modifiers lie within the transforming growth factor-beta (TGF-β) signaling pathway, which is known to be dysregulated in DMD patients, thus increasing its significance in disease pathophysiology. We used a TGF-β pathway driven approach to help investigate potential genetic modifiers and their interactions in a large natural history cohort of DMD patients (CINRG) followed longitudinally for 5-8 years. Key single nucleotide polymorphisms (SNP's) in TGF-β pathway (*SPP1*, *LTBP4*, *TGFBR2*, and *IBSP*) genes were genotyped and tested for associations with outcome measures. Outcome measures chosen were those that could predict disease progression before loss of ambulation occurred i.e. time to run/walk, grip strength. In accordance with our previous studies, dominant *SPP1* rs28357094 genotype showed significant association with grip strength and appeared to be dependent on steroid use. We identified a novel genetic modifier in DMD; patients with the recessive *IBSP* rs2616262 TT genotype showed a significant association (p=0.01, n=150) with disease progression in the time to run/walk (TTRW) test. *IBSP* genotype also showed significant association with grip strength in the non-ambulatory patients (p=0.04, n=111). Preliminary data for SNP interaction studies between *SPP1* and *IBSP* polymorphisms show evidence for interaction of the two loci. Providing sufficient statistical confidence of genetic modifier effects in different patient cohorts has been problematic due to differences in steroid treatment regimens, ethnicity, outcome measures, age of study population, and study sites tested. In our cohort, there were differences in functional outcomes between different ethnic groups as well as different standards of care provided. Validation of SNP association studies requires selection of a homogenous population to reduce confounding factors and hidden stratification. Further analyses of multi-locus modifier approaches will help identify the key players in transition from successful to unsuccessful remodeling in muscular dystrophy.

1791S

Linkage disequilibrium clustering can improve power of weighted-sum-type multi-marker tests for genetic association analysis. *Y. Yoo[1,2], S. Kim[1], L. Sun[3,4], S. Bull[3,5].* 1) Mathematics Education, Seoul National University, Seoul, South Korea; 2) Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, South Korea; 3) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 4) Department of Statistical Science, University of Toronto, Toronto, Canada; 5) Prosserman Centre for Health Research, The Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Canada.

Background For gene-based analysis of genome-wide association study, a multi-SNP regression analysis using common and low frequency SNPs can be performed for a gene unit and global hypothesis tests can be constructed. Methods constructed as quadratic forms are especially effective in the presence of both deleterious and protective causal SNPs. On the other hand, weighted-sum type tests benefit from smaller df when causal SNPs are all deleterious. The power of both tests suffers in the presence of many neutral SNPs relative to causal SNPs. In this study, we investigate how clustering of correlated SNPs can improve power of a hybrid quadratic-linear combination test using weights incorporating the correlations between SNPs. **Methods** The Wald statistic is defined as a quadratic sum of beta coefficients with covariance matrix weights. The LC statistic is linear sum using weights incorporating variances and covariances between SNPs. We propose the MLC statistic defined as a quadratic sum over per-cluster linear combinations, and apply a spectral clustering algorithm to cluster closely correlated SNPs. The coding for base and risk alleles within each cluster is assigned using a pairwise LD coding method so that the number of pairs of SNPs with positive value of linkage disequilibrium measure r is as many as possible . We compared the power of Wald, LC, MLC tests with numerical calculation of asymptotic power over different disease model scenarios. For comparison over realistic correlation patterns, we selected 1000 genes, and specified genotype-phenotype models using the structure of SNPs from HapMap data for these genes. Three types of disease model scenarios were examined: A: one causal SNP, B: two causal SNPs in different clusters both with deleterious effects, C: two causal SNPs in different clusters with one deleterious effect and one protective effect. **Results and Conclusions** For a one causal SNP scenario with the number of neutral SNPs between 1~10, the power of the LC test exceeds that of the Wald test when the correlations between SNPs reach a certain positive value between 0.1~0.6. The power of the MLC test is usually higher than the other tests under all three disease model scenarios with one or two causal SNPs. These results suggest that clustering of highly correlated SNPs can improve the power of linear combination tests using weights incorporating the correlations with other SNPs.

1792M

A unified analysis approach for X-chromosome that accounts for random, skewed and escaping of X-chromosome inactivation. *J. Wang, R. Yu, S. Shete.* Dept Biostatistics, UT MD Anderson Cancer Ctr, Houston, TX.

X-chromosome inactivation (XCI) on female X-chromosome loci states that in females during early embryonic development, 1 of the 2 copies of the X-chromosome present in each cell is randomly selected to be inactivated to achieve dosage compensation of X-linked genes in males and females. Due to the random XCI, about half of the cells have one allele inactive while the other half of the cells have the other allele inactive. As a result, the homozygous genotypes in females have the similar effects as the hemizygous allele types in males. In general, the XCI process is random; however, recent studies have suggested that skewed XCI is a biological plausibility in which more than 75% of cells have the same allele active. Another complexity in analyzing X-chromosome data is the escaping of XCI outside the pseudo-autosomal regions on the female X-chromosome, that is, both alleles are active in all cells. Current statistical tests for X-chromosome genetic association studies either account for random XCI only (e.g., Clayton's approach) or escaping of XCI only (e.g., PLINK software). Because for females, the true XCI process is unknown and differs across different regions on X-chromosome, we proposed a unified approach for analyzing X-chromosomal genetic data, which will account for all such biological possibilities: random XCI, skewed XCI and escaping of XCI. Simulation studies were conducted to compare the performance of the proposed approach with Clayton's and PLINK approaches. For the scenarios where the XCI is skewed, the proposed approach has higher powers to identify the associated genetic variants than the Clayton's and PLINK approaches. We also analyzed the head and neck cancer X-chromosome genetic data using the proposed and existing approaches. The proposed approach identified an SNP rs12388803 associated with head and neck cancer risk: meta-analysis-based p values = 2.04×10^{-6}, 2.83×10^{-6} and 2.83×10^{-6} for Fisher's, fixed effect and random effect models, respectively.

1793T

Efficient Detection of Allelic Imbalance from SNP microarrays. *C. Hahn[1,4], S. Vattathil[2,4], Z. Weber[3], L. Huang[4], R. Xia[5], G.E. Davies[3], E. Ehli[3], P. Scheet[1,2,4].* 1) Biomathematics and Biostatics Program, The University of Texas Graduate School of Biomedical Sciences, Houston, Texas 77030, USA; 2) Human & Molecular Genetics Program, The University of Texas Graduate School of Biomedical Sciences, Houston, Texas 77030, USA; 3) Avera Institute for Human Genetics, Sioux Falls, South Dakota, USA; 4) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA; 5) Biostatistics Program, The University of Texas School of Public Health, Houston, Texas 77030, USA.

Studies of tumor genomes in settings of high levels of normal DNA contamination or clonal heterogeneity require computational methods sensitive enough to detect subtle forms of allelic imbalance (AI) resulting when a small portion of the cells contain copy-number loss of heterozygosity, deletions or duplications. Detection of these forms can be enhanced by leveraging haplotype information to identify regions where there has been relative loss or gain of one of an individual's inherited chromosomes. Such regions may be representative of subclones or may indicate clonal mosaicism in non-tumor tissues. Due to lower levels of perturbation from the tumor genome in the BAFs of low-frequency mixed samples, the called genotypes from such samples represent the germline genotypes and we can apply these calls towards a single sample analysis. We present a framework to model the observed microarray data by combining parametric distributions for the magnitude of B allele frequencies (BAFs) with haplotype estimates informing the correlation expected in the BAFs in regions of AI. Specifically, we apply a mixture of bivariate normal distributions at consecutive heterozygous sites. Our framework allows for a probabilistic characterization of each bi-marker interval across a tumor genome, which can then be used to formulate emission probabilities for use in a hidden Markov model. Our method thus retains the speed of simple models that make use of haplotype information while capturing the important feature of correlated BAFs. We compared our method to other state-of-the-art packages (e.g. hapLOH, BAFsegmentation) to identify aberrations and improve detection performance in mixtures with aberrations in less than < 5% of the sample. We focus our comparisons on a novel data set, constructed from series of mixtures of DNA from cell lines derived from tumor and paired blood of a lung (T: HTB-172, N: CRL-5948) and breast (T: CRL-2343, N: CRL-2363) cancer. Methods such as hapLOH that capture haplotype information alone excel over those that simply model the magnitude of BAFs. However, the more subtle inflections in signal at lower tumor proportions (~5%) are difficult to recognize using hapLOH but in several instances are detectable by our method, as we incorporate a more use of the available data. Our flexible framework can be utilized in a variety of contexts, including testing specific regions of a priori interest for AI.

1794S

Imputing phenotypes for genome wide association studies. *F. Hormozdiari[1], E. Kang[1], C. Vulpe[4], S. McLachlan[5], A. Lusis[2,3], E. Eskin[1,2].* 1) Computer Science, UCLA, Los Angeles, CA; 2) Department of Human Genetics, University of California at Los Angeles; 3) Department of Medicine, University of California, Los Angeles, Los Angeles; 4) Department of Nutritional Science and Toxicology, University of California Berkeley; 5) Centre for Population Health Sciences, University of Edinburgh.

Genome wide association studies (GWAS) are extremely successful to detect the associated variants for various phenotypes. However, the power to detect these associated variants depends on the strength of the associated signals and the number of individuals whom their phenotypic values are collected. Thus, for phenotypes that are difficult to collect, we are likely to lack the sufficient number of individuals to have the desired power. In this work, we utilize a set of multiple phenotypes, which are easy to collect to predict (impute) an additional phenotype (target phenotype), which is difficult to collect to achieve a desired power for the association study of the target phenotype. We leverage the correlation structure between phenotypes to impute the phenotypic values of the target phenotype. The correlation structure can be estimated from the training data, where we collect all phenotypes including target phenotype. Then, we impute the target phenotype of the individuals whose target phenotype is not collected. We derive an analytical formula to compute power of the association study with imputed phenotype and show the correctness through rigorous simulations. We show that we can increase the power of detecting the significant variants by 9-25% through imputing the missing individuals in the Northern Finland Birth Cohort data. In addition, we show the utility of our method through applying our method to Hybrid Mouse Diversity Panel (HMDP) data. We use two versions of HMDP that were collected in two different times. We use the older version of the HMDP for training and then we impute the phenotypes in the newer version of HMDP. We manage to show that the correlation between the imputed phenotype and the real phenotypic values is high for most phenotypes.

1795M

Quality and accuracy assessment for NGS data analysis and interpretation. *J. Li[1], M. Mohiyuddin[1], A. Kiani[2], M. Carneiro[3], P. Jiang[3,4], S. Tabrizi[3,4], S. Schaffner[3,4], I. Shlyakhter[3,4], N. Asadi[1,2], P. Sabeti[3,4], W.H. Wong[5,6], H.Y.K. Lam[1].* 1) Department of Bioinformatics, Bina Technologies, Redwood City, CA; 2) Department of Engineering, Bina Technologies, Redwood City, CA; 3) Program in Medical and Population Genetics, the Broad Institute, Cambridge, MA; 4) Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA; 5) Department of Statistics, Stanford University, Stanford, CA; 6) Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA.

Next-generation sequencing (NGS) has become the mainstream for characterizing genetic variations for both targeted genomic regions and whole genome. The quickly decreasing cost at increasing coverage is enabling numerous possible applications of large-scale studies, ranging from discovering genetic disease-causing mutations and pathways to identifying cancer somatic alterations.

To conduct meaningful research and generate correct interpretation based on the identified variations, quality assessment (QA) of the data becomes extremely important, especially in clinical settings. Accumulated cohort study results, in turn, provide learning resources for differentiating true variations from errors. Here we present a comprehensive, user-friendly quality control pipeline for NGS secondary and tertiary analysis. This pipeline, synergized with the secondary analysis framework, provides across-the-board quality measurements addressing potential quality issues residing in NGS reads, such as sequence quality, length, and adaptor detection; in alignment, such as coverage statistics, mapping distribution, and insert estimation; and in variant calling, such as known variant ratio, heterozygous/homozygous ratio, and transition/transversion ratio.

For tertiary analysis, to ensure the relevant biological interpretations are not caused by low quality or conflicting data, our pipeline provides a novel scheme for quality assurance by learning from accumulated homogeneous samples. A wide range of quality control is performed on the identified variations from the new samples, such as calculating missing data rate and heterozygosity rate, detecting duplicated or related individuals, checking for ethnicity, gender and blood type, identifying mitochondrial haplotype, and computing Hardy-Weinberg log p-value.

Visualization methods such as clustering heatmap, VQSR LOD score plots, and a novel method we have developed - QCRadar, are used in QC report for users to easily identify potential problems at a glance. We have applied our QC pipeline to a large number of samples, including recently to 20 Lassa fever patient genomes of a cohort in which we successfully identified the optimal VQSR LOD score to maximize the sensitivity of the high-confidence variant call set.

1796T

A comprehensive survey of genetic variation in 20,769 subjects from the Harvard Cohorts. *S. Lindstrom[1], S. Loomis[2,3], C. Chen[1], H. Huang[1], J. Huang[1], A. Chan[4], H. Choi[5], G. Curhan[6], I. De Vivo[1, 6], C. Fuchs[2,6,7], F. Hu[1], C. Kabrhel[2,4,6], L. Pasquale[2,3,6], E. Rimm[1], R. Tamimi[1,6], S. Tworoger[1,6], D. Hunter[1,6], P. Kraft[1].* 1) Harvard School of Public Health, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Massachusetts Eye and Ear Infirmary, Boston, MA; 4) Massachusetts General Hospital, Boston, MA; 5) Boston University School of Medicine, Boston , MA; 6) Brigham and Women's Hospital, Boston, MA; 7) Dana-Farber Cancer Institute, Boston, MA.

The Nurses' Health Study (NHS), Nurses' Health Study II (NHSII), Health Professionals Follow Up Study (HPFS) and the Physicians Health Study (PHS) have collected detailed longitudinal data on a wide range of exposures and traits for more than 310,000 study participants over the last 35 years. Across the four cohorts, more than 90,000 study participants donated a blood sample between 1982 and 1999 allowing for prospective analysis of biomarkers and genetic variation. Over the last eight years, 20,769 subjects have been genotyped as part of genome-wide association studies (GWAS) of 12 primary traits within the cohorts. However, these studies have utilized 6 different GWAS platforms, making it difficult to conduct analyses of secondary phenotypes or borrow controls from other studies. To allow for secondary analysis in this data, we have created new datasets merged by platform and imputed them to a common reference panel, the 1,000 Genomes Phase I release. We describe the methodology behind the data merging and imputation and present imputation quality statistics and association results from two GWAS based on the merged and imputed data: one of body mass index (BMI) and one of venous thromboembolism (VTE). Using a significance level of p=0.05, 19 out of 32 known BMI SNPs showed association with BMI with the strongest association for the FTO SNP rs55872725 (β=0.44, $p=1.31\times10^{-21}$). For VTE, 3 out 4 known VTE SNPs showed association with VTE with the rs144737447 SNP (OR=2.15, 95% CI: 1.78-2.61, $p=2.96\times10^{-15}$), located downstream of F5 showing the strongest association. This pooled resource can be used to boost power and study rare phenotypes and genotypes.

1797S

Heteroscedastic Extreme Sampling Strategy in Target Sequencing Studies. *W. Ouyang, HW. Deng, H. Qin.* Biostatistics & Bioinformatics, Tulane University, New Orleans, LA.

Judiciously selecting carriers of target genetic variants is crucial for the success of a target deep sequencing study. Extreme phenotype sampling (EPS) has been widely employed to identify complex trait genes. EPS can enrich rare genetic causal variants compared to naïve random sampling (NRS). An intuitive alternative is extreme residual sampling (ERS) - selecting individuals with extreme phenotypic residuals after adjusting for some covariates. However, it remains unclear when local genetic covariates within target genomic regions should be adjusted for to enrich target genetic variants at sampling stage. Heteroscedasticity is ubiquitous in real data but largely ignored in gene mapping for complex traits, especially that of admixed subjects. People pay most of attention on the phenotypic mean instead of phenotypic variance. In this article, we first formulated the powers of ERS with adjusting for the local genetic covariate and EPS to include carriers of the target variant under heteroscedasticity model. In terms of inclusion power, we specified diverse superiorities of EPS and ERS. Second, given budget restriction,, Our findings provide genomic researchers with novel sequencing extremes strategy combining EPS and ERS to determine the truncation threshold and design cost-effective target sequencing studies. Last, we apply our novel method in real data to identify genetic causal variants for complex diseases.

1798M

Use of exome sequencing data for the analysis of population structures, inbreeding, and familial linkage. *V. Pedergnana[1,2], A. Belkadi[1], A. Abhyankar[3], Q. Vincent[1], Y. Itan[4], B. Boisson[4], J.L. Casanova[1,4,5], L. Abel[1,4].* 1) Laboratory of Human Genetics of Infectious Disease, INSERM U1163, University Paris Descartes. Imagine Institute, Paris, France; 2) Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 3) New York Genome Center, New York, NY, USA; 4) St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, the Rockefeller University, New York, NY, USA; 5) Howard Hughes Medical Institute, New York, NY, USA.

Numerous methods have been proposed to analyze whole exome sequencing (WES) data in order to discover potential causal variants in Mendelian disorders and in more complex traits. These methods could benefit from additional information such as linkage studies in the study of Mendelian diseases. Population stratification could also be an issue in the analysis of WES data when focusing on complex traits. Both linkage and population structure analyses are classically conducted through genome-wide (GW) SNP arrays. Here, we compared the information yielded by WES data to that provided by SNP array data in terms of analyses usually performed by SNP array data such as principal component analyses (PCA), linkage studies, and homozygosity rate estimation. We analyzed 123 subjects originating from six world regions, including North Africa and Middle East which are regions poorly covered by public database and presenting a high consanguinity rate. A number of quality control (QC) filters were tested and applied to the WES data. Compared to results obtained with SNP array data, we found that WES data provided accurate prediction of population substructure and led to highly reliable estimation of homozygosity rates (correlation > 0.94 with the estimations provided by SNP array). Linkage analyses showed that the linkage information provided by WES data was on average 53% lower than the one provided by SNP array at the GW level, but 58% higher in the coding regions. In conclusion, WES data could be used after appropriate QC filters to perform PCA analysis and adjust for population substructure, to estimate homozygosity rates, and to perform linkage analyses at least in coding regions.

1799T

GENESIS: a French national resource to study the missing heritability of breast cancer. *N. ANDRIEU[1,2,3], M-G. DONDON[1,2,3], S. EON-MARCHAIS[1,2,3], F. DAMIOLA[4], M. MARCOU[1,2,3], L. BARJHOUX[4], M. GAUTHIER-VILLARS[5], B. BUECHER[5], A. CHOMPRET[6], O. CARON[6], P. VENNIN[7], I. COUPIER[8], C. NOGUES[9], V. BONADONA[10], P. GESTA[11], C. DUGAST[12], C. MAUGARD[13], J-P. FRICKER[14], L. FAIVRE[15], E. LUPORSI[16], M. LONGY[17], P. BERTHET[18], C. DELNATTE[19], A. CHEVRIER[20], C. COLAS[21], Y-J. BIGNON[22], CGEI. Platform[1,2,3], S. MAZOYER[4], D. STOPPA-LYONNET[5,23,24], O. SINILNIKOVA[4,25], GENESIS investigators.* 1) Institut Curie, rue d'Ulm 26, Paris cedex 05, 75248, France; 2) Inserm U900, rue d'Ulm 26, Paris cedex 05, 75248, France; 3) Mines ParisTech, rue St Honoré 35, Fontainebleau Cedex, 77305, France; 4) Genetics of Breast Cancer group, Cancer Research Center of Lyon, INSERM U1052, CNRS UMR5286, Université de Lyon, Centre Léon Bérard, rue Laennec 28, Lyon cedex 08, 69373, France; 5) Genetic oncology service, Institut Curie, rue d'Ulm 26, Paris cedex 05, 75248, France; 6) Gustave Roussy, Department of Medical Oncology, Villejuif, F-94805 France; 7) Breast Department, Oscar-Lambret Regional Cancer Center, Rue Frédéric Combemale 3, Lille, 59000, France; 8) Unit medical genetics and oncology, Hôpital Arnaud de Villeneuve CHU Montpellier, avenue du Doyen Gaston Giraud 371, Montpellier Cedex 5, 34295, France; 9) Department of Public Health, Institut Curie Hôpital René Huguenin, rue Dailly 35, Saint Cloud, 92210, France; 10) Unit of genetic epidemiology and prevention, Centre Léon Bérard, rue Laennec 28, Lyon cedex 08, 69373, France; 11) Oncology center for the regional cancer genetics consultation Poitou-Charentes, CH Georges Renon, avenue Charles de Gaulle 40, Niort Cedex, 79021, France; 12) Unit Genetics, Centre Eugène Marquis, avenue Bataille Flandres Dunkerque, Rennes Cedex, 35042, France; 13) Laboratoire de Diagnostic Génétique, Nouvel Hôpital Civil, place de l'hôpital 1, BP 426, Strasbourg cedex, 67091, France; 14) Unit of oncology, Centre Paul Strauss, rue de la porte de l'Hôpital 3, Strasbourg, 67000, France; 15) Genetic center, CHU Hôpital d'enfants, boulevard Maréchal de Lattre de Tassigny 10, Dijon Cedex, 21034, France; 16) Unit of oncologenetic, ICL Alexis Vautrin, Avenue de Bourgogne 6, Vandœuvre-lès-Nancy, 54511, France; 17) Laboratory of molecular genetics, Institut Bergonié, Cours Argonne 229, Bordeaux, 33000, France; 18) Unit of gynecological pathology, Centre François Baclesse, Avenue Général Harris 3, Caen, 14000, France; 19) Centre René Gauducheau, Boulevard Jacques Monod, St Herblain Cedex, Nantes, 44805, France; 20) Centre Henri Becquerel, rue d'Amiens, Rouen, 76038, France; 21) Department of Genetic, Groupe hospitalier Pitié Salpétrière, APHP, Boulevard de l'Hôpital 83, Paris Cedex 13, 75651, France; 22) Department of Oncogenetic, Centre Jean Perrin, University of Clermont-Ferrand, rue Montalembert 58, BP 392, Clermont-Ferrand Cedex 01, 63011, France; 23) Unit Genetics, Inserm U830, rue d'Ulm 26, Paris cedex 05, 75248, France; 24) Université Paris-Descartes, rue de l'Ecole de Médecine 15, Paris, 75006, France; 25) Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon - Centre Léon Bérard, rue Laennec 28, Lyon cedex 08, 69373, France.

Less than 20% of familial breast cancer (BC) cases who undergo genetic testing for *BRCA1* and *BRCA2* carry a pathogenic mutation in one of the two genes. The GENESIS (GENE SISter) study was designed to identify new BC susceptibility genes in women attending the cancer genetics clinics and with no *BRCA1/2* mutation. The study is based on enrichment of the study population in genetic factors through case selection based on familial criteria, but also on the consideration of environmental factors and endophenotypes to assess potential genetic heterogeneity. For instance the study sample may be stratified on characteristics such as mammary density or tumour characteristics. The study involves the French national network of genetic counsellors, and the enrolment of patients, the collection of their clinical, epidemiological, familial and biological data are centralized by a coordinating centre. The purpose of this communication is the description of the study design and of the available resources. Index cases (and their affected sisters) were eligible when diagnosed with infiltrating mammary or ductal adenocarcinoma, having no *BRCA1/2* mutation, and having a sister with BC. In addition, unrelated controls and unaffected sisters were recruited. Inclusion of participants started in February 2007 and ended in December 2012. A total of 1,711 cases and 2,829 sisters and controls were included. 98% of participants completed the epidemiological questionnaire and provided a blood sample, 68% of them were able to provide mammograms. Index cases were on average 59 years old at inclusion, were born in 1950, and were 49.7 years at BC diagnosis. The mean age at diagnosis of affected sisters was slightly higher (51.4 years). One of the initial aims of GENESIS was to recruit affected sibpairs. A total of 785 affected sisters accepted to participate. The representativeness of the control group has been verified. The size of the study (over 4500 participants), the availability of biological specimen and the clinical data collection together with the detailed and complete epidemiological questionnaire makes it a unique national resource to investigate the missing heritability of BC by taking into account environmental and lifestyle factors and stratifying data on endophenotypes to decrease heterogeneity. Molecular analyses including high-throughput genotyping and massive-parallel sequencing projects have started.

1800S

SURVIVAL MONITORING DURING THE FIRST YEAR OF LIFE OF INFANTS WITH BIRTH DEFECTS IN A HIGH COMPLEXITY HOSPITAL OF THE CITY OF CALI, COLOMBIA, 2012-2013. *F. Ruiz[1], Y. Ariza[1], H. Pachajoa[1, 2].* 1) Centro de Investigaciones en Anomalías Congénitas y Enfermedades Raras (CIACER), Universidad Icesi, Cali, Colombia; 2) Fundación Clínica Valle del Lili, Cali, Colombia.

Introduction: Congenital anomalies are structural defects that can be fatal or produce physical and / or mental disabilities. In the literature numerous studies about the prevalence and etiology of birth defects are reported, that include mortality analysis that focus on the contribution of these conditions to the perinatal and infant death rates, and in few opportunities focuses on survival of those affected. This study aims to describe survival during the first year after birth in a group of children with birth defects. Methodology: A retrolective cohort study with internal comparison group was conducted. Primary and secondary sources of information were used. For patient identification it was resorted to the database of the birth defects surveillance registry that takes place in a clinical institution of higher complexity in the city of Cali. Identifying individuals born with birth defects in the period between t October 1, 2012 and May 31, 2013. The obtained data were analyzed by the Kaplan - Meier to determine the probability of survival. Results: This study showed that the probability of survival during gestation ranged from 98.4 % at week 20 to 79.0% during the first 24 hours after birth. In the subgroup of patients who survived the first day after birth, only one death was recorded, so by the end of week 96 of follow-up, the group had a survival probability of 91.7% average. Conclusion: It is of importance to extend the network of surveillance and monitoring of birth to other health institutions in the city of Cali, for the purpose of establishing a surveillance registry of birth defects that is representative of the city, and that allows an adequate sampling that provides an accurate overview of the current state of epidemiology, health care and survival of patients with birth defects in the city of Cali.

1801M

The simulation of the confounding effect on cryptic relatedness for environmental risks in cohort studies. *K. Shibata, A. Tatsukawa, S. Ogishima, K. Shimokawa, S. Nagaie, N. Nakamura, J. Nakaya.* Tohoku University.

The impact of cryptic relatedness (CR) on genomic association studies is well studied and known to inflate false-positive rates as reported. In contrast, conventional epidemiological studies for environmental risks, the confounding effect of CR is still uninvestigated. In our previous study, we showed that heritability for liability might reflect on the estimation of regression coefficients between systolic blood pressure and environmental risk factors using real data. On the other hand, it is difficult to examine the confounding effect on CR in more detail using real data under the changes to the setting. In our simulation study, we detail an investigation on the confounding effect of unadjusted CR among a rural cohort in the quantitative phenotype mean effect using simulation data under a different heritability setting. Using SIBSIM, we generated combined independent individual samples and family data contained with two generations, and one-way analysis of variance to determine whether the confounding effect of CR was affected by environmental variance as environmental risk factors in a different setting: sample size (<2000), the ratio of CR to the population, and heritability for liability (10%, 20%, 30%, 40%) in a simulation cohort population. We present a simple model to illustrate the effect of CR on the estimation of phenotype mean effect. We are now extensively analyzing this issue in different settings incorporated with the prevalence in the subpopulation without CR, the prevalence in CR, and the population prevalence using family data contained with two , three or four generations.

1802T

Assessing the potential impact of low participation in DNA buccal swab collection on the validity of effect estimates. *M.M. Jenkins[1], J. Reefhuis[1], H. Razzaghi[1,2], A. Herring[3], M.L. Gallagher[1], M.A. Honein[1], The National Birth Defects Prevention Study.* 1) Centers for Disease Control and Prevention, Atlanta, GA; 2) Oak Ridge Institute for Science and Education, Oak Ridge, TN; 3) The University of North Carolina, Chapel Hill, NC.

Low participation and potential selection bias for specimen collection in gene-environment interaction studies threaten validity. To better understand if effect estimates obtained from participants who submitted DNA buccal swab samples will be impacted by selection bias, we examined the associations between self-reported pregestational diabetes mellitus (PGDM) and four birth defects among families who did and did not submit samples, stratified by maternal race-ethnicity (non-Hispanic white, non-Hispanic black, and Hispanic). Following a telephone interview in a multi-site, population-based, case-control study (National Birth Defects Prevention Study), mothers with estimated dates of delivery between 1997 and 2009 were asked to collect and submit buccal swab samples. Samples were submitted for 52% of interviewed mothers, their infants, or both. For each birth defect, maternal race-ethnicity-stratified ratio of odds ratios (ROR) and corresponding 95% confidence intervals (CI) were calculated using logistic regression, comparing the ORs for the PGDM/birth defect association of families who did not submit samples to the ORs for the PGDM/birth defect association of families who did submit samples, while adjusting for maternal age at delivery. Analyses included control families with live born infants who had no major structural birth defects (N=8775), families with pregnancies affected by limb reduction defects (N=958), neural tube defects (N=1615), orofacial clefts (N=3509), or septal heart defects (N=3762). 1% of control mothers reported PGDM versus up to 7% of case mothers. Submitting samples was not associated with reporting PGDM among non-Hispanic black and Hispanic control families but non-Hispanic white control families who submitted samples reported PGDM less often. Statistically significant differences in the effect estimates for the PGDM/birth defect association between families who did or did not submit samples were observed among non-Hispanic black families whose child had an orofacial cleft (ROR=7.58, 95% CI 1.02-56.49, P=0.048) and non-Hispanic white families whose child had a septal heart defect (ROR=0.32, 95% CI 0.13-0.76, P=0.010). We found no evidence to support different effect estimates of PGDM and birth defects in the 10 other analyses among these three race-ethnicities. Further analyses will be completed with other exposures and birth defects.

1803S

The Million Veteran Program (MVP): A National Resource for Genomic and Epidemiological Research. *S. Muralidhar[1], J.M. Gaziano[2], J.P. Concato[3], S. Whitbourne[2], J. Deen[2], J. Moser[1], K. Schaa[1], S. Pyarajan[2], R. Przygodzki[1], T.J. O'Leary[1].* 1) Office of Research and Development , Department of Veterans Affairs, Washington, DC; 2) VA Boston Healthcare System, 150 S. Huntington Ave, Boston, MA-02130,; 3) VA Connecticut Healthcare System, 950 Campbell Ave. Bldg 35A (151B), West Haven, CT 06516.

Genomic discoveries are critically dependent on having a well-defined cohort with a sufficiently large number of subjects to ensure statistically and clinically meaningful associations. The Department of Veterans Affairs Million Veteran Program (MVP) was launched in 2011 to create a cohort of Veterans, one-million strong to leverage the excellent electronic medical record resource and combine it with genomic data and self-reported survey data to allow for both original and replication studies on diseases and conditions prevalent in Veterans. With mail-based recruitment, face-to-face enrollment at approximately 50 VA medical centers and broad consent for research, MVP has enrolled over 275,000 Veterans to date. Future models of online enrollment are currently under development. A centralized resource for computation and data analysis called the Genomic Information System for Integrative Science (GenISIS), is under development to facilitate data analysis in a secure environment. Approximately 200,000 samples are currently being genotyped on a customized Affymetrix Axiom chip and smaller subsets of samples are undergoing whole exome and genome sequencing, all through contracted services. The initial focus of analysis will be on mental health conditions such as PTSD, schizophrenia and bipolar disorder, other conditions predominant in the MVP cohort, as well as the exceptionally aged. Approved researchers will access data within the GenISIS environment. Results of the recruitment strategies, overview of demographics and diseases prevalent in the cohort, and updates on the informatics infrastructure as well as genetic analysis will be presented.

1804M

Genotype imputation performance in multiple ethnicities via comparing with whole-genome sequencing data. *H. Zhan, N. Bing, S. Paciga, J. Wendland, S. John, B. Zhang.* Worldwide Research & Development, Pfizer Inc., Cambridge, MA., USA.

It is well known that genotype imputation can boost the power of genome-wide association studies by imputing genotypes not observed in the sample. It facilitates the identification of susceptibility loci for complex diseases. Over the past decade, many imputation algorithms and software have been developed to impute genotype for samples from various populations. In this study, we investigated the genotype imputation accuracy of IMPUTE2 and MaCH-Admix using 148 subjects from multi-ethnic populations (African, Asian, European and Hispanic). The subjects were genotyped in Illumina HumanOmniExpress plus exome array and at the same time whole genome sequenced at high coverage (30X). Instead of masking the genotypes in some specific loci, we directly assessed imputation accuracy by comparing the imputed genotype with the real genotype from sequencing data. In order to address the potential imputation bias, imputation accuracy is based on the concordance for heterogeneous genotype. We found that IMPUTE2 is more computationally efficient than MaCH-Admix. For common variants (MAF>0.01), IMPUTE2 and MaCH-Admix have similar imputation accuracy. However, MaCH-Admix has relatively higher imputation accuracy than IMPUTE2 for the rare variant (MAF<0.01). Both IMPUTE2 and MaCH-Admix show better imputation accuracy in common variants than rare variants. The imputation results also provide valuable information for post-imputation quality control. Optimal threshold for selecting imputed SNPs can be determined for one fixed imputation accuracy. In addition, we also imputed HLA classical alleles and compared with HLA loci called from sequencing data using HIBAG (an R package). HIBAG imputes seven HLA loci with increased accuracy across ethnicities correlated with increased posterior probability of imputed HLA type. In general, we demonstrated that genotype and HLA classical alleles can be accurately imputed for multiple ethnicities using current imputation methods via comparing sequencing data. IMPUTE2 and MaCH-Admix have similar accuracy performance, and IMPUTE2 has better computation efficiency.

1805T

A Strategy for Testing Zero Variance Components with Application to QTL Association Mapping in Admixture Population. *J. Zhou[1], H. Zhou[2].* 1) University of Arizona, Tucson, AZ; 2) North Carolina State University, Raleigh, AZ.

It is widely known that population substructure can distort the results of genetic association analysis, especially when rare variants are present such as in next-generation sequencing (NGS) studies. Linear mixed effects model (LMM) were proposed before for adjusting population substructure while testing typically common variant as a fixed effect. To improve the power of testing multiple rare variants while adjusting for population substructure, one can jointly test them as a random effect in conjunction with the random effect from population substructure. This brings challenge to the current testing methods that are computationally intensive and hard to scale to NGS. We propose a powerful and scalable strategy for testing zero variance component in presence of multiple variance components in LMM. Our approach combines the recent development of exact (restrictive) likelihood ratio test with single variance component and a strategy for reducing LMM with multiple variance components to that with a single one. The performance of the strategy is demonstrated on both synthetic and real data.

1806S

Estimating base-calling error rates in next-generation sequencing data using overlapping read pairs. *Y.Y. Lo[1], S. Zöllner[1,2], The BRIDGES Consortium.* 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Psychiatry, University of Michigan, Ann Arbor, MI.

Genotype calling from short read sequencing data requires accurate estimates of base errors in the reads. Such errors can occur in multiple steps of the sequencing experiment, particularly during the PCR and sequencing steps. Base errors from all steps appear as mismatches to the reference genome on aligned reads, hence the error models typically aggregate errors from all sources. Besides, base information from non-independent sequence reads are often discarded or wrongly incorporated into the base-calling error model intended for independently sequenced reads. Disentangling the sources of base-calling errors would allow more efficient use of all sequencing reads.

Here we aim to distinguish between base-calling errors generated by PCR and by the sequencing machine (machine error), using overlapping read pairs (ORPs) arising from short PCR fragments in paired-end sequencing. ORPs replicate PCR errors in the overlapping region, while each read in the pair is independently sequenced. Therefore, per genomic position, discordant bases from an ORP are the result of at least one machine error but not PCR error, regardless of the true genotype. On the other hand, if an ORP reports the same base, but different from the reference allele, either the underlying base is a true variant, or the ORP has at least one PCR error or two machine errors. Based on these observations, we derive maximum-likelihood estimators for PCR error and machine error using the counts of discordant and concordant bases from ORPs. We calculate these estimates, as functions of read cycle and base quality score, using an expectation-maximization algorithm.

We apply our method to a subset of data from a whole-genome sequencing study (median 8x coverage, 806 individuals) with 40-60% ORPs. From these ORPs, we estimate that base-calling errors generated by the sequencing machine are approximately twice as common as PCR errors. We quantify the contribution of read cycle and base quality score to these base-calling error estimates. We evaluate the mean and variance of these predicted error rates across individuals. Understanding the contribution of both error sources to base-calling allows more efficient use of ORPs, hence better modeling of errors and improved accuracy in variant calling.

1807M

Overcoming Systematic Miscalibration of Linear Mixed Model Test Statistics in Genetic Association Studies by Leveraging Ancestry Representative Principal Components. *M.P. Conomos, T. Thornton.* Biostatistics, University of Washington, Seattle, WA.

A plethora of linear mixed model based procedures have been developed recently to control Type I error rates in genetic association studies of complex traits by simultaneously accounting for both population and pedigree structure among sample individuals. The general approach taken by these methods is to compute a genome-wide genetic similarity matrix to be included as part of the covariance structure of the phenotype when fitting the model. These mixed model based approaches control genomic inflation genome-wide; however, it has recently been shown that they may not adequately control for false positives at SNPs that are unusually differentiated between ancestral populations (Price et al., 2010). In reality, the differentiation in allele frequencies between ancestral populations varies greatly across the genome, and proper calibration of test statistics from these methods should be a concern even at SNPs that are not unusually differentiated. Through the use of simulation studies of samples from structured populations with related individuals, we demonstrate that adjusting for a genetic similarity matrix calculated genome-wide controls Type I error well at SNPs with an average amount of differentiation, but it also leads to a systematic inflation or deflation of test statistics for SNPs that are more or less differentiated. By including ancestry representative principal components as fixed effect covariates in the mean model in addition to the genetic similarity matrix included in the covariance structure, we are able to correct for this issue and obtain well-calibrated test statistics genome-wide. Additionally, we demonstrate that fitting the mixed model with an ancestry adjusted genetic similarity matrix may increase power to detect true associations with traits that are correlated with ancestral background. We apply our testing procedure to a variety of blood phenotypes measured on African American and Hispanic women in the Women's Health Initiative SNP Health Association Resource (WHI-SHARe) study, and we see increased significance over existing methods at SNPs associated with white blood cell count.

1808T

Evaluation of population stratification in a large biobank linked to Electronic Health Records. *M. de Andrade[1], G. Thromp[2], A. Burt[3], D.S. Kim[3,4], S.S. Verma[5], A.M. Lucas[5], S.M. Armasu[1], J.A. Heit[6], M.G. Hayes[7], H. Kuivaniemi[2], M.D. Ritchie[5], G. Garvik[3,4], D.R. Crosslin[3,4].* 1) Div Biomed Statistics & Informatics, Mayo Clinic, Rochester, MN; 2) The Sigfried and Janet Weis Center for Research, Geisinger Health System, Danville, PA; 3) Dept Genome Sciences, Univ of Washington, Seattle, WA; 4) Dept Medicine, Div Med Genet, Univ of Washington, Seattle, WA; 5) Center for Systems Genomics, Dept Biochemistry and Mol Biology, Pen State Univ, University Park, PA; 6) Div Cardiovascular Diseases, Mayo Clinic, Rochester, MN; 7) Div Endo, Metabolism and Mol Medicine, Fienberg School of Medicine, Northwestern Univ, Chicago, IL.

For genomic association studies, combining samples across multiple studies in Networks or “Big Science” is standard practice. Increasing the number of subjects allows for power needed to assess association. Controlling for genomic ancestry is common, but there is a need to standardize the approach when calculating principal components (PCs) across cohorts such as elimination of SNPs with linkage disequilibrium (LD) pruning at $r = 0.5$ and a MAF < 0.03. Due to heterogeneity between sites, adjusting for PCs only, does not remove the site and platform bias. Therefore, we propose an alternative approach of generating PCs for our cohort to control for site and platform bias in addition to ancestry difference. Our approach consists on deriving the PCs using the loadings calculated from reference samples, much like generating PCs using the founders of families. We applied our approach using the electronic Medical Records and Genomics (eMERGE) Venous Thromboembolism African ancestry cohort that consists of four adult sites and four genotyping platforms that had previously been analyzed controlling for site, platform and ancestry. Our results showed that our approach provided similar association results while both controlling for inflation ($\theta = 1.01$ and 1.02 for standard and loadings, respectively) with the advantages of controlling for fewer covariates, thus less degrees of freedom. Therefore, we expect this approach will serve as a “Best Practices” for similar projects, and as a reference for assessing and controlling for confounders in addition to ancestry in genetic association studies.

1809S

Effects of Genotype Uncertainty on Statistical Analysis of Variant Association Studies. *C. Palmer, I. Pe'er.* Columbia University Medical Center, New York, NY.

Background: Standard genotype variant association methods assume high certainty in underlying genotype calls. The extent to which high certainty is actually present varies between genotyping method: array, sequencing, or genotype imputation. In all cases, widespread use of rounding methods ignores missingness in genotypes. Established statistical missingness theory shows that unmodeled bias in missingness patterns may introduce artifacts in downstream analysis. We have undertaken an evaluation of the effects of mishandling missingness in genotype association. **Methods:** We use as a case study genotype imputation, where genotype probabilities are readily accessed from existing software. We have simulated a multi-platform meta-analysis of an association study for complex traits in a simulated cohort of ~20000 individual genomes of stratified European ancestry. Each sample consists of variants at all 1000 Genomes European segregating sites. Sites not present on a particular simulated platform are masked and then imputed using standard genotype imputation with prephasing. We evaluate existing methods for simple linear and logistic regression on imputed genotypes, which handle missingness to varying degrees, and compare these results to similar analyses on the unmasked variants. We extend this testing to other more complex analysis tools (EMMAX, generalized estimating equations, ...), for which alternative strategies for handling genotype uncertainty are not currently available. **Results:** Existing analysis tools operating on uncertain genotype estimates generate directionally biased estimates of variant effect size, where the magnitude of bias tends to increase with increasing genotype uncertainty. Monte Carlo methods for sampling from the posterior distribution of genotypes outperform the best methods for handling missingness currently in use. We furthermore find that imputation probabilities in particular are not consistent, in the sense that they do not accurately reflect the true probability distributions they are estimating. Methods for analyzing imputed data must model and correct this bias. **Conclusion:** The current standard approximation of probabilistic genotype estimates with “best-guess” genotypes is suboptimal. Fortunately, this approximation can be readily avoided with simple resampling techniques. We recommend that uncertainty and bias correction be considered in all future analyses of uncertain genotype data.

1810M

Assessing the power of the Affymetrix Axiom® CEU array for studying rare and low-frequency variants in a European population sample. *B. Schormair[1,2,3], E. Tilch[1,2], D. Czamara[4], B. Pütz[4], J.S. Ried[5], P. Lichtner[1,2], C. Trenkwalder[6], W. Paulus[7], B. Högl[8], K. Berger[9], I. Fietze[10], W.H. Oertel[11], T. Meitinger[1,2], C. Gieger[5], B. Müller-Myhsok[4,12,13], J. Winkelmann[1,2,3,12,14], EU-RLS-GENE-Consortium.* 1) Institute of Human Genetics, Helmholtz Zentrum München, Munich, Germany; 2) Institute of Human Genetics, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; 3) Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; 4) Max Planck Institute of Psychiatry, Munich, Germany; 5) Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 6) Paracelsus Elena Klinik, Kassel, Germany; 7) Department of Clinical Neurophysiology, University Medical Center, Georg August Universität Göttingen, Göttingen, Germany; 8) Neurologische Klinik, Medizinische Universität Innsbruck, Innsbruck, Austria; 9) Institut für Epidemiologie und Sozialmedizin, Westfälische Wilhelms Universität Münster, Münster, Germany; 10) Zentrum für Schlafmedizin, Charité Universitätsmedizin, Berlin, Germany; 11) Neurologische Klinik, Philips Universität Marburg, Marburg, Germany; 12) Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; 13) Institute of Translational Medicine, University of Liverpool, Liverpool, UK; 14) Department of Neurology and Neurosciences, Stanford Center for Sleep Medicine and Sciences, Stanford University, Palo Alto, California, USA.

Genome-wide association studies (GWAS) were initially designed to capture common variants with minor allele frequencies (MAFs) of ≥ 5%. However, for most complex diseases, risk loci identified by GWAS explain only a small proportion of the heritability. Rare (MAF < 1%) and low-frequency variants (1% ≤ MAF < 5%) are suspected to account for a part of the missing heritability. They require large sample sizes for sufficient study power, currently limiting the application of next-generation sequencing. Therefore, information about their coverage by commercial genotyping arrays based on experimental data is important to aid researchers in study design. The Axiom® CEU1 array is optimized for European populations, but only 20% of its content are variants with a MAF < 5%. Autosomal genotype data of 8,488 individuals of European descent, originating from Germany and Austria, was used to assess its performance for rare and low-frequency variants. Standard GWAS quality control was applied. Exclusion criteria for markers included a call rate <98%, discordant calls in duplicates, deviation from Hardy-Weinberg equilibrium with $p<10^{-5}$, and insufficient cluster quality. Subsequently, imputation to 1000Genomes Phase I v3 content was performed. Genotypes were prephased using SHAPEIT, followed by imputation with IMPUTE2 in 5MB-chunks. To optimize representation of rare haplotypes, samples were phased together with 8,752 samples from 7 other European populations genotyped on the same platform. Genotyping quality control had a stronger impact on rare variants compared to low-frequency and rare variants. Whereas 90% and 91.2% of these were retained, only 82% of rare variants were kept. Quality of imputation was judged using IMPUTE2's info criterion, using a value of ≥ 0.8 to indicate high-quality data. For rare variants, only 4% of the 1000 Genomes dataset were imputed with high quality. For low-frequency variants, this increased to 59.7% and for common variants to 85.9%. As a subset of the study samples were also genotyped on the Illumina HumanExome® array, concordance rates for the three categories were calculated. These were all > 99.9%, indicating comparable quality across MAF bins. Low-frequency variants are reasonably well covered by the Axiom® CEU 1 array, rare variants only very fragmentary. Imputation using a reference panel build from both Axiom and HumanExome array data might improve coverage in further samples.

1811T

Addressing population-specific multiple testing burdens in genetic association studies. *R.S. Sobota[1,2], D. Shriner[3], N. Kodaman[1,2], R. Goodloe[1], W. Zheng[4], Y.T. Gao[5], T.L. Edwards[1,4], C.I. Amos[2], S.M. Williams[2].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Genetics, Dartmouth College, Hanover, NH; 3) Center for Research on Genomics and Global Health, NIH, Bethesda, MD; 4) Division of Epidemiology, Department of Medicine, Vanderbilt University, Nashville, TN; 5) Shanghai Cancer Institute, Shanghai, China.

The number of effectively independent tests performed in genome-wide association studies and the corresponding genome-wide significance level varies by population. Therefore, a common p-value threshold may be inappropriate. To assess this, we estimated the number of independent SNPs for all Phase 3 HapMap samples using the LD pruning function in PLINK. We also used an autocorrelation-based approach to verify the HapMap findings, and tested it on 1000 Genomes data to estimate the number of independent tests in whole genome sequences. The number of independent tests is consistent with autocorrelation corresponding to LD pruning at r^2= 0.3. Using the number of independent tests from both methods, we calculated appropriate population-specific thresholds by Bonferroni correction. African populations had the most stringent thresholds (e.g., 1.49×10^-7 for YRI, r^2=0.3), followed by European/Indian populations (3.12×10^-7 for CEU, r^2=0.3), and East Asian populations (3.75×10^-7 for JPT, r^2=0.3). Applying our methods to a previously published GWAS of melanoma, we identified two additional SNPs as significant when data were LD-pruned at r^2=0.3, and five more when pruned at r^2=0.1. These SNPs all mapped to genes previously associated with melanoma; two SNPs were not significant at 5×10^-8, likely representing type 2 errors. Applying our method to a Chinese breast cancer GWAS (r^2=0.3) yielded 50 additional significant SNPs, 19 of which were in or near genes previously associated with the phenotype. Our methods indicate that the conventional correction threshold is not appropriate for many association studies, especially in populations that were recently founded.

1812S

Statistical tests for GWAS in small, admixed populations. *L. Skotte[1], I. Moltke[1,2], A. Albrechtsen[1].* 1) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 2) Department of Human Genetics, University of Chicago, Chicago, IL, USA.

It has recently been demonstrated that genetic association studies in small and isolated populations offer increased statistical power due to extended linkage equilibrium and increased genetic drift over many generations. Many historically small and isolated populations, like the Greenlandic, have recently experienced substantial admixture, which can complicate the association studies, but also further inform them. At a causal SNP locus, the risk allele assumably has the same effect regardless of the carrier's ancestry. However, if a nearby variant is not causative and the linkage disequilibrium pattern differs between the ancestral populations, the effect size may depend on which population the risk allele was inherited from. We introduce a new statistical method for association testing in small, admixed populations, where the effect sizes are allowed to depend on the unknown local ancestry of the allele, which enters as a latent variable. Our method does not rely on accurate inference of local ancestry, yet gives a dramatic increase in statistical power to detect associations and in addition allows for testing for difference in effect size between ancestral populations. Simulation studies show that the statistical power increase from 39% to 96% for a SNP locus with a modest effect size of 0.3 standard deviations in one of the ancestral population and no effect in the other (5000 samples with equally frequent admixture proportions in {0, 0.25, 0.5, 0.75, 1} and a p-value cutoff of 5e-10). In small populations, the samples included in the association study constitute a relatively large proportion of the population and therefore there will be a substantial amount of relatedness between the samples. Admixture and relatedness can lead to decreased power and increased type 2 error rates if not appropriately accounted for. To avoid increased type 2 error rates we use a linear mixed model approach that takes both admixture and relatedness into account.

1813M

The Power Comparison of the Haplotype-based Collapsing Tests and the Variant-based Collapsing Tests for Detecting Rare Variants in Pedigrees. *W. Guo, Y. Yao.* Unit of Statistical Genetics, NIMH, Bethesda, MD.

Abstract Background: Both common and rare genetic variants have been shown to contribute to the etiology of complex diseases. Recent genome-wide association studies (GWAS) have successfully investigated how common variants contribute to the genetic factors associated with common human diseases. However, understanding the impact of rare variants, which are abundant in the human population (one in every 17 bases), remains challenging. A number of statistical tests have been developed to analyze collapsed rare variants identified by association tests. Here, we propose a haplotype-based approach. This work inspired by an existing statistical framework of the pedigree disequilibrium test (PDT), which uses genetic data to assess the effects of variants in general pedigrees. We aim to compare the performance between the haplotype-based approach and the rare variant-based approach for detecting rare causal variants in pedigrees. Results: Extensive simulations in the sequencing setting were carried out to evaluate and compare the haplotype-based approach with the rare variant methods that drew on a more conventional collapsing strategy. As assessed through a variety of scenarios, the haplotype-based pedigree tests had enhanced statistical power compared with the rare variants based pedigree tests when the disease of interest was mainly caused by rare haplotypes (with multiple rare alleles), and vice versa when disease was caused by rare variants acting independently. For most of other situations when disease was caused both by haplotypes with multiple rare alleles and by rare variants with similar effects, these two approaches provided similar power in testing for association. Conclusions: The haplotype-based approach was designed to assess the role of rare and potentially causal haplotypes. The proposed rare variants-based pedigree tests were designed to assess the role of rare and potentially causal variants. This study clearly documented the situations under which either method performs better than the other. All tests have been implemented in a software, which was submitted to the Comprehensive R Archive Network (CRAN) for general use as a computer program named rvHPDT.

1814T

Variation in estimates of kinship observed between whole-genome and exome sequence data. *E. Blue.* Div Med Gen, Unversity of Washington, Seattle, WA.

Genotypic variation is often combined with Mendel's laws of inheritance and population genetic theory to estimate the relationships between individuals. These relationships are clearly important when identifying genetic variants influencing familial disease, when performing quality control to confirm pedigree relationships and when testing the co-segregation of a variant with the phenotype. However, it is also important when testing association of a genetic variant with case/control status in a set of "unrelated" subjects: this is why principal components or estimates of kinship are included as covariates to minimize the effects of population stratification on results. Reliable estimates of the relationship between subjects are necessary.

With the advent of next-generation sequencing technology and the focus on exome data for disease gene discovery, we need to know whether estimates of relationships between samples based on exome data are reliable. The exome represents ~1% of the genome, and definitely does not represent a random subsample of genomic variation. In this study, we compare a method-of-moments and the KING-robust estimator of kinship applied to analyze Illumina HumanOmni1-Quad SNPchip data, whole exome sequence data, and whole genome sequence data for four subjects with known pedigree relationships. SNPchip-based estimates of kinship are quite similar to the pedigree-based expectation, with the KING-robust estimates deviating slightly more than the method-of-moments estimate. However, the exome-based estimates are much more variable: overestimating some relationships by as much as a third and underestimating others by nearly a quarter of the pedigree-based expectation.

We explore the effects of minor allele frequency, linkage disequilibrium, and the number of markers on estimates of kinship drawn from whole genome sequence data. These results suggest that investigators must account for the non-random distribution of variation in the exome when estimating relationships for pedigree verification and disease gene identification.

1815S

Quality control procedures for Whole Exome Sequencing Studies with application to a large family-based study: The International Consortium of Prostate Cancer Genetics (ICPCG) Study. *SK. McDonnell[1], M. Larson[1], S. Middha[1], NB. Larson[1], S. Riska[1], Z. Fogarty[1], S. Baheti[1], M. DeRycke[2], B. Eckloff[3], A. French[2], L. Cannon-Albright[4], J. Stanford[5], WB. Isaacs[6], KA. Cooney[7], J. Schleutker[8], JD. Carpten[9], O. Cussenot[10], G. Giles[11], C. Maier[12], AS. Whittemore[13], F. Wiklund[14], WJ. Catalona[15], W. Foulkes[16], D. Mandal[17], R. Eeles[18], J. Cunningham[2], DJ. Schaid[1], SN. Thibodeau[2] on behalf of the ICPCG.* 1) Dept of Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Dept Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 3) Medical Genome Facility, Mayo Clinic, Rochester, MN; 4) Dept of Internal Medicine, Univ of Utah School of Medicine, Salt Lake City, UT; 5) Fred Hutchinson Cancer Research Center, Seattle, WA; 6) Dept of Urology, The Johns Hopkins Hospital, Baltimore, MD; 7) Dept of Internal Medicine and Urology, Univ of Michigan Medical School, Ann Arbor, MI; 8) Medical Biochemistry and Genetics, Univ of Turku, Turku, Finland; 9) Integrated Cancer Genomics Division, The Translational Genomics Research Institute, Pheonix, AZ; 10) Service Urology-Batiment Gabriel, Hospital Tenon, Paris, France; 11) Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria; 12) Dept of Urology, University of Ulm, Ulm, Germany; 13) Dept Health Research and Policy, Standford University, Stanford, CA; 14) Dept of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; 15) Northwestern University Feinberg School of Medicine, Chicago, IL; 16) Depts of Oncology and Human Genetics, Montreal General Hospital, Montreal, QC; 17) Dept of Genetics, LSU Health Sciences Center, New Orleans, LA; 18) Genetics and Epidemiology, Institute of Cancer Research, Sutton Surrey, UK.

With continued advances in Next-Generation Sequencing (NGS) technologies, Whole-Exome Sequencing (WES) is rapidly emerging as a popular approach to evaluate associations between genetic variation and complex phenotypes. WES provides a powerful tool to uncover disease-causing mutations and, when combined with a family-based study design, increases the chances of finding rare mutations. WES also introduces challenges due to the vast amount of data generated, complexity of bioinformatics processing, and potential batch effects. As relatively large sample sizes become more commonplace, there is an increasing need for fully automatable and methodical quality control procedures. We present a comprehensive yet flexible QC analysis pipeline for WES studies, illustrating its application on a large study of familial prostate cancer. The ICPCG has identified and sampled the most informative high-risk prostate cancer (PC) pedigrees known throughout the world. With the goal of identifying PC susceptibility loci utilizing this extraordinary collection of families, we have performed WES on 539 familial cases of PC derived from 366 families all having at least 3 affected men with PC: 257 cases from 84 families (the majority having 3 sequenced/family) and 282 singleton cases. 490 samples also have HumanOmniExpress array data for comparison of genotype calls. Sequencing used one of two Agilent capture kits (V3 or V4_UTR) and bioinformatic processing was completed using Novoalign alignment and GATK. Extensive quality control procedures were developed aimed at identifying poor quality samples, artifacts with the bioinformatics pipeline, and batch effects due to different sequencing conditions. QC analyses include: 1) coverage of the Agilent capture region; 2) alignment summaries; 3) comparison of variant call quality and counts by functional status; 4) call rates in the capture region; 4) concordance and discovery of array variants with WES variants; 5) sex verification; 6) relationship checks; 7) sample contamination; and 8) population stratification. Final quality control results from the WES of the 539 men will be presented.

1816M

Reduction of systematic bias in transcriptome data from human peripheral blood mononuclear cells for transportation and biobanking. *H. Ohmomo[1], T. Hachiya[1,2], Y. Shiwa[2], R. Furukawa[2], K. Ono[1], S. Ito[3], Y. Ishida[3], M. Satoh[1,2,4,5], J. Hitomi[6,7], K. Sobue[8,9], A. Shimizu[1].* 1) Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Iwate Medical University Disaster Reconstruction Center, 2-1-1 Nishitokuda, Yahaba-cho, Shiwa-gun, Iwate 028-3694, Japan; 2) Division of Biobank and Data Management, Iwate Tohoku Medical Megabank Organization, Iwate Medical University Disaster Reconstruction Center, 2-1-1 Nishitokuda, Yahaba-cho, Shiwa-gun, Iwate 028-3694, Japan; 3) Division of Hematology and Oncology, Department of Internal Medicine, Iwate Medical University School of Medicine, 19-1 Uchimaru, Morioka, Iwate, 020-8505, Japan; 4) Community Medical Supports and Health Record Informatics, Iwate Tohoku Medical Megabank Organization, Iwate Medical University Disaster Reconstruction Center, 2-1-1 Nishitokuda, Yahaba-cho, Shiwa-gun, Iwate 028-3694, Japan; 5) Division of Cardioangiology, Department of Internal Medicine and Memorial Heart Center, Iwate Medical University School of Medicine, 19-1 Uchimaru, Morioka, Iwate, 020-8505, Japan; 6) Deputy Executive Director, Iwate Tohoku Medical Megabank Organization, Iwate Medical University Disaster Reconstruction Center, 2-1-1 Nishitokuda, Yahaba-cho, Shiwa-gun, Iwate 028-3694, Japan; 7) Department of Anatomy, Iwate Medical University School of Medicine, 2-1-1 Nishitokuda, Yahaba-cho, Shiwa-gun, Iwate 028-3694, Japan; 8) Executive Director, Iwate Tohoku Medical Megabank Organization, Iwate Medical University Disaster Reconstruction Center, 2-1-1 Nishitokuda, Yahaba-cho, Shiwa-gun, Iwate, 028-3694, Japan; 9) Department of Neuroscience, Institute for Biomedical Sciences, Iwate Medical University, 2-1-1 Nishitokuda, Yahaba-cho, Shiwa-gun, Iwate 028-3694, Japan.

Transportation of samples is essential for large-scale biobank projects. However, RNA degradation during pre-analytical operations prior to transportation can cause systematic bias in transcriptome data, which may prevent subsequent biomarker identification. Therefore, to collect high-quality biobank samples for expression analysis, specimens must be transported under stable conditions. In this study, we examined the effectiveness of RNA-stabilizing reagents to prevent RNA degradation during pre-analytical operations with an emphasis on RNA from peripheral blood mononuclear cells (PBMCs) to establish a protocol for reducing systematic bias. To this end, we obtained PBMCs from 11 healthy volunteers and analyzed the purity, yield, and integrity of extracted RNA after performing pre-analytical operations for freezing PBMCs at -80°C. We randomly chose 7 samples out of 11 samples, and systematic bias in expression levels was examined by real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR) and RNA sequencing (RNA-Seq). Our data demonstrated that omission of stabilizing reagents significantly lowered RNA integrity, suggesting substantial degradation of RNA molecules due to pre-analytical freezing. qRT-PCR experiments for 19 selected transcripts revealed systematic bias in the expression levels of five transcripts. RNA-Seq for 25,223 transcripts also suggested that about 40% of transcripts were systematically biased. These results indicated that appropriate reduction in systematic bias is essential in protocols for collection of RNA from PBMCs for large-scale biobank projects. Among the seven commercially available stabilizing reagents examined in this study, qRT-PCR and RNA-Seq experiments consistently suggested that RNALock, RNA/DNA Stabilization Reagent for Blood and Bone Marrow, and 1-Thioglycerol/Homogenization solution could reduce systematic bias. On the basis of the results of this study, we established a protocol to reduce systematic bias in the expression levels of RNA transcripts isolated from PBMCs. We believe that these data provide a novel methodology for the collection of high-quality RNA from PBMCs for biobank researchers.

1817T

A Two Step Framework for Integrative Analysis of Genome Wide Methylation and Genotyping Studies. *N. Zhao, M. Wu.* Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA.

Integrative analysis of "multidimensional genomics data", especially genotypes (G) and DNA methylation (M), has been shown to be promising in elucidating the biological processes underlying the phenotype of interest. Currently many large GWAS consortia are expanding to simultaneously examine the joint role of DNA methylation. In this paper we propose a two-step model where in the first step we test for the joint genetic and epigenetic effect on certain phenotype and in the second step we evaluate the potential causal relationship between these variables. We develop statistical approaches that work at gene level for both genotype and methylation, allowing for common units between these two different data types. Powerful kernel machine framework was used in which pair-wise similarity in the trait values between individuals are compared to pairwise similarity in methylation and genotype values for a particular gene, with correspondence suggestive of association. Similarity in methylation and genotype is found by constructing an optimally weighted average of the similarities in methylation and genotype. For a significant gene, we continue to a second step which uses mediation analysis procedure which enables elucidation of the manner in which the different data types work, or do not work, together. We demonstrate through simulations and real data applications that our proposed testing approach often improves power to detect trait associated genes, while protecting type I error. Additionally, the approach can be easily applied to analysis of rare variants and sequencing studies.

1818S

Measuring population stratification in the Brazilian population: how accurate can we be? *L. AlvaradoArnez[1], T.P. Monteiro[2], C.S. Marques[1], C.C. Cardoso[3], M.O. Moraes[1], A.G. Pacheco[2].* 1) Laboratório Hanseníase, Instituto Oswaldo Cruz-RJ. Fiocruz, Brazil; 2) Programa Computação Científica-RJ, Fiocruz, Brazil; 3) Laboratório Virología Molecular, Departamento Genética, UFRJ, Brazil.

Stratification may affect genetic association studies in admixed populations. In Brazil, no study has thoroughly assessed the accuracy of different baseline populations used in controling stratification. We used simulations to compare the accuracy of 3 baseline populations (African and European from 1000 Genomes Browser and Amerindian from Northern Brazil). A stationary population of 30,000 individuals was evolved for 10 generations with starting SNP frequencies calculated from the original baseline. Migration was allowed between populations at varying rates with random mating and 20% fixation rate/generation. For each individual, 3 variables tracked ancestry background. Each scenario was repeated 100 times to account for simulation variability. Data was analyzed with the Admixture software set to obtain 3 clusters both considering as baseline population the same one utilized for simulation and using a different baseline population. Correlation coefficients and an extension of Kaplan-Meier curves were employed for comparisons. Fst was calculated to compare the degree of admixture considering 0.15-0.25 as great, 0.05-0.15 as moderate and 0-0.05 as small genetic divergence. The simulation process was implemented in Python with the library SimuPOP, and statistical analyzes were performed with the R environment. Our results indicate that the baseline population composed of individuals of African ancestry (YRI, LWK,ASW), European ancestry (FIN, GBR, IBS, TSI) and Amerindian (Northern Brazil) performed best in comparison to other baseline populations. Median correlation coefficients for this baseline population ranged from 0.97, 0.95 and 0.90 at migration rates of 0.005 (Fst= 0.45), 0.01 (Fst=0.34) and 0.02 (Fst=0.20) respectively. These results were robust when the other populations were used to start the simulation. The difference between the correlations from the 3 clusters using this baseline population presented a narrower interval as compared to other baseline populations tested. As expected, increased migration rates of 0.03 (Fst= 0.11) and 0.04 (Fst=0.06) reduced median correlation coefficients to 0.78 and 0.79, respectively. These results were also supported by Kaplan-Meier curves with a similar tendency. Our results point that one of the baseline population set is superior to other 2 tested and the accuracy decreases as admixture increases. Further studies are needed to address the impact in a scenario of an association study with stratification.

1819M

Genotyping of the UK Biobank resource, a large extensively phenotyped population collection. *D. Petkova[1], S. Murphy[2], C. Bycroft[1], C. Freeman[1], P. Donnelly[1].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 2) UK Biobank, Units 1-4 Spectrum Way, Adswood, Stockport, Cheshire, SK3 0SA, UK.

The UK Biobank project is a large prospective cohort study of 500,000 individuals aged 40-69 from the UK. With its large sample size and wealth of phenotypic information, the UK Biobank is a powerful resource to investigate how genetic and environmental factors influence a wide range of complex diseases. Samples of the participants' DNA are being genotyped at the Affymetrix Research Service Laboratory on the custom designed UK Biobank Affymetrix Axiom® array with over 800,000 SNPs, including content to provide good genome-wide imputation coverage for variants with frequency >1%, genome-wide content such as eQTLs, missense and loss of function variants, and content for cancers, metabolic and neurological disorders (including Alzheimer's). To introduce the UK Biobank resource, we describe results from quality control and analysis of the first 100,000 samples. In addition to standard QC methods, we discuss strategies for quality control of a very large dataset genotyped in multiple batches, with a large proportion of rare variants. We also characterize various aspects of the UK Biobank dataset: population structure, relatedness, gender misidentification and other sample statistics. The extension of the resource to include genome-wide genetic information on each individual is currently underway.

1820T

Spurious cryptic relatedness can be induced by population substructure, population admixture and sequencing batch effects. *D. Zhang, S. Li, G.T. Wang, S.M. Leal.* Center for Statistical Genetics, Baylor College of Medicine, Houston, TX.

It is important to identify cryptically related individuals in population-based association studies, since inclusion of related individuals can increase type I & II errors. Mixed models have been proposed to analyze data when related individuals are included in the study, but these methods can be computationally intensive and do not always sufficiently control type I & II errors. Another option is to only include one individual from a relative pair/group in the analysis. In population-based studies cryptically related individuals should be identified when performing data quality control. However, caution should be used, since population substructure/admixture and sequence data batch effects can cause detection of spurious relatedness. In order to investigate the problem we systematically evaluated the relatedness of 1,092 samples in the 1000 Genomes project and 2,300 African-American (AA) subjects from the NHLBI-Exome Sequencing project (ESP) via two published methods for kinship inference: (i) the PLINK algorithm which is based on identical-by-descent (IBD) statistic under the assumption of homogeneous population, and (ii) the KING-robust algorithm which uses an estimate of the genome-wide average heterozygosity across individuals to compute an estimator of kinship coefficient. For the 1,000 Genomes project we analyzed the data by country of origin, e.g. Japan, continent of origin e.g. Asia and sequencing batch by continent of origin and for the ESP data we analyzed the data for all AA individuals and separately by sequencing batch. We identified spurious cryptically related pairs of samples due to population substructure/admixture and sequencing batch effects with both methods, but the problem was more extreme for PLINK. For example, an excess of 3rd degree relatives were observed for AA individuals, when sequencing batch was not considered and also for 1000 Genomes project data when population substructure was introduced by analyzing the data by continent. Additionally, the kinship coefficients varied depending on how the analysis was performed, e.g. accounting for sequencing batch or not, caused reclassification of individuals, e.g from 1st degree to 2nd degree relatives. In addition to presenting the results of these analyses and showing the severity of the biases in the kinship coefficients, we also demonstrate strategies which can be used to evaluate if individuals are truly cryptically related or spurious relationships have been identified.

1821S

Correcting for population stratification in secondary genetic association studies using subsamples. *M.C. Babron[1,2], S. Benhamou[1,2,3], E. Génin[4], R. Kazma[5].* 1) Inserm U946, Variabilité Génétique et Maladies Humaines, Paris, France; 2) Université Paris-Diderot, Sorbonne Paris-Cité, UMRS-946, Paris, France; 3) Service de Biostatistique et d'Epidémiologie, Gustave Roussy, Villejuif, France; 4) Inserm U1078, Génétique, Génomique Fonctionnelle et Biotechnologies, CHU Morvan, Brest, France; 5) Centre National de Génotypage, Institut de Génomique, CEA, Evry, France.

Population stratification (PS) is a potential cause of false positive results in genome-wide association studies, when cases and controls are drawn from a population comprising multiple groups with different disease prevalences. To correct for PS, the inclusion of several principal components (PCs) of genome-wide genotypes as covariates has been shown to control for the inflation of association statistics and has become standard procedure. After testing for genetic association with a primary phenotype of interest, secondary hypotheses are often tested using a fraction of the initial sample for which the secondary phenotype is available. In practice, correction for subsample stratification is done using the PCs that were calculated for the full initial sample. This approach makes the assumption that the subsample has a population structure similar to that of the initial sample, used to calculate the PCs. However, this might not always be the case. In fact, the strategy that uses PCs calculated on the initial sample to correct for PS, when testing secondary hypotheses on a subsample, has not yet been evaluated. Here, we assess the robustness of PS correction using PCs calculated with the full initial sample or a subsample that comprises a distribution of subpopulations that is different from the initial population. First, the collection of 5,811 individual genome-wide genotypes (Illumina 317k) from 13 European countries is considered as the initial case-control sample. Then, to simulate subsamples with a different population structure, we randomly draw sets of cases and controls with different proportions from extremes. These extremes match up with the North-East and South-West quadrants of the space defined by the first two PCs calculated on the initial sample. Simulating 100 replicates of a subsample with 75% of cases and 25% of controls belonging to the North-East quadrant (the rest belonging to the South-West quadrant), the Q-Q plots show a strong type 1 error inflation ($\theta_{GC} = 1.3513$) without PS correction which is corrected satisfactorily with as low as 2 PCs calculated on the initial sample ($\theta_{GC} = 1.0019$) or on the subsample ($\theta_{GC} = 1.0014$). However, the use of PCs calculated on the subsample corrects PS better than PCs calculated on the initial sample for specific variants in genes which are strongly stratified along this axis (*LCT*). Further ongoing simulations will assess these approaches in different situations of subpopulation structure.

1822M

Correction for population stratification and relatedness in case-control studies using logistic mixed models. *H. Chen, C. Wang, Z. Li, X. Lin.* Biostatistics, Harvard School of Public Health, Boston, MA.

In genome-wide association studies (GWAS), population stratification and relatedness can result in spurious association findings. Principal component analysis (PCA) has been widely used to adjust for population structure. However, PCA does not work well in the presence of familial or cryptic relatedness in addition to population stratification. Linear mixed models have been proposed as an alternative method for correcting for population stratification and relatedness for continuous phenotypes. However, the normality assumption required by linear mixed models is violated for binary phenotypes in case-control studies. We propose to use logistic mixed models to control for both population stratification and familial or cryptic relatedness in GWAS for case-control data. We develop a computationally efficient score test for association analysis using logistic mixed models. Our simulation studies show that our method performs well in controlling type I error rates in the presence of population stratification and relatedness for case-control data. We illustrate our method in a real data example.

1823T

Control of population stratification in family data using pedigree information and ancestry principal components. *C. Wang, H. Chen, X. Lin.* Biostatistics, Harvard School of Public Health, Boston, MA.

Genetic data collected from families with disease affected members are more likely to enrich disease causal variants. Population-based association tests, when applied to family data, need to control for both population structure and family relatedness. Principal components analysis (PCA), which has been widely used to control for population structure in unrelated samples, does not work well in capturing the population structure in family data. We have previously developed a method called LASER, which can estimate individual ancestry in a reference principal component space based on a set of unrelated reference individuals. The LASER method is robust to the presence of family relatedness in the study sample. Based on ancestry coordinates estimated from LASER, we propose to control for population structure and family relatedness using mixed model and generalized estimation equation (GEE) approaches. We adjust the population structure by regressing on ancestry coordinates and account for family relatedness using the kinship matrix to specify the covariance structure in mixed models and GEEs. Using both simulations and empirical data from the Framingham Heart GWAS Studies, we show that this approach performs well for both continuous traits and binary traits. When genome-wide genotypes are available, our approach outperforms the standard PCA method. Furthermore, because LASER can accurately estimate individual ancestry using extremely low coverage sequencing data, our approach can be applied to targeted sequencing or exome sequencing studies where genome-wide data are not available.

1824S

Optimal strategies for studying singletons associated with quantitative traits. *S. Rashkin, G. Jun, G. Abecasis.* Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

As the focus genetic association studies shifts from common variants to rare variants, it is important to determine the optimal strategies for conducting studies of very rare variants. A common argument is that, in order to detect rare variants with high power, it is necessary to deeply sequence samples. However, some studies have shown that low-pass sequencing can also detect many rare variants while enabling larger sample sizes. We aim to identify sequencing strategies that maximize power (for a fixed sequencing effort) for studies that explore the role of rare variants in human disease. We specifically focused on variants that appear in a single sequenced sample as they are the most difficult type of variant to identify; designs with sufficient power to detect associations with disease for singletons should also work well for more common variants. Our previous work on binary traits indicates that power to detect association is maximized at 10x coverage. Here, we extend this analysis to quantitative traits.

We assessed the power to detect singletons using a simulated multi-sample caller using different read depths, sample sizes, sequencing error rates, and false positive rates. The power to detect singletons increases as coverage increases, reaching saturation at coverage about 15-20x. We validated these simulations by down-sampling deeply sequenced exome samples and assessing our ability to detect previously called singletons for different sample sizes and read depths.

We performed simulations to assess the power to detect association between the count of singletons in a gene and a quantitative trait, using a simple linear regression framework. We estimated association study power across a range of values for read depth (2-50x), sample size (100-5000), population frequency of singletons (0.001-0.5% per gene per person), effect size, baseline trait value, sequencing error rates (0.001-0.01), and false positive rate for detecting singletons (0.0001-0.05). Preliminary results indicate a similar trend as seen in binary traits: power to detect an association between singletons and a quantitative trait seems to be maximized around 10x and coverage should only be increased beyond this threshold if it does not require a decrease in sample size.

1825M

Accurate Error Rates: Calculating Reproducibility by Minor Allele Frequency. *J. Romm, P. Zhang, H. Ling, K.F. Doheny.* CIDR/JHU, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. Many commercially available genotyping arrays (HumanExome, HumanOmni2.5, HumanOmni5 and all Illumina "plus exome" arrays) contain SNP content that enables users to assay low minor allele frequency (MAF) variants (<1%). Although overall sample reproducibility rates have not changed with the use of arrays containing a fraction of low MAF content, there are concerns about the accuracy of low MAF SNPs given the nature of the genotype cluster and calling algorithms. In order to get a better picture of the entire project, minor allele reproducibility rates from study duplicate pairs have been evaluated in different MAF bins (0-1%, 1%-5%, 5%-10%, 10%-50%). We exclude major allele homozygotes in the computation of reproducibility in order to alleviate the "overwhelming" effect when computing reproducibility on rare variants. As expected, SNPs in the lower MAF bins have lower reproducibility rates, especially when the sample call rate is lower for 1 sample in the pair being evaluated. As the MAF of the SNPs increase, the sample reproducibility rate increases, and this seems to be less dependent on sample call rate. We have also computed reproducibility using similar statistics, such as kappa coefficient, which allow us to use all of the data (including major allele homozygotes) and we have received similar results. Our data can be used to guide decisions on expected realistic error rates for variants of different allele frequencies. It can also be used to guide decisions on sample call rate cut-off decisions for data cleaning, based on the type of analysis being done - i.e. rare versus common variant analyses may require different sample level call rate threshold cut-offs.

1826T

Factors affecting relative telomere length measurements by quantitative PCR. *C.L. Dagnall[1,2], B.D. Hicks[1,2], A.A. Hutchinson[1,2], S.M. Gadalla[1], L. Mirabello[1], B.J. Ballew[1], J.D. Figueroa[1], C. Bodelon[1], J. Liu[1], J. Hofmann[1], S.A. Savage[1], M. Yeager[1,2].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI, National Institutes of Health (NIH), Bethesda, MD; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD.

As variation in telomere length have been associated with risk of cancer, heart disease, and various other illnesses it is important to understand the pre-analytic and analytic factors affecting relative telomere length measured by quantitative PCR (qPCR). We have conducted multiple studies comparing various DNA extraction methods, methods of removing potential residual PCR inhibitors, and sample storage conditions as factors that may influence relative telomere length measurements obtained from qPCR assays which lead to discrepancies within study results. We have confirmed that the DNA extraction method affects the measurement of telomere length defined by T/S ratio (concentration of telomeric repeat sequence normalized by concentration of single copy gene sequence). We studied 40 individuals had DNA extracted from the same buffy coat source specimen by three extraction techniques. The dynamic ranges of the T/S ratios from Qiagen QIAamp- and QIAsymphony-extracted DNA samples were lower (0.39-0.87, and 0.29-0.74, respectively) than the Promega ReliaPrep-extracted DNA samples (0.51-1.46). We observed only moderate correlation between paired individuals and different extraction methods (R-squared ranges from 0.32 to 0.46). In addition, we evaluated the differences in telomere length measurements due to inter-plate variability and show that analysis techniques effectively standardize results across assay plates, significantly reducing the average %CV of replicates (from 17.4% to 8.2% in a data set of 170 replicates from 3 samples all extracted using the same extraction technique). Due to these large differences in dynamic range and correlation between extraction methods, it is advisable to only measure relative telomere length when samples have been extracted using the same technique. Eliminating the factors which cause variation in relative telomere length measurements increases the reliability of results when investigating associations between telomere length and various diseases. Results from inhibitor removal methods and sample storage conditions will also be presented.

1827S

Identifying relative pairs within large datasets. *Z. Zeng[1], D.E. Weeks[1,2], W. Chen[3], N. Mukhopadhyay[1], E. Feingold[1,2].* 1) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 3) Division of Pediatric Pulmonary Medicine, Allergy and Immunology, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA.

When a large number of people from a small geographical region are sequenced (e.g., by a regional healthcare system), it is of interest to identify individuals within the datasets who are closely related. The first step in this process is to identify regions of identity-by-descent (IBD) between pairs of individuals. Shared IBD segments between individuals can be inferred by identity-by-status of streaks of SNPs using dense marker data. Existing methods for detecting IBD segments, such as implemented in BEAGLE, GERMLINE and PLINK, are likelihood-based. Some of these methods are extremely time-consuming (BEAGLE), and some require extra assumptions, such as independence of SNPs (PLINK) or correctly phased genotypes as input (GERMLINE). To strike a balance between computational time, accuracy, and extra requirements on the data, we propose an empirical method for detecting IBD segments and indentifying recombination events in close relative pairs. Our method first determines confident IBD status in small windows and then fills in the unsure gaps between the confident ones to generate the intact IBD segments. We explored combinations of different strategies (large sliding window vs. small fixed window; reference panel vs. no reference panel; windows based on physical distance vs. those based on a fixed number of SNPs) and developed a new algorithm that is computationally efficient and does not require knowledge of putative relationships. Our algorithm can be used as a building block for detecting relationships in large datasets of putatively unrelated individuals or for testing relationships in pedigree datasets. We applied our algorithm on a group of grandparent-grandchild pairs and evaluated its performance by comparing with a relationship-aware algorithm we are developing as well as on artificially synthesized IBD data. This research was funded under NIH (grant number: R01 38979).

1828M

Leveraging Family Structure for the Analysis of Rare Variants in Known Cancer Genes from WES of African American Hereditary Prostate Cancer. *C.D. Cropp[1], S.K. McDonnell[2], S. Middha[2], M.S. DeRycke[2], D. Karyadi[3], D. Schaid[2], S.N. Thibodeau[4], W.B. Isaacs[5], E.A. Ostrander[3], J. Stanford[6], K.A. Cooney[7], J.E. Bailey-Wilson[1], J.D. Carpten[8] On behalf of the International Consortium for Prostate Cancer Genetics (ICPCG).* 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute/National Institutes of Health, Baltimore, MD; 2) Departmant of Health Science Research, Mayo Clinic, Rochester, MN; 3) Cancer Genetics Branch, National Human Genome Research Institute/National Institutes of Health, Bethesda, MD; 4) Departmant of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 5) Johns Hopkins University School of Medicine, Johns Hopkins Hospital, Baltimore, MD; 6) Public Health Sciences Division, Epidemiology Program, Fred Hutchinson Cancer Research Center, Seattle, WA; 7) University of Michigan Comprehensive Cancer Center, Ann Arbor, MI; 8) Integrated Cancer Genomics Division, Translational Genomics Research Institute (TGen), Phoenix, AZ.

Prostate cancer (PRCA) is the second leading cause of cancer death in North American men and it disproportionately affects African American (AA) men, who have higher incidence and mortality rates compared to men without known African ancestry. Although several studies of hereditary PRCA exist resulting in a few significant and reproducible loci associated with an increased risk of PRCA across different study populations (e.g. *HOXB13*), disentangling the environmental and genetic contributions towards the persistent health disparity in AA with hereditary PRCA remains elusive. As part of the international effort to address this disparity, the African American Hereditary Prostate Cancer Study (AAHPC) was developed as a national collaboration to explore the role of genetics in the causation of hereditary PRCA in AA. AAHPC is in partnership with the International Consortium for Prostate Cancer Genetics (ICPCG), which conducts collaborative studies of PRCA genetics in multiplex families. As part of an ICPCG sequencing study of 539 affected individuals from 366 PRCA pedigrees, we performed whole exome sequencing on 16 AAHPC affected men from 12 pedigrees. In the combined ICPCG AA cohort (N=22), there were 60/141 rare and/or novel coding variants (minor allele frequency, MAF $\leq$ 5%) in 15 known cancer-associated genes, 35 of which were non-synonymous. Post-variant calling AAHPC quality control was implemented using Golden Helix SVS 8 software with filters set for removal of variants with Read Depth < 10, Quality Score < 20, Quality Score:Read Depth Ratio < 0.5, Call Rate < 0.75. Variants were additionally filtered by MAF based on the NHLBI ESP650051-V2 exomes variant frequencies for AA population using a MAF threshold of 5%. After QC, 207,741 variants remained for further analysis. In these analyses, we focused on the same aforementioned cancer causing genes. Two AAHPC sequenced families had > 1 affected members (3 per family). Under the dominant model, all affecteds in Family 1 shared 2 variants in two of these genes (25 shared by 2 affecteds). All affecteds in Family 2 shared 9 variants in two different known cancer genes (29 shared by 2 affecteds). Of all 65 shared rare variants in these 2 families, 40 are shared by other ICPCG AA PRCA-affected men. Additional QC is underway to validate these variants and bioinformatic analyses are being used to predict effects of the variants in an effort to unravel the complex genetic heterogeneity of hereditary PRCA in AA.

1829T

Detection of meiotic breakpoints in families using dense genotyping data. *N. Mukhopadhyay[1], F. Begum[2], Z. Zeng[3], E. Feingold[3,4].* 1) Oral Biology, Univ Pittsburgh/Sch Dental Med, Pittsburgh, PA; 2) Dept of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 3) Dept. of Biostatistics, School of Public Health, Pittsburgh, PA; 4) Dept. of Human Genetics, School of Public Health, Pittsburgh, PA.

Meiotic recombination is important for proper segregation of chromosomes; errors in recombination may result in chromosomal abnormalities and non- disjunction. Both the total number and the pattern of recombination events are known to vary genome-wide and from person to person. Using genome-wide genotype data to detect locations of recombination in probands is the first necessary tool to study recombination. Earlier methods, e.g. CRIMAP, used linkage-style modeling on three-generation families and sparse microsatellite markers to detect recombination events. More recently, methods that look for "streaks" of SNPs showing IBD status on dense GWAS SNP data have been used to score recombination locations in sibships. We have now developed a new SNP streak method and software to score recombination locations in pedigree types not previously handled, such as half-sibling pedigrees, and pedigrees with one or more ungenotyped individuals. Our method can be used to substantially increase sample sizes for performing GWAS on recombination phenotypes. We discuss ways to handle inherent problems such as uninformative mating combinations, genotyping errors and missing genotypes. Recombination scoring results on real family study datasets will be presented.

1830S

PIX-LRT: A parent-informed test for SNPs on the X chromosome using case-parent triads. *C.R. Weinberg[1], A. Wise[1,2], M. Shi[1].* 1) Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC., Select a Country; 2) Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA.

The X-chromosome is generally understudied in association studies, in part because the analyst has limited methodological options. For nuclear-family-based association studies, most current methods extend the transmission disequilibrium test (TDT) to the X-chromosome. We present a new method to study association in case-parent triads: the parent-informed likelihood ratio test for the X-chromosome (PIX-LRT). Our method takes advantage of parental genotype information and the sex of the affected offspring to increase statistical power to detect an effect. Under a parental exchangeability assumption for the X, if case-parent triads are complete, the parents of affected offspring provide an independent replication sample for findings based on transmissions to the affected offspring. For each offspring sex we combine the parent-level and the offspring-level information to form a likelihood ratio test statistic; we then combine the two to form a composite test statistic. Our method can estimate relative risks under different modes of inheritance or a more general co-dominant model. In triads with missing parental genotypes, the method accounts for missingness with the Expectation-Maximization algorithm. We calculate non-centrality parameters to demonstrate the power gain and robustness of our method compared to alternative methods. We apply PIX-LRT to publically-available data from an international consortium of genotyped families affected by the birth defect oral cleft and find a strong, internally-replicated signal for a SNP marker related to cleft lip with or without cleft palate.

1831M

Simulation Analysis to Assess Linkage Results of Class III Malocclusion and Human Chromosome 11. *L.K AlOthman[1], N. Mukhopadhyay[1], L. Otero[2], R.M. Cruz[3], J. Turner[4], L.A. Morford[5,6], S. Oliveria[7], J.K. Hartsfield, Jr[5,6], M. Govil[1].* 1) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 2) Centro de Investigaciones Odontológicas, Pontificia Universidad Javeriana, Bogota, Colombia; 3) Private Orthodontic Practice, Brasilia, Brazil; 4) Private Orthodontic Practice North Carolina, USA; 5) Division of Orthodontics, University of Kentucky College of Dentistry, Lexington, KY; 6) Hereditary Genomics/Genomics Core Laboratory, Center for Oral Health Research, University of Kentucky College of Dentistry, Lexington, KY; 7) Department of Genetics and Morphology, Universidade De Brasilia, Brasilia, Brazil.

A simulation study was conducted to evaluate multipoint linkage results for chromosome 11 genetic data from two multi-generational family-based cohorts located in South America with a high prevalence of class III malocclusion. Based on previous data implicating a region of chromosome 11 in the Class III phenotype, 4 single nucleotide polymorphisms (SNPs; rs666723, rs578169, rs1386719 and rs12416856) were genotyped within 17 Brazilian families (178 individuals) and 15 Colombian families (248 individuals). The families in each cohort varied greatly in the size, structure and the number of affected individuals. Class III affected and unaffected individuals were diagnosed based on cephalometric measurements, models, photographs and/or oral examination. MMLS multipoint HLODs maximized over different levels of heterogeneity and two genetic models (reduced penetrance dominant and recessive) were generated using SimWalk2. To estimate the empirical significance of these multipoint HLODs, 1000 replicates of unlinked genotype data based on real data pedigree structures, affection status and pattern of missing genotypes were simulated for the Brazilian and Colombian cohort using SLINK and SIMULATE respectively. These replicates were then analyzed using SimWalk2 with the original maximizing mode of inheritance (MOI). Power was estimated similarly for each cohort by generating 1000 replicates of pedigree data linked to the SNP with the highest HLOD. The corresponding cohort-specific MOI was used for the power simulation genetic parameters. For the Brazilian cohort: (a) the empirical HLOD for α= 0.05 was 0.51, (b) empirical p-value for HLOD=1.84 was <0.001, and (c) power for suggestive linkage (HLOD≥2) was 72%. For the Colombian cohort: (a) the empirical HLOD for α=0.05 was 0.29, and (b) an empirical p-value= 0.023 for HLOD=0.51 was observed. These simulation results support potential linkage on chromosome 11. The current linkage results are for the MMLS maximized (by definition) over two genetic models; work is underway to compute MODs, which will take into consideration the entire genetic parameter space. Due to complexity of the Colombian cohort pedigrees, power calculations are pending.

1832T

Rare Variant Association Test for Nuclear Families. *Z. He[1], N. Krumm[2], G. Wang[1], B. O'Road[3], E. Eichler[2], S. Leal[1], Simons Simplex Sequencing Consortium.* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR.

The development of next-generation sequencing (NGS) technologies, has led to many population-based complex trait association studies of rare variants (RVs). These studies are vulnerable to spurious associations due to population stratification. Approaches employing trio (two parents and an affected proband) data such as the transmission disequilibrium test [TDT (Spielman et al. 1993)] can avoid the problems of population stratification. Recently the TDT was extended to analyze RVs [RV-TDT (He et al. 2014)]. It was shown that RV-TDT controls type I error and also has considerable power to detect associations. The TDT analyses only employ information on an affected offspring and their parents. When there are additional siblings, it is advantageous to include them in analysis since they also provide association information. We extended the RV-TDT to analyze all types of independent nuclear families (NF) with at least one affected offspring (RV-NF) and test for RV associations using four commonly used methods: CMC (Li & Leal, 2008), WSS (Madsen & Browning 2009), BRV (Morris & Zeggini, 2010) and VT (Price et al. 2010). We demonstrate that for all RV-NF tests type I error is well controlled even when there is a high level of population stratification or admixture. The power of the RV-NF test was evaluated using a number of disease models and nuclear pedigree configurations. The RV-NF is considerably more powerful than the RV-TDT to detect associations. For example using the CMC version of the RV-TDT and RV-NF the power was evaluated by generating variant data for a 1,500bp gene for which the causal RVs [minor allele frequency (MAF)<1%] have an odds ratio of 2.0. The power to detect and association is: 0.49 for 1,000 trios: 0.58 for 1,000 nuclear families with one affected child and an unaffected child; and 0.65 for 1,000 nuclear families with two affected children. In order to illustrate the application of the RV-NF methods, the exome data from 600 autism spectrum disorder nuclear families with one affected child and one unaffected child were analyzed. RV associations with autism were found for several genes. Given the problem of adequately controlling for population stratification and admixture in RV association studies, the capability of analyzing all types of nuclear families and the growing number of nuclear family studies with NGS data, the RV-NF method is extremely beneficial to elucidate the involvement of RVs in the etiology of complex traits.

1833S

The collapsed haplotype pattern method for linkage analysis of next-generation sequencing data. *G.T. Wang, D. Zhang, B. Li, H. Dai, S.M. Leal.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Traditionally, linkage analysis was used to map Mendelian diseases and genes within the linked regions were sequenced to identify the causal variants. Recent advances in next generation sequencing (NGS) make it possible to directly sequence genomes and exomes of individuals with Mendelian diseases and identify causal mutations by filtering variants in an affected individual(s) family member(s), removing those variants with higher allele frequency, e.g. >0.1% in variant databases. Linkage analysis of SNP data are sometimes used in conjunction with NGS to increase the success of identifying the causal variant. With the reduction in cost of NGS, DNA samples from entire families can be sequenced and linkage analysis can be performed directly using NGS data. Inspired by "burden" tests which are used for complex trait rare variant association studies, we developed the collapsed haplotype pattern (CHP) method to generate markers from sequence data for linkage analysis. To demonstrate the power of the CHP method compared to analyzing individual variants, we analyzed and performed empirical power calculations using the allelic architecture for several known non-syndromic hearing loss genes, i.e. GJB2, SLC26A4, MYO7A & MYH6. Power analysis demonstrated that the CHP method is substantially more powerful than analyzing individual SNVs in the presence of inter-allelic familial heterogeneity, i.e. families have different pathological variants within a gene or intra-familial heterogeneity e.g. compound heterozygotes. Specifically for an autosomal recessive model with allelic heterogeneity and locus heterogeneity of 50%, it requires 12 families for the CHP method to achieve a power of 90% for the SLC26A4 gene, while analyzing individual SNVs requires >50 families to achieve the same power at a genome-wide significance level of α=0.05. Unlike the commonly practiced filtering approaches used for NGS data, the CHP method provides statistical evidence of the involvement of a gene in Mendelian disease etiology. Additionally because it incorporates inheritance information and penetrance models it is less likely than filtering to exclude causal variants in the presents of phenocopies and/or reduced penetrance. We recommend the use of the CHP method in parallel to filtering methods to take full advantage of the power of NGS in families. The CHP method is incorporated in the SEQLinkage software which is freely available http://www.bioinformatics.org/seqlink/.

1834M

Identifying rare variants in linkage regions through pedigree-based conditional linkage analysis. *C.W. Bartlett[1], S. Buyske[2,3], S.L. Wolock[1], L.M. Brzustowicz[3].* 1) Battelle Ctr Mathematical Med, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) Department of Statistics, Rutgers University, NJ; 3) Department of Genetics, Rutgers University, NJ.

Linkage analysis is sensitive for detection of rare mutations, even when multiple rare variants in that region are causal in the population, as long as the same single variant segregates with any single family. However, linkage analysis has poor resolution, implicating large genomic regions with many candidate genes and numerous candidate mutations within each candidate gene. Here we develop a method, based on a variation on likelihood ratio testing, to assess if a set of candidate variants in a linkage region underlie or produce the observed linkage. We show that our conditional linkage test can be used to determine which variants are responsible for an observed linkage signal. Our hypothesis that pedigrees contained information useful for conditional linkage was first validated using ball and urn model probability calculations to develop exact analytical demonstrations and to inform downstream modeling choices. Our approach involves calculating a pedigree likelihood stratifying subjects by liability classes determined by a candidate variant "propensity" score. When variants contributing to the propensity score are primarily responsible for the linkage signal, the linkage signal drops once liability classes based on that score are incorporated, demonstrating that the tested variants are required to observe the linkage signal. We use lasso logistic regression of all loci in the gene against affection status to generate subject-level propensity scores and apply a decision rule to create liability classes, though in in our simulations we did not observe ambiguous case that required more than 3 liability classes. Using an integrated (i.e., model-averaged) LOD score, we have derived an empirical null distribution for the variability in the integrated LOD with inclusion of liability classes relative to baseline, to allow us to define p-values for significance testing. Power was estimated through simulation over a range of generating conditions with varying numbers of causal and non-caudal SNPs within a linked region. In general, power increases as the absolute number of non-causal candidate variants decreases and the ratio of causal variant to total test variants increases. In combination with modest bioinformatic filtering of genes and variants, our pedigree method for conditional linkage analysis may provide statistical evidence for a specific gene, or gene set, to help reduce the testing in model organisms and systems.

1835T

A general framework for group-wise transmission/disequilibrium tests for identifying rare variant associations. *R. Chen[1], W. Chen[2], X. Zhan[3], B. Li[1].* 1) Molecular Physiology and Biophysics, Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA; 3) Department of Clinical Science, University of Texas Southwestern Medical Center, Dallas, TX.

A major focus of current sequencing studies for human genetics is to identify rare variants association with complex disease. Aside from reduced power of detecting associated rare variants, controlling for population stratification is particularly challenging for rare variants. Transmission/disequilibrium tests (TDT) based on family designs have been proved to be immune to population stratification and admixture, and therefore provide an effective approach to rare variant association studies to eliminate spurious associations. Moreover, group-wise strategies, originally proposed for population-based studies, have been extended to family designs to collectively analyze multiple rare variants as a single unit. In this study we describe a general framework for group-wise TDT (gTDT) and develop an efficient tool that implemented several common genetic models, including additive, dominate and compound heterozygous models. Another important feature of gTDT is that it's a flexible haplotype-based framework that can encompass a variety of other genetic models as well. For example, we implemented a hybrid model, which combines compound heterozygous and weighted additive models to test for recessive models while incorporating prior knowledge assigned to different variants. Unlike existing methods, gTDT constructs haplotypes by transmission when possible and inherently takes into account the LD among variants. Through extensive simulations we showed that type I error is correctly controlled for rare variants under all models investigated, and it is also true in the presence of population stratification and admixture. Under a variety of genetic models gTDT is more powerful than the classical TDT. We hope that gTDT is useful to analyze pedigree data with the flexibility to test for different models for genetic studies of complex disease.

1836S

Rare Variant Association Analysis of Quantitative Traits in Pedigrees of Arbitrary Size and Structure. *Y. Jiang[1], K. Conneely[2], M. Epstein[2].* 1) Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA; 2) Department of Human Genetics, Emory University, Atlanta, GA.

Family-based rare-variants sequencing studies have attractive characteristics: they can be more powerful compared to case-control and population-based studies due to increased genetic load and can further enable the implementation of rare-variant association tests that, by design, are robust to confounding due to population stratification. However, current methods for family-based studies mainly focus on nuclear families or parent-offspring trios while ignoring the information provided by other types of relatives. With this in mind, we propose a powerful rare-variant test for analysis of quantitative traits in extended pedigrees that can accommodate any family size, and is robust to population stratification. For each non-founder, our method partitions rare variants in a region of interest into two components: a between-family component (sensitive to population stratification but less sensitive to genotyping error) based on information from all ancestors as well as an orthogonal within-family component (which is robust to population stratification). We then test for association between the robust within-family component and the trait of interest using a variance-component score test that accounts for relatedness among family members. As ignoring the between-family component may lead to power loss, we further adopted a two-stage test to enhance power: at the first stage, we construct a variance-component score test using the between-family component to identify top hits; at the second stage, we follow up these top hits by constructing the variance-component score test on the independent within-family component. Our method maintains power when a region harbors variants acting in different directions on outcome, accommodates covariates, and efficiently calculates p-value using an asymptotic distribution. As more and more studies have re-sequenced large pedigrees from previous linkage studies, our method is in need. We will show the robustness and power of our method using simulated data and further illustrate the approach using sequencing data from the SardiNIA Medical Sequencing Discovery Project.

1837M

Dissecting the Genetic Architecture of Longevity with Millions of Individuals. *J. Kaplanis[1,2], B. Markus[1], M. Gershovits[1], M. Sheikh[1], A. Price[2,3], D. MacArthur[4], Y. Erlich[1].* 1) Whitehead Institute for Biomedical Research, Cambridge, MA; 2) Harvard School of Public Health, Boston, MA; 3) Broad Institute of Harvard and MIT, Cambridge MA; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA.

Understanding the genetic architecture of complex traits is one of the top missions of human genetics. Emerging lines of studies have highlighted the entangled etiologies of these traits, which can include epistasis, parent-of-origin effects, sex and age interactions, and environmental risk factors. To conduct robust genetic epidemiological analysis, statistical models require sampling substantial amount of data from large families. However, the recruitment of large cohorts of extended kinships is both logistically challenging and cost-prohibitive. Here, we present a big data strategy to address this challenge: harnessing existing, free, and massive Web 2.0 social network resources to trace the aggregation of complex traits in extremely large families. We collected millions of public profiles from Geni.com, the world's largest genealogy-driven social network. Using this information, we constructed a single pedigree of 13 million individuals spanning many generations up to the 15th century. In addition, Natural Language Processing was used to convert genealogical information into birth and death locations to obtain a proxy for environmental factors. We obtained multiple of phenotypes from this resource including longevity, fertility, and migration patterns. This dataset provides a wide range of kinships for familial aggregation studies. We compiled this data in a new resource called FamiLinx which will be free available for the entire scientific community. We will show the utility of FamiLinx for heritability estimates across distant relatives to disentangle analysis of epistasis, parent-of-origin, and shared environments.

1838T

Adaptive Combination of P-values for Family-based Association Testing with Sequence Data. *W. Lin.* Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan.

Family-based study design will play a key role in identifying rare causal variants, because rare causal variants can be enriched in families with multiple affected subjects. Furthermore, different from population-based studies, family studies are robust to any bias induced by population substructure. It is well known that rare causal variants are difficult to detect with single-locus tests. Therefore, burden tests and non-burden tests have been developed to combine signals of multiple variants in a chromosomal region or a functional unit. This inevitably incorporates some neutral variants into the test statistics, which can dilute the power of statistical methods. To guard against the noise caused by neutral variants, we here propose an 'adaptive combination of P-values method' (abbreviated as 'ADA'). This method combines per-site P-values of variants that are more likely to be causal. Variants with large P-values (which are more likely to be neutral variants) are discarded from the combined statistic. In addition to performing extensive simulation studies, we applied these tests to the Genetic Analysis Workshop 17 data sets, where real sequence data were generated according to the 1000 Genomes Project. We show that this ADA method outperforms the burden tests and non-burden tests when the percentage of causal variants is small. ADA is recommended for its potential to guard against the noise induced by neutral variants.

1839S

Exome sequencing in an isolated population reveals multiple rare variants affecting both high-density lipoprotein cholesterol and the levels of certain blood metabolites. *E.M. van Leeuwen[1], A. Demirkan[1,2], N. Amin[1], A. Verhoeven[3], A. Meissner[3], R.W.W. Brouwer[4], W.F.J. van Ijcken[4], A. Isaacs[1], T. Hankemeier[5,6], K. Willems van Dijk[2,7], C.M. van Duijn[1].* 1) Epidemiology, Erasmus MC, Rotterdam, Zuid-Holland, Netherlands; 2) Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands; 3) Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, the Netherlands; 4) Center for Biomics, Erasmus MC, Rotterdam, the Netherlands; 5) Leiden Academic Center for Drug Research, Division of Analytical Biosciences, Leiden University, Leiden, The Netherlands; 6) The Netherlands Metabolomics Centre, Leiden University, Leiden, the Netherlands; 7) Department of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands.

Finding rare variants implicated in complex diseases has proven to be difficult. Large family-based studies in isolated populations yield increased power for identifying rare coding variants among others due to underlying founder effects. We explored this in the ERF study, a genetic isolate from the Netherlands. We tested if rare variants coding for high density lipoprotein cholesterol (HDL-C) also influence other metabolic traits which are also known to be under strong genetic control (30-80%). Participants of the ERF study include 3000 related individuals selected on the basis of genealogy. Exome sequencing was conducted in 1300 individuals. HDL-C was determined by the enzymatic method, whereas 400 metabolites (glycerophospholipids, triglycerides, amino-acids, acyl-carnitines, small metabolic compounds) were determined by NMR or MS. We tested around 550,000 coding variants within the exomes of 1300 samples for associations with the HDL-C level, followed by segregation analysis in families. We found both common (minor allele frequency (MAF) > 0.1) and rare variants (MAF < 0.02) to be associated with HDL-C. The common variants are all located within the CETP region, a region which is known to be associated with HDL-C level. We found nine rare variants to be significant (p-value < $2.57 \bullet 10^{-6}$, Bonferroni corrected for the number of unique genes tested) after adjusting for multiple testing and family structure. Each of the rare variants segregated within families. Of note is that all carriers of the rare variants have a (extremely) high HDL-C, and thus the rare variants may be of interest for therapy development. Each rare variant was also associated with several metabolites (p-value < $3.09 \bullet 10^{-6}$) which may help us understand the mechanisms involved. This study shows that combining next-generation sequencing with metabolomics within large family studies can help us unravelling the process from variant into biological processes influencing a clinical measurement. The fact that we found variants associated with high HDL-C levels opens up opportunities to translational research for medication.

1840M

Efficiently Incorporating Annotation Information into the Analysis of Genomic Sequence Variants in Pedigree Samples. *Q. Li, J. Bailey-Wilson.* NHGRI/NIH, 333 Cassell Dr. Suite 1200, Baltimore, MD.

Family-based analyses are an effective way to detect disease predisposing variants, especially in rare variants. However, when analyzing sequencing data, a long list of plausible variants that may account for disease susceptibility are usually identified and further filtering is needed to narrow down the list to a more manageable handful of variants for follow-up. Incorporating biological information about genetic markers in the analysis and interpretation of results has been shown to be a useful and successful approach. Here we propose a streamlined R process to analyze sequence data in pedigree samples for quantitative traits based on famSKAT software and variant annotation files from the Combined Annotation Dependent Depletion (CADD) server [1]. The famSKAT program is designed to analyze sequence data in family samples for quantitative traits. It has the advantage of using biological information as weights to adjust the statistical evidence for the variants. To get the weights, we extracted the CADD scores from the annotation files on the CADD server, because the scores measure the deleteriousness of single nucleotide mutations as well as insertion/deletion mutations in the human genome.

Due to the large amount of data produced by whole genome sequence (WGS) and whole exome sequence (WES) studies and the long processing time to conducting famSKAT tests in large studies, the analysis process is likely to be terminated prematurely due to technical problems even on fully redundant servers and clusters. Therefore, it is important to implement a recapture mechanism in our program, to find the endpoint of an aborted analysis session and restart the analysis from that point without redundantly repeating work already completed. In our implemented R programs, we periodically write out analysis results and enable the program to restart after the last completed analysis.

We tested our computer program on a simulated Mini-Exome dataset for extended families and show the performance in terms of type-I error and power. We also recorded metrics to assess the processing time for WES data analysis.

[1] http://cadd.gs.washington.edu/.

1841T

Efficient gene-gene interaction test for discordant sib pairs in genome-wide association studies. *R. Chung[1], P. Song[2], Y. Wang[2], P. Lin[2], W. Tsai[1].* 1) National Health Research Institu, Zhunan, Taiwan; 2) National Tsing-Hua University, Hsinchu, Taiwan.

Genome-wide association studies (GWAS) have been a popular strategy to identify single nucleotide polymorphisms (SNPs) associated with complex diseases. Gene-gene interactions may play an important role in complex diseases. Therefore, many methods have been developed for gene-gene interactions for GWAS. However, most of the methods are developed for case-control studies, and only a few methods are available for family-based interaction analysis for complex diseases. Moreover, current family-based interaction methods are computationally intensive, which are not applicable to genome-wide interaction studies, which test all possible pairs of SNPs across the genome. We propose an efficient family-based gene-gene interaction test, which compares the difference in log odds ratios for a pair of SNPs between cases and controls. Extended from the sib TDT (S-TDT), cases and controls are defined as genotypes in affected and unaffected sibs in discordant sib pairs (DSP), respectively. A multivariate hypergeometric distribution is used to calculate the variance and covariance for the test statistic. We used simulations to demonstrate that the proposed test has correct type I error rates under different scenarios, such as different sample sizes and minor allele frequencies. We also performed power studies to evaluate the power for the proposed test with other family-based interaction methods under different scenarios. The results suggested that the proposed test is a valid test and has comparable power to the existing tests. We also used simulations to show that the proposed test can run 20 times faster than the regression-based interaction test. Finally, we applied the test to a family GWAS study for hypertension and several promising SNP-SNP interactions were identified for the disease.

1842S

Robust and powerful family test for rare variant association. *K. Lin[1], S. Zöllner[1,2].* 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Psychiatry, University of Michigan, Ann Arbor, MI.

Modern sequencing technology allows investigating the impact of rare variant on complex disease. However, in a conventional case-control design, large samples are needed to achieve sufficient power for testing association of suspected loci to complex disease. In such large samples, population structure can cause spurious signals. One approach to overcome low power and population stratification is family-based study design since the count of risk variants can be enriched in a family containing multiple affected members with the same genetic background. Traditional methods such as the TDT and the FBAT are not very powerful when applied to rare variants. Thus, there is a need to develop a method to directly incorporate the advantage of family-based design to discover association between rare variants and disease phenotypes.

We propose a novel test for a sample of families. In each family, we first determine the chromosomes shared IBD/non-IBD at the locus of interest for all family members. Conditional on the sharing status, we evaluate the number of risk alleles shared IBD among affected family members. Since risk variants are more likely to be shared among affected members, we test for an excess of shared rare variants among affected members by comparing the observed count of shared variants to the expected count of shared variants under the null. This expectation can be calculated conditional on the founders' haplotypes. The proposed test is robust to population stratification since each family is a matched unit. We evaluate the power of this approach analytically as well as with computer simulations using a general model for the effect size of rare risk variants. We compare our method with existing methods for family-based data as well as the conventional case-control design. Considering different models from rare to common variants, we show that the proposed approach is especially powerful for rare variants as compared to the conventional case-control design.

1843M

Family-based rare variant association methods for quantitative traits in the presence of population structure. *W-M. Chen, A. Manichaikul, S.S. Rich.* Ctr Pub Hlth Genomics, Univ Virginia, Charlottesville, VA.

Technological advances in high-throughput sequencing platforms have made it possible to extend genome-wide association studies to rare variants by exome, whole-genome, and targeted sequencing. Custom genotyping arrays such as the ExomeChip have been utilized for their low cost in genotyping rare variants. Family data can enrich rare variant association analyses, and allow the use of well characterized pedigree data that were collected previously in linkage and association studies. Population structure (such as in admixed populations) presents great challenges to statistical methods that analyze datasets with rare variants. The standard approach of including principal components of ancestry as covariates in the regression of a trait may not be sufficient to eliminate the effect of population structure on the rare variant association.

We extend the sequence kernel association test (SKAT) to family studies, and present several novel kernels for rare variant association in the presence of population structure: 1) weighted linear kernel (to confirm the existing results from SKAT); 2) distance kernel that adjusts pedigree structure; 3) distance kernel that adjusts pedigree and global population structure; and 4) distance kernel that adjusts pedigree and local population structure. In addition to including principal components of ancestry as covariates in the linear regression model, we standardize the distance between any pair of individuals at a gene by dividing their global or local distance across the genome. Extensive computer simulations show that our distance kernels can effectively eliminate population stratification as well as increase the power of association test. Finally we apply our methods to a data set of rare variants genotyped in African American and Hispanic populations.

1844T

TIMBER - personalized computationally-efficient filtering of GERMLINE-discovered putative IBD segments. *M. Barber[1], K. Noto[1], Y. Wang[1], R.E. Curtis[2], J.M. Granka[1], J.K. Byrnes[1], N.M. Myres[2], C.A. Ball[1], K.G. Chahine[2].* 1) AncestryDNA, San Francisco, CA; 2) AncestryDNA, Provo, UT.

Discovering regions of pairs of genomes that are identical by descent (IBD) is an important part of many genetic analyses. Given a large sample size (over 100K people and beyond), the problem is for IBD discovery to be both computationally feasible and accurate. GERMLINE is a computationally efficient algorithm (as it uses a hashing approach and a separate extension algorithm) but it lacks the accuracy of other approaches on its own. RefinedIBD is an accurate algorithm, as it evaluates all putative IBD segments with a haplotype model. However, RefinedIBD is not computationally efficient, even though it uses the GERMLINE algorithm to discover the putative IBD segments. TIMBER is an algorithm that decides if a putative IBD segment discovered by GERMLINE has a sufficient level of evidence for IBD to be retained. TIMBER uses GERMLINE-discovered putative IBD segments to filter the very same segments. Within GERMLINE, the genome is split up into non-overlapping windows, where TIMBER calculates a weight for each of these windows. Each weight provides the relative level of evidence of IBD from that window in a GERMLINE-discovered putative IBD segment. TIMBER's main action is to down-weight windows of the genome that show excessively high degree of putative IBD segments. TIMBER is made possible given the large amount of putative IBD segments that are discovered in running GERMLINE on a large data set (over 100K people). We have evidence that TIMBER is a very useful IBD filter from both real and simulated data given a large data set of over 100K people. While TIMBER is not as accurate as a method such as RefinedIBD, it provides a significant improvement in accuracy over running GERMLINE on its own and it is computationally efficient for a large data set.

1845S

Statistical Tests for Co-Segregation of Genetic Variants with Disease in Pedigrees. *D. Schaid[1], S. McDonnell[1], J. Sinnwell[1], L. Cannon Albright[2], C. Teerlink[2], J. Stanford[3], E. Ostrander[4], W. Isaacs[5], J. Xu[6], K. Cooney[7], E. Lange[8], J. Schleutker[9], J. Carpten[10], I. Powell[11], J. Bailey-Wilson[12], O. Cussenot[13], G. Cancel-Tassin[13], G. Giles[14], L. FitzGerald[14], C. Maier[15], A. Whittemore[16], CL. Hsieh[17], F. Wiklund[18], W. Catolona[19], W. Foulkes[20], D. Mandal[21], R. Eeles[22], Z. Kote-Jarai[22], S. Thibodeau[23].* 1) Dept Hlth Sci Res, Mayo Clinic, Rochester, MN; 2) Dept. Internal Medicine, Univ Utah School of Medicine, Salt Lake City, Utah; 3) Fred Hutchinson Cancer Research Center, Seattle, WA; 4) National Human Genome Research Institute, Besthesda, MD; 5) Johns Hopkins Hospital, Department of Urology, Baltimore, MD; 6) Center for Human Genetics, Wake Forest University, Winston-Salem, NC; 7) Departments of Internal Medicine and Urology, University of Michigan Medical School, Ann Arbor, MI; 8) Department of Genetics, University of North Carolina, Chapel Hill, NC; 9) Medical Biochemistry and Genetics, University of Turku, Finland; 10) Integrated Cancer Genomics Division, The Translational Genomics Research Institute, Phoenix, AZ; 11) Wayne State University, Detroit, MI; 12) Statistical Genetics Section, National Human Genome Research Institute, Bethesda, MD; 13) Service Urology-Batiment Gabriel, Hopital Tenon, Paris, France; 14) Cancer Epidemiology Centre, Cancer Council Victoria, Austrailia; 15) Department of Urology, University of Ulm, Ulm, Germany; 16) Dept. Health Research and Policy, Stanford Univ, Stanford, CA; 17) Department of Urology , University of Southern California, Los Angeles, CA; 18) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 19) Northwestern University Feinberg School of Medicine, Chicago, IL; 20) Depts. Of Oncology and Human Genetics, Montreal General Hospital, Montreal QC, Canada; 21) Department of Genetics, LSU Health Sciences Center, New Orleans, LA; 22) Genetics and Epidemiology, Institute of Cancer Research, Sutton Surrey, UK; 23) Department of lab Medicine/Pathology, Mayo Clinic, Rochester, MN.

Sequencing studies for rare variants in pedigrees often evaluate whether affected pedigree members carry the same variant - an evaluation of co-segregation of variant with disease. This approach, however, requires conditional probabilities, because pedigrees without any carriers are not informative. That is, statistical assessment of co-segregation requires comparing the observed number of affected carriers with random expectation, where expectation is based on Mendelian segregation of variants, conditional on at least one affected pedigree member carrying a variant. We developed new statistical methods to test for co-segregation of genetic variants with disease, for studies that genotype only affected pedigree members. Our methods, a simple comparison of observed and expected carrier counts, as well as methods based on quasi-likelihood score (QLS) statistics, provide was to perform gene-level tests, important for situations where different pedigrees have different segregating variants, yet all within the same gene. Our gene-level test allows use of weights for different variants, such as weights based on allele frequencies or based on likely function. Furthermore, our QLS method accounts for correlations in the data, such as correlations from related subjects, or correlation among multiple variants (e.g., linkage disequilibrium).We will present simulations to illustrate the power of our approach, as well as results from application to a whole exome scan of familial prostate cancer.

1846M

Cross pedigree shared ancestry reveals rare, disease-causing variants in the presence of locus heterogeneity. *H.J. Abel, MA. Province.* Genetics, Washington University School of Medicine, St. Louis, MO.

Currently, there is great interest in the use of family studies to identify rare variants underlying complex disease. Attempts at fine mapping, however, are often confounded by locus heterogeneity, which can produce noisy and poorly localized linkage signals, as well as by an abundance of rare variants, which frequently segregate with phenotype by chance. Because rare variants shared across pedigrees are likely to be of recent origin, we have developed an approach to leverage identity-by-descent (IBD) sharing between pedigree founders in order to better localize linkage signals in the presence of heterogeneity. Our method relies on the lengths of segments shared identically-by-state (IBS) across pedigrees, using as test statistic a score based on the sum of maximal pairwise shared lengths at each locus, and optimized over all genotyped pedigree members. Use of unphased IBS renders it both computationally efficient, so that pedigree-based permutation tests assessing significance are tractable, and robust to genotyping/sequencing and haplotype phase switch errors. Further, our approach produces a cross-family metric to facilitate local clustering of families near IBD regions which, in simulations, accurately recovers ancestral relationships and permits stratification by recent shared ancestry. We have evaluated the performance of our method using a combination of coalescent simulation of founder individuals, followed by gene-dropping onto pedigrees. Under a variety of scenarios, with rare causal variants (population MAF<0.01) and modest effect sizes (OR=5-10), our approach achieves 60-90% power, and is able to detect shared ancestral segments harboring rare causal variants when multipoint linkage and rare-variant burden tests fail.

1847T

Increasing Power to Detect Rare Variant Associations by Integrating Linkage Data: A Bayesian Approach. *s. Lutz, A.L. Peljto, T.E. Fingerlin.* Biostatistics, University of Colorado, Denver, CO.

Large resequencing studies often combine familial cases of disease from linkage studies with sporadic cases to achieve the large sample sizes necessary for detecting low-frequency alleles. In general, these studies include only a single case from each family. As a result, family-based association tests cannot be applied and standard tests for rare variants generally ignore family-specific allele-sharing information that may be available, resulting in an inefficient use of resources and loss of power for this common study design. We propose a Bayesian approach for testing associations between groups of variants and a phenotype that uses family-specific linkage information in the prior to increase power to detect associations with rare variants and the phenotype of interest. Our method directly combines family-specific linkage and rare variant data by giving more weight in the prior to familial cases with evidence for excess allele sharing at the position of the variant being tested. To assess the performance of our method, we simulated genetic data for familial cases, singleton cases, and controls under various genetic models. We then compared our Bayesian approach that incorporates family-specific linkage information to the sequence kernel association test (SKAT) and several burden tests. We show that our method has increased power compared to the standard methods (SKAT and burden tests), which do not incorporate the family-specific allele-sharing information for several underlying genetic models. Our method provides a framework for tests of association with rare variants that are more statistically efficient when family-specific linkage information is available on at least a subset of cases, while leveraging the resources that have been devoted to recruiting, describing, and genotyping familial linkage cohorts.

1848S

Impact of screening for precancerous lesions on family-based genetic association tests: an example of colorectal polyps and cancer. *S.L. Stenzel[1], J.C. Figueiredo[1], V.K. Cortessis[1,2], D.C. Thomas[1].* 1) Dept of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 2) Dept of Obstetrics and Gynecology, Keck School of Medicine, University of Southern California, Los Angeles, CA.

Estimates of the association between genetic variants and cancer risk may be biased by screening and subsequent removal of detected precancerous lesions. We conducted a simulation study using a discordant sibship case-control design assuming various screening and prevention parameters and determined the association between genetic variation and risk of colorectal cancer (CRC) under four analysis strategies: 1) CRC cases vs. all non-cases with no adjustment for polyp history; 2) same as (1) with adjustment; 3) CRC vs. polyp-free non-cases; 4) CRC or polyp cases vs. neither. We considered genes involved in polyp development (G1), in cancer risk but not polyps (G2), and in progression from polyps to carcinoma (G3). We then applied these strategies to a family-based case-control study of folate-related genes and risk of CRC. In our simulation, estimates of the association between G2 or G3 and CRC comparing CRC cases against all non-cases without adjustment for polyp history were not biased, but estimates were underestimated for G1 if detection of a polyp led to reduction in CRC risk. Analysis strategies 2 through 4 led to substantial biases, with directions varying across strategy and gene type. Type I errors were correct for all approaches, but strategy 1 was the most powerful. In our application, estimates of relative risk differed for selected variants when controls with polyps were excluded or when individuals with polyps were treated as cases. Although estimates of gene-CRC associations may be biased due to detection of precancerous lesions, the degree of bias appears to differ only modestly by analytic strategy in applied studies with population-based recruitment. The impact may be more unpredictable in studies that are not population-based.

1849M

TITLE: Familial aggregation of blood pressure in Ramadasia population of north-west Punjab. *R. KUMAR[1], B. DOZA[2].* 1) GHS MARI KAMBO KE, TARNTARAN; 2) GURU NANAK DEV UNIVERISITY, AMRITSAR.

In India, the burden of hypertension has increased many folds in recent times due to increase of westernized diets, life styles and the increasing mean age of populations. Therefore, it is essential to understand how blood pressure is influenced by familial factors (both genetic and environmental) and how these contribute to the risk for cardiovascular diseases. The major objective of the study was to describe the genetic heritability and familial household contribution to the phenotypic variation for cardiovascular risk factors especially for SBP and DBP. Three generations i.e. grandparental, parental and offspring generation, aged 7 years and above, from Ramadasia, a scheduled caste population, were selected for the study. The questionnaire grossly included socio-economic lifestyle variables, physiometric measurements and anthropometric measurements. A total of 600 families with 1827 individuals were sampled. Results demonstrated familial aggregation of CVD risk factors. The brother-sister, father-offspring (male/female), mother-offspring (male/female) correlations were all greater than spouse correlations (brother-sister: 34% for SBP, 31% for DBP, 38% for BMI; father-male offspring: 31% for SBP, 30% for DBP, 21% for BMI; father-female offspring: 21% for SBP, 20% for DBP, 15% for BMI; mother-male offspring: 17% for SBP, 15% for DBP, 23% for BMI; female offspring: 20% for SBP, 18% for DBP, 29% for BMI; spouse: 4% for SBP, 5% for DBP, 9% for BMI).. The magnitudes of these correlations for SBP, DBP and BMI were lower in grandparent-offspring. Therefore, the higher estimates of correlations among parent-offspring generations and siblings suggested the genetic closeness. But the correlation between siblings was even larger than correlation between parent-offspring and others. Familial aggregation of blood pressure was observed to be largely due to genes rather than familial environment from the heritability explained by the genetic variation in blood pressure. Therefore, the present study has a great potential for the study of quantitative genetic variations within and between the generations. The estimates of heritabilities through variance-component approach produced a wide range of heritabilities (10%-98.7%) for selected anthropometric and physiometric phenotypes between different combinations of the three generations.

1850T

Mixed Model Association Mapping on the X Chromosome. *C. McHugh, T. Thornton.* Department of Biostatistics, Univ Washington, Seattle, WA.

Genetic variants on the X chromosome could potentially play an important role in some complex traits. However, statistical methods for genetic association studies have primarily been developed for variants on the autosomal chromosomes with significantly less attention given to the X chromosome. Mixed linear models (MLMs) have recently emerged as a powerful and effective method of choice for association mapping in the presence of sample structure such as population structure and/or relatedness. Existing MLM approaches are not directly applicable to association testing on the X chromosome. Here, we propose a MLM approach for genetic association testing on the X chromosome in samples with related individuals. In simulation studies with both unrelated and related individuals, we demonstrate that our proposed MLM has the correct type I error for association testing with SNPs on the X chromosome. We also show that existing mixed model approaches are not properly calibrated and have inflated type I error for X-linked traits. We further demonstrate that our proposed MLM method for the X has high power to detect genetic association with X-linked markers.

1851S

lincRNA-mRNA transcriptional regulatory co-expression network from RNA-seq data in 624 Sardinian individuals. *P. Forabosco[1], M. Pala[1,2,6], M. Marongiu[1], A. Mulas[1], R. Cusano[1], F. Crobu[1], F. Reinier[3], R. Berutti[3,6], M.G. Piras[1], C. Jones[3], D. Schlessinger[5], G. Abecasis[4], A. Angius[1], S. Sanna[1], S.B. Montgomery[2,7], F. Cucca[1,6,7].* 1) Research Institute of Genetics and Biomedicine, CNR, Cagliari, Italy; 2) Pathology and Genetics Department, Stanford University, Stanford, CA, USA; 3) CRS4, Advanced Genomic Computing Technology, Pula, Italy; 4) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 5) Laboratory of Genetics, National Institute on Aging, Baltimore, MD, USA; 6) Department of Biomedical Science, University of Sassari, Sassari, Italy; 7) co-senior authors.

Long non-coding intergenic RNAs (lincRNAs) are thought to play important roles in regulating gene expression, but the exact function of the majority of them remains unknown. RNA-seq represents an effective tool for high-throughput analysis of lincRNA expression, as it provides a far more precise and accurate measurement of transcript levels than other methods. Bioinformatics approaches might be useful in providing insight into lincRNA functions. In particular, co-expression network analysis of lincRNA and mRNA transcripts enables a systematic and global interpretation of expression data, by assessing co-expression patterns (through correlation matrices), and by identifying functionally relevant modules. Annotation of lincRNAs probable functions can be predicted and linked to biological pathways based on module sharing, i.e. by assigning functions according to the functional enrichment of coding transcripts in the same module. Analysis of co-expression of lincRNAs with mRNA can therefore help prediction of their functional role as a foundation for further mechanism studies. Here we present the construction of a regulatory network comprising both coding and long non-coding transcripts. Gene expression levels have been quantified with RNA-seq from PolyA(+) PBMC samples in 624 Sardinian individuals. We construct a co-expression network, where nodes are transcripts and edges represent co-expression strength, using the Weighted Gene Co-expression Network Analysis package (WGCNA) implemented in R. We inspect the network in detail in terms of connectivity and density patterns, network hubs, and module identification. Compared to microarray technologies, RNA-seq data results in more interconnections among transcripts (e.g. higher mean connectivity is observed in a subsample of 188 unrelated subjects when using the suggested soft thresholding power in the adjacency transformation in order to meet scale free topology). Using gene significance (a measure correlated with intramodular connectivity) encoding lincRNA status, we identify modules where lincRNAs are most relevant. This study is the first to explore genome-wide lincRNA expression and co-expression with mRNAs using RNA-seq technology.

1852M

Sample-specific gene co-expression networks controlling for confounding effects. *C. Gao[1], S. Zhao[2], B.E. Engelhardt[1,3].* 1) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC; 2) Department of Computational Biology and Bioinformatics, Durham, NC, USA; 3) Department of Statistical Science, Duke University, Durham, NC, USA.

It is well known that genes form sparse sets of co-regulated modules to achieve unique biological functions. Biological complexity suggests that there could be thousands of co-regulated modules interacting simultaneously in a single cell and therefore extracting a large number of small co-regulated clusters of genes is essential. Two types of methods have been proposed to address this challenge. First, gene co-expression networks identify regulatory interactions across all genes based on correlations of those genes. These methods are generally sensitive to confounding effects and the parameter settings. Sparse latent factor models, on the other hand, identify a small number of unstructured clusters of co-regulated genes and may account for confounding effects; however, due to computational tractability and signal recovery accuracy, sparse latent factor models only focus on small numbers of clusters, and, to our knowledge, have not been used to identify networks for individual genes. To address these discrepancies, we developed a Bayesian statistical model to recover large number of small clusters of co-regulated genes from gene expression data. We use this framework to cluster both the genes and the samples in order to allow heterogeneous samples with subtype-specific gene interactions; furthermore, we used the recovered sparse gene sets to reconstruct biologically meaningful gene co-expression networks that control for confounders and are specific to sample subtypes. Our method uses Bayesian sparse factor analysis, with the gene expression matrix is decomposed into a sparse loading matrix and a corresponding factor matrix, where the genes loaded on the same factor have the interpretation of being jointly regulated by the same covariate. To induce sparsity, we used a flexible Three Parameter Beta (TPB) prior that scales with the sparsity in the data. To address the latent biological and technical confounders, we modeled each factor and loading as a mixture of sparse and dense components, where dense vectors capture technical or biological confounding when estimating the sparse co-regulated gene sets. We used simulated data to quantify and compare our method relative to other biclustering methods. We analyzed gene expression data from the GTEx project, extracting large numbers of sparse groups of co-regulated genes. We discuss the implications of the recovered gene networks in revealing important biological, and potentially disease-related, mechanisms.

1853T

Random forest for genetic analysis: Integrating the X chromosome. *G. Jenkins, J. Biernacka, S. Winham.* Mayo Clinic, Rochester, MN.

Despite the fact that it is known to play a major role in many Mendelian disorders, the X chromosome is routinely excluded from genome-wide association studies. Data mining methods such as Random Forest (RF) have been proposed to investigate complex genetic models involving many variants. However, for traits associated with sex, inclusion of X chromosome SNPs in RF analysis yields biased results, as we illustrate using simulations and an example from a case-control study of alcohol dependence. We propose three extensions of the RF algorithm to include X chromosome SNPs, based on (1) the principle of X chromosome inactivation (XCI), (2) stratification of the forest by sex, and (3) incorporation of sex as a variable in the RF. We compare the performance of these three new approaches to the traditional RF implementation using simulations and application to data from the Study of Addiction: Genes and Environment (SAGE). Comparison of the SAGE analysis results for autosomal vs. X SNPs shows that traditional RF consistently ranks the X SNPs highly, whereas the three new approaches all rank the X SNPs similarly to the autosomes. To better understand statistical properties of the alternative approaches to incorporate X-chromosome data into RF, we compare the importance measures for X chromosome and autosomal SNPs produced by the different methods, using data simulated with and without genetic effects. Specifically, we examine performance using data simulated under varying degrees of trait or sex imbalance, effect of sex on the trait, and patterns of linkage disequilibrium. Our results show that all methods rank X SNPs similarly to the autosomes if sex was not associated with the trait (with no genetic effect). But if sex is associated with the trait, traditional RF leads to inflated variable importance for the X chromosome. Incorporating sex data in RF produces biased X importance across scenarios. However, the methods based on XCI and stratified forests do not inflate the importance of the X chromosome, although X importance was slightly lower than the autosomes if the number of cases and controls are not equal. If sex is not associated with the trait in the sample, regular RF may be appropriate to analyze X chromosome data. If sex is associated, either stratification of the forest or extension based on XCI will not inflate the importance of the X chromosome, although these methods may underestimate X importance if the data is unbalanced based on the trait.

1854S

Integrative Metabolomics of Asthma Severity using Bayesian Networks. *J. Lasky-Su[1], M. McGeachie[1], W. Qiu[1], J. Savage[1], A. Dahlin[1], A. AlGarawi[1], D. Croteau-Chonka[1], J. Sordillo[1], A. Wu[1], E. Chen[1], D. DeMeo[1], A. Litonjua[1], C. Clish[2], B. Raby[1], S. Weiss[1].* 1) Brigham & Women's Hosp, Boston, MA; 2) Broad Institute, Cambridge, MA.

Asthma is a heritable disease with both environmental and genetic components. Although a number of molecular determinants have been identified, much remains to be understood about how these variants impact the disease. Metabolites have the distinct advantage of being more proximal markers of disease processes than transcriptional, translational or post-translational changes. Therefore the integration of metabolomics, with genetics, genomics, and epigenetics may more effectively link disease outcomes with genetic determinants. We generated lipidomic pilot data using liquid chromatography tandem mass spectrometry (LC-MS) using plasma samples from 20 Caucasian individuals from the Asthma BioRepository for Integrative Genomic Exploration (Asthma BRIDGE), an open-access repository from 1,435 asthmatics. A total 64 metabolites were generated of which 26 were identifiable. We used two asthma severity phenotypes, a binary measurement, “hospitalized overnight for asthma” and a continuous, composite measurement, “acute asthma control.” We performed the following analyses for all 64 metabolites: 1) GWAS analyses for each metabolite and all genetic variants; 2) regression analyses for each metabolite and genome-wide gene expression variants; 3) regression analyses for each metabolite and the genome-wide CpG methylation data. We then identified the top variants from each of these analyses including the top SNPs, mRNA transcripts, and CpG sites that were associated with any metabolites. We then applied a Conditional Gaussian Bayesian Network analysis to the top findings from these analyses, using the asthma severity phenotypes as the outcome of interest. We first used 5-fold cross-validation to arrive at hyper-parameters for the Bayesian likelihood calculations. We then performed 100 bootstrap iterations of the dataset, and learned an exhaustive network for each bootstrap iteration. From the sample of 100 networks, we built a consensus network by including the edges present in more than 25% of the bootstrap networks. From this, we developed networks with predictive accuracy as high as 98% for asthma severity phenotypes. This approach illustrates the strength of integrating metabolomics data with genetic, gene expression, and epigenetic data for disease prediction.

1855M

Closed-form Wald tests for genome-wide analysis of gene-gene interactions. *Z. Yu[1], M. Demetriou[2], D. Gillen[1].* 1) Department of Statistics, University of California, Irvine, CA; 2) Department of Neurology, University of California, Irvine, CA.

Despite the successful discovery of hundreds of variants for complex human traits using genome-wide association studies, the degree to which genes jointly affect disease risk is largely unknown. One obstacle toward this goal is that the computational effort required for testing gene-gene interactions is enormous. As a result, numerous computationally efficient tests were recently proposed. However, the validity of these methods often relies on unrealistic assumptions such as additive main effects, main effects at only one SNP, and no linkage disequilibrium. Here we propose to use closed-form Wald tests. The Wald tests are asymptotically equivalent to the corresponding likelihood ratio tests, largely considered to be the gold standard tests but generally too computationally demanding for genome-wide interaction analysis. Simulation studies show that the Wald tests have very similar performance with their computationally intensive counterparts. Applying the proposed tests to a genome-wide study of multiple sclerosis, we identify interactions within the major histocompatibility complex region. In this application, we found that (1) focusing on pairs where both single nucleotide polymorphisms (SNPs) are marginally significant leads to more significant interactions when compared to focusing on pairs where at least one SNP is marginally significant; and (2) parsimonious parameterization of interaction effects might decrease, rather than increase, statistical power.

1856T

Gene based analyses of sympathetic nervous system genes on long term blood pressure: The GenSalt study. *C. Li[1], J. He[1,3], J.E. Hixson[4], D. Gu[2], D.C. Rao[5], L.C. Shimmin[4], J. Huang[2], C.C. Gu[5], J. Chen[2], F. Liu[2], J. Li[2], T.N. Kelly[1].* 1) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 2) State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; 3) Department of Medicine, Tulane University School of Medicine, New Orleans, LA; 4) Department of Epidemiology, Human Genetics and Environmental Sciences, University of Texas School of Public Health, Houston, TX; 5) Division of Biostatistics, Washington University School of Medicine, St. Louis, MO.

Although genomic factors are known to influence BP, much of the heritability of the BP phenotype remains unexplained. Gene-based analysis of long term BP may increase power to detect genetic factors underlying BP phenotypes. The objective of the current study was to examine the influence of sympathetic nervous system (SNS) genes and gene-gene interactions on long term average BP among 1,881 Han Chinese participants of the family-based Genetic Epidemiology Network of Salt Sensitivity (GenSalt) follow-up study. Nine BP measures were collected at a 2003 baseline examination and at each of two follow-up visits in 2007 and 2011. Single marker associations of 448 tag SNPs in 20 SNS genes and the pairwise interactions of SNPs between each pair of genes were assessed for association with long term BP using mixed models to accommodate the repeated BP measures and GenSalt family structure. For single gene analysis, single-marker analysis *P*-values for SNPs within each gene were combined using the truncated product method (TPM) to estimate an overall gene-based *P*-value. TPM was also used to combine *P*-values for all SNP interaction terms within each gene pair to generate a gene-gene interaction *P*-value. Nominally significant findings from GenSalt were evaluated for replication using publicly available data from the Multi-Ethnic Study of Atherosclerosis (MESA). Findings from both studies were combined using meta-analysis. Statistical significance was determined by nominal significance in both GenSalt and MESA studies, along with significance after Bonferroni correction to meta-analysis results (α-theshold=0.0025 and 0.00026 for single-gene and gene-gene interaction analyses, respectively). In single-gene analysis, *CACNA1C* was associated with long term systolic BP (P=0.0014), while *ADRA1D* was associated with long term diastolic BP (P=0.0004). In addition, multiple gene-gene interactions influencing BP phenotypes were identified, including: *CACNA1A*×*ADRA1A*, *CACNA1A*×*CACNA1C*, *CACNA1C*×*CACNA1B*, *CACNA1C*×*ADRA1A*, *CACNA1C*×*ADRA1D*, *CACNA1D*×*ADRA1D*, *CACNA1D*×*SLC6A2* for both systolic and diastolic BP (all $P<1.0\times10^{-4}$); *CACNA1A*×*ADRA1D*, *CACNA1C*×*ADRB2*, *ADRA1A*×*ADRA1B*, *ADRA1A*×*SLC6A2* for systolic BP (all $P<1.0\times10^{-4}$); and *CACNA1A*×*CACNA1D*, *CACNA1A*×*SLC6A2*, *CACNA1C*×*SLC18A1*, *ADRA1D*×*ADRA1A*, *ADRA1D*×*DBH*, *ADRA1D*×*SLC18A2* for diastolic BP (all $P<1.0\times10^{-4}$); and These findings provide strong evidence for influence of SNS genes and gene-gene interaction.

1857S

Rapid Variance Component Aggregation Test (RVCAT) for evaluating interaction effects of rare-variants. *R. Marceau[1], W. Lu[1], F.-C. Hsu[2], J.-Y. Tzeng[1,3].* 1) Department of Statistics, North Carolina State University, Raleigh, NC; 2) Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest University Health Sciences, Winston-Salem, NC; 3) Bioinformatics Research Center, North Carolina State University, Raleigh, NC.

Studying gene-environment (GxE) and gene-gene (GxG) interactions in genetic association studies allows us to better understand complex disease etiology, uncover missing heritability, and opens the door for personalized medicine. Marker-set aggregation approaches have been shown to be powerful alternatives to the more traditional single SNP GxE tests, especially for rare variants. There are two categories of aggregation approaches that are commonly used: fixed effects methods, which collapse at the genotype level (e.g., the burden test), and random effects methods, which collapse at the similarity level (e.g., kernel machine regression and similarity regression). For assessing interaction effects, random effects models have found to be more robust to model misspecifications than fixed effect approaches. That is, these approaches are still valid when the genetic main effect term and/or the GxE term are misspecified. However, because random effects models use variance components (VCs) to capture the effects, such robustness comes at the cost of increased computational burden from having to estimate nuisance VCs for each effect when assessing GxE effect. These nuisance VCs are computationally intensive to calculate when the trait of interest is non-quantitative. In this work, we propose a rapid variance component aggregation test (RVCAT) which sidesteps the need to estimate the nuisance VCs by treating the genetic main effect as a fixed effect term. We show that while RVCAT maintains the validity and power of the traditional VC tests, it is also computationally efficient—for a quantitative trait analysis with a sample of 5000 subjects, it was over 6.5 times faster than the fully random effect model. Moreover, the proposed algorithm can be conducted on any current software package that is designed for main effect analysis (i.e., no estimation of nuisance VCs). We demonstrate the utility of our proposed method to quantitative, survival, and binary traits using extensive simulations, and apply our method to the Vitamin Intervention for Stroke Prevention (VISP) trial to better understand the effect of genotype on recurrent stroke and its risk factors.

1858M

Studying the elusive exposome and its interaction with the genome in large-scale. *C.J. Patel[1], J.P.A. Ioannidis[2].* 1) Center for Biomedical Informatics, Harvard Medical School, Boston, MA; 2) Stanford Center for Prevention Research, Stanford University School of Medicine, Stanford, CA.

It is hypothesized that greater than 50% of complex disease risk is attributed to differences in an individual's environment, but we lack ways of investigating the exposome -- the totality of exposure load that occurs throughout a lifetime - in disease risk. Further, we have yet to incorporate the exposome in genome-wide association studies (GWAS) to assess how environment modifies genetic risk for complex disease. Furthermore, investigating one or a handful of exposures at a time has led to a highly fragmented literature of epidemiologic association and much of that literature is not reproducible. A new model is required to discover environmental exposures, and how they interact with the genome, in disease.

Here, we discuss ways of to remedy this problem through a strategy known as the "Environment-wide association study" (EWAS), where investigators assess 100s-1000s of personal exposures simultaneously. Analogous to GWAS, we will show how multiple personal exposures can be assessed simultaneously in terms of their association with diseases such as type 2 diabetes (T2D), heart disease risk factors, and mortality in cohorts from the United States National Health Nutrition Examination Survey and the International Study of Macronutrients in Blood Pressure (N~5K-30K). In these studies, we show how an array of exposures ranging from pollutants, nutrients, and pesticides are associated with these diseases and have effect sizes that are comparable or exceed GWAS findings. We will discuss the hurdles, including biases such as reverse causality, confounding, and challenges of inferring independence in midst of the dense correlation structure of the exposome.

We outline and show preliminary findings that combine the exposome with genome to assess how genetic risk for disease may be dependent on environmental exposure, known as gene-environment interactions (GxE). Because searching for interactions requires prohibitively large sample sizes and hypothesis tests, we describe how to increase chances for discovery and replication by consideration of interactions between top EWAS and GWAS findings. Specifically, we describe a strategy that tests the interaction between a summary of genetic risk (via a validated polygenic risk score) in interaction with EWAS-derived exposures. We show examples of the strategy in complex traits including body mass index and blood pressure as well as a disease, T2D.

1859T

To evaluate the determinants of pre-hypertension and hypertension among Punjabi adolescent population using path analysis and structural equation modelling. *S.K. Brar, H. Kaur, B. Badaruddoza.* Deptt. of Human Genetics, Guru Nanak Dev University, Amritsar, India.

Punjabi ethnic population in India is at high risk for obesity and hypertension. It is believed that these disorders begin in childhood especially in adolescent period. Keeping in mind the complex nature of these disorders, a cross-sectional study with a sample of 3060 (1530 boys and 1530 girls) adolescents aged 10 to 18 years was carried out. The study has applied a path analysis of structural equation modeling for all studied anthropometric and socio-economic lifestyle variables to evaluate the determinants of pre-hypertension and hypertension. The study first tested the adequacy of hypothesized path model which contain the inter-correlated variables, such as body mass index (BMI), waist circumference (WC), waist to hip ratio (WHR), sum of skinfolds (SumSF), food habits (FH), exercise (EX), screen time (ScT) and family status (FS). Next, the study tested the relative strengths of association of independent variables with SBP and DBP on the basis of the magnitude of path coefficients of regression, covariances and variance. The study adopted maximum likelihood estimation, 95% confidence limit and critical ratio in generating the estimate of path coefficients and its significance. The regression path coefficients of BMI, waist circumference and WHR were found significant among both boys and girls. The estimates of variance and critical ratio for all the studied risk factors for SBP and DBP have been found significant on SBP and DBP for both boys and girls. The whole analysis confirmed that the waist circumference, BMI and family income status have significant contribution to increase blood pressure among Punjabi adolescents.

1860S

Study design and statistical tests for detecting gene-environment interaction on environmental exposure-defined phenotypes. *C. Chen[1], P. Kraft[2], B. Neale[1], J. Smoller[1,2].* 1) Massachusetts General Hospital, Boston, MA; 2) Harvard School of Public Health, Boston, MA.

There is an increasing interest in investigating gene-environment interaction for phenotypes with a prerequisite of certain environmental exposure (e.g. post-traumatic stress disorder, substance dependence). However, the nature of these phenotypes, which exclude the possibility that an individual becomes a case without the required exposure, poses unique challenges in choosing the optimal study design and statistical test for gene-environment interaction analysis. We explored the best strategy to detect gene-environment interaction for environmental exposure-defined phenotypes, where the exposure had four ordered levels (unexposed; and low, medium, high exposure). We compared the statistical power and Type 1 error of 4 different tests that are commonly used in this context: 1) gene main effect test (G); 2) conventional gene-environment interaction test (GE); 3) case-only gene-environment interaction test (CO); 4) joint test of both the gene main effect and the interaction (GGE). We considered factors that may affect the power and Type 1 error of these tests, including: genetic risk allele frequency; environmental exposure frequency; genotype-exposure correlation in the population; effect sizes of exposure, gene and gene-environment interaction; and exposure misclassification. Finally, we compared analyses for non-exposure-defined phenotype and exposure-defined phenotype with both exposed and non-exposed controls, or with exposed controls only. When only main effects were present with no interaction effect, the G test showed higher power than the GGE test. When main and interaction effects were present, the CO test showed higher power than the GE test for non-exposure-defined and exposure-defined phenotypes. We note that the power of GE test reduced dramatically when restricting to exposed controls for exposure-defined phenotypes. For non-exposure-defined phenotype, the GGE test showed higher power than the G test (2% increase in power). However, for exposure-defined phenotype, the G test showed higher power than the GGE test (12% increase in power). Restriction to exposed controls improved the power of the G test and the GGE test for exposure-defined phenotype. Based on these results, we recommend the CO test for investigating gene-environment interaction, with a caveat on inflated type 1 error due to G-E correlation, and restriction to the exposed controls for detecting genetic effects on exposure-defined phenotypes.

1861M

A simulation study of gene-by-environment interactions in GWAS implies ample hidden effects. *U.M. Marigorta, G. Gibson.* School of Biology, Georgia Institute of Technology, Atlanta, GA., USA.

The recent switch to a modern lifestyle has coincided with a rapid increase in prevalence of obesity and other diseases. Paradoxically, these conditions present large heritability values. The shifts in prevalence could be explained by the release of genetic susceptibility for disease in the form of gene-by-environment (GxE) interactions. However, the detection of interaction effects requires large sample sizes, little replication has been reported, and a few studies have demonstrated environmental effects only after summing the risk of GWAS alleles into genetic risk scores (GRSxE). We performed extensive simulations of a quantitative trait controlled by 2,500 common causal variants to inspect the feasibility to detect gene-by-environment interactions in the context of GWAS. To mimic the switch in human lifestyle, the simulated individuals were assigned either to an ancestral or a modern setting that alters the phenotype by increasing the effect size by 1.05 to 2-fold at a varying fraction of perturbed SNPs (from 1% to 20%). We report two main results. First, and for a wide range of realistic scenarios, we confirm that GxE interactions may remain hidden at the level of individual loci but be easily detected through GRSxE analyses. Second, we observe that a pervasive hidden presence of GxE interactions can heavily reduce the power to discover susceptibility variants by GWAS upon mixed cohorts. We conclude that a pervasive presence of gene-by-environment effects can remain hidden even though it contributes to the genetic architecture of complex traits.

1862T

A novel functional data analysis approach to detecting gene by longitudinal environmental exposure interaction. *P. Wei[1], H. Tang[2], D. Li[2].* 1) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 2) Dept of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX.

Most complex diseases are likely the consequence of the joint actions of genetic and environmental factors. Identification of gene-environment (GxE) interactions not only contributes to a better understanding of the disease mechanisms, but also improves disease risk prediction and targeted intervention. In contrast to the large number of genetic susceptibility loci discovered by genome-wide association studies, there have been very few successes in identifying GxE interactions which may be partly due to limited statistical power and inaccurately measured exposures. While existing statistical methods only consider interactions between genes and static environmental exposures, many environmental factors, such as air pollution and diet, change over time, and cannot be accurately captured at one measurement time point. There is a dearth of statistical methods for detecting gene by time-varying environmental exposure interactions. Here we propose a powerful functional logistic regression (FLR) approach to model the time-varying effect of longitudinal environmental exposure and its interaction with genetic factors on disease risk. Capitalizing on the powerful functional data analysis framework, our proposed FLR model is capable of accommodating longitudinal exposures measured at irregular time points and contaminated by measurement errors. We use extensive simulations to show that the proposed method can control the Type I error and is more powerful than alternative ad hoc methods. We demonstrate the utility of this new method using data from a candidate gene-based case-control study of pancreatic cancer to identify the windows of vulnerability of lifetime body mass index (BMI) on the risk of pancreatic cancer as well as genes which may modify this association. SNP rs8050136 in the FTO gene was found to be nominally significantly modifying the association between the lifetime BMI, especially early-adulthood overweight, and the pancreatic cancer risk (p-value = 0.02 based on the FLR model).

1863S

GWAS for a longitudinal trait with non-uniform errors: Recovery of CD4 cell counts after initiation of anti-retroviral therapy in two Ugandan cohorts. *J. Mefford, J. Micheli, J. Witte, D. Kroetz.* UC San Francisco,.

We present results and an assessment of alternative models for a genome-wide association study of a longitudinal trait -- CD4 T-cell recovery trajectories from a total of 516 subjects with advanced HIV disease. Subjects were from two cohorts: 396 subjects from UARTO (Uganda AIDS Rural Treatment Outcomes), and 120 from ARKS (Anti-Retrovirals for Karposi's Sarcoma). Subjects were genotyped, with 681230 markers passing QC and imputed to a 1KG reference panel. A GWAS was then run using a longitudinal model to search for genetic variants that are associated with either the rate of CD4-cell recovery after initiation of ART, or overall levels of CD4 counts. As count data, the CD4 measurements have sampling variances dependent on the CD4 level, so models are considered where a variance-stabilizing square-root transformation is applied to the CD4 measurements, or where the variance is stabilized while keeping observations on the original scale through weighting observations by their inverse variance. On either scale, the rate of increase in CD4 counts appears to be rapid in the first year and then to taper off, so piece-wise linear mixed models were used to show the time course of CD4 recovery, with breakpoints at 3 and 12 months. CD4 counts vary widely between subjects so mixed models with random intercepts and adjustments for fixed covariates were used (cohort, age, sex, and ten genetic principal components). Genetic effects were modeled as a genetic main effects and interactions of genotype with the piecewise linear time effect. Multiple degree-of-freedom likelihood ratio tests were used to compare the full models to reduced models with no genetic effects. The component genetic effects (main genetic effect and the time interaction) are examined to help interpret the GWAS results from the global genetic test. Analyses of the CD4 trajectories on the square-root transformed scale and the variance-weighted analysis on the untransformed scale show similar patterns of results with suggestive effects on chromosomes 1, 9 and 10. The total genetic effects seem driven by the gene-time interaction more than the genetic main effects. The weighted analysis on the untransformed scale has the advantage of allowing an interpretation of the gene*time effect as an additive contribution to the CD4 trajectories.

1864M

Detecting clusters of disease-associated SNPs. *D. Swanson, J. Laramie.* GNS Healthcare, Cambridge, MA.

Many methods have been developed for associating SNPs with disease, and databases exist that catalogue these SNP-phenotype associations. While genome-wide knowledge of these SNP-phenotype associations can enable researchers to locate clusters of disease-associated SNPs, thereby helping to relate them to an underlying feature such as all being found in a common gene, there do not exist methods for principled identification of such clusters. Identification of SNP-disease association clusters is made slightly more complicated because existing databases such as those found on NCBI are informed by GWAS, whose hypothesis tests are generally performed in the context of univariate regression analysis and, as a result, do not control for the non-trivial LD structure present between many adjacent SNPs--a seeming ‘cluster’ of disease-associated SNPs could therefore be driven by a single primary disease-associated SNP in LD with many other SNPs that have no association with the outcome, controlling for the primary SNP. We propose 2 methods for principled cluster detection of disease associated SNPs, one in which the measure of SNP-disease association is continuous, such as a Z-statistic, and one in which the measure of association is binary (i.e., the SNP is or is not associated with the outcome). Both methods draw on existing work coming from the disease surveillance and spatial statistics literature. We apply the two methods to two publicly available sets of GWAS results, one in which genome-wide significant SNPs are simply identified for rheumatoid arthritis and one in which there exists a continuous test statistic for each analyzed SNP for Alzheimers disease. In both cases, we detect clusters of disease-associated SNPs.

1865T

Biochemical network-driven analysis of genetic control of human metabolome. *Y.A. Tsepilov[1,2], K. Strauch[3,4], C. Gieger[3], Y.S. Aulchenko[1,2].* 1) Institute of Cytology and Genetics SD RAS, Novosibirsk, Russia; 2) Novosibirsk State University, Novosibirsk, Russia; 3) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 4) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany.

Genome-wide association analysis (GWAS) is one of the most popular methods of identification of alleles that affect complex traits. Recent accumulation of functional genomics data ("ome" data, e.g. transcriptome, metabolome etc) could give a new insight into the functional role of specific changes in the genome. Such data require special statistical methods for their analysis, as they are characteristically high dimensional and contain statistical interactions and functional relationships between individual components, which usually reflect biological interactions. Development of such methods is of current importance as the progress of molecular biology techniques marches on. We introduce a new approach to evaluate the impact of genetic variants to "omics" phenotypes with relatively high power and rather modest computational intensities. Our approach is based on the idea that if biological pathways and relationships in data could be reconstructed, we can use the knowledge about biochemical neighbors for chosen trait and include this information into analysis of this trait. The sources of biological relationship information can vary: biological pathway databases or reconstructed net from the data. We assessed our approach using real population study data from big German study KORA (n=1,784, 2M SNPs) with measured metabolomics data (151 metabolites). For reconstruction of the pathways we used Gaussian Graphical Models (GGM) which was shown as effective tool for the unbiased reconstruction of metabolic reaction. Previously it was demonstrated that using GGM-driven ratios instead of all ratios leads to comparable power with much less computational expenses. Our approach as well has shown the comparable power with all ratio approach with computational complexity similar to the analysis of original concentrations only (n instead of n^2). Using of biochemically related covariates decreases the phenotype variances and the noise driven by network influences between metabolites with the dissolution of induced signals and the improvement of actual associations that helps with further functional annotation.

1866S

Joint Analysis of Genetic Interaction and Imprinting in Family Studies. *C.C. Wu[1], S. Shete[2].* 1) Environmental/Occupational Health, National Cheng Kung University, Tainan, Taiwan; 2) Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

Genomic imprinting is considered a primary epigenetic phenomenon, which results in 2 alleles at a locus being expressed differentially depending on the parental origin of the chromosome. The role of imprinted genes and gene-gene interactions in disease susceptibility is well recognized. Imprinting leads to manifestation of parent-of-origin effects. More than 1000 individual genetic variants associated with common diseases have been detected by recent whole genome studies; the efforts to map interacting genes have been less successful. In response, we proposed a method to detect genetic interaction effects in the presence of genomic imprinting. We extended our previous allele-sharing method that accounts for parent-of-origin effects in the affected sib pairs framework. We derived 3 mathematical two-locus models that incorporate parent-of-origin effects: additive, multiplicative, and general models. We further proposed to use a novel test to assess the gene-gene interaction and parent-of-origin effects individually and jointly. The statistical properties and performance of the proposed method will be evaluated using simulated data.

1867M

Gene-gene interactions in admixed populations. *E. Ziv, D. Hu, L. Fejerman.* Department of Medicine, Institute for Human Genetics, Helen Diller Family Comprehensive Cancer Center, UCSF, San Francisco, CA.

Identifying gene-gene interactions is a challenge in human genetics due to the limitations of power after adjusting for the large number of potential interactions tested. We examine the approach to identifying gene-gene interactions in an admixed population. We model a dichotomous trait (disease vs. no disease) in an admixed population consisting of two ancestral populations. First, we consider how the type of genetic interaction and the range of allele frequencies in the ancestral populations affect the association between genetic ancestry and disease. We demonstrate that there are distinct non-linear patterns of the association between disease risk and genetic ancestry depending on the allele frequency differences between ancestral populations. Furthermore, we consider how, by capitalizing on these patterns, investigators can design studies that enhance the power to detect gene-gene interactions. We illustrate the approach using simulations and a study of breast cancer in a U.S. Latina population.

1868T

Integration of multiple types of functional annotation with genotype data in genetic association studies at gene and pathway levels. *Y. Guo[1,2], D.V. Conti[1,2], G.K. Chen[1,2], K. Wang[1,3,4].* 1) Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA; 2) Division of Biostatistics, Department of Preventive Medicine, University of Southern California, Los Angeles, CA; 3) Department of Psychiatry & Behavioral Sciences, University of Southern California, Los Angeles, CA; 4) Division of Bioinformatics, Department of Preventive Medicine, University of Southern California, Los Angeles, CA.

Association tests on multiple levels of regions, such as genes and biological pathways provide an alternative to conventional genome-wide association studies (GWAS). Compared with single SNP-based methods, this approach may have increased power, can facilitate direct comparison across different populations or genotyping platforms, and can ease interpretation. Quintana and Conti (Genetic Epidemiology, 2011) proposed to use a combination of Bayesian model uncertainty and a hierarchical model to construct a Bayesian risk index as a measurement for strength of regional association. In their method, various sources of biological annotations can be incorporated into the model to help select variants and Metropolis-Hastings (MH) and Gibbs Sampling techniques are used for inference. There are, however, two major hurdles for this approach. First, computational cost associated with MH method grows cubically with the number of variants, limiting its scalability with next-generation sequencing data. Second, the approach is limited to a few sources of biological information and it is unclear how to select annotations from thousands of options available. Here we alter the algorithm to gain efficiency in computation. At the beginning of the MH procedure, reversible jump Markov Chain Monte Carlo is used to select a subset of variants, which constitute a more plausible model, as a starting point. In each iteration of the stochastic search, a proposal kernel is more likely to add a variant that is close to elevated signal of activity or in high linkage disequilibrium with a nearby GWAS hit. To select prior information, we applied the statistical model (fgwas) by Pickrell (American Journal of Human Genetics, 2014) to pick a few annotations that most differentiate associated variants from non-associated variants. Using this approach, we incorporated multiple DNase, chromHMM, ChIP annotation data tracks from UCSC genome browser. Our method has several conceptual advantages than existing approaches and can be easily scaled up from GWAS to sequencing studies on large population cohorts. Through simulation and real data analysis, we demonstrate the applicability of our method and show performance improvements compared to competing methods.

1869S

Identification of shared genetic aetiology between epidemiologically linked disorders with an application to obesity and osteoarthritis. *J. Asimit[1], K. Panoutsopoulou[1], E. Wheeler[1], S. Berndt[4], A.P. Morris[2,3], I. Barroso[1], E. Zeggini[1], the GIANT consortium, the arcOGEN consortium.* 1) Human Genetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 3) Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom; 4) US Department of Health and Human Services, Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA.

Diseases often co-occur in individuals more likely than expected by chance. Shared genetic aetiology may explain this, and the discovery of shared variants may suggest a common biological process. A common approach to a genetic overlap analysis is to select a p-value threshold and compare the signals from the genome-wide association study (GWAS) of each trait. However, p-values do not account for differences in power, whereas Bayes' factors do, and may be approximated using summary statistics. We use simulation studies to compare the power of frequentist and Bayesian approaches to overlap analyses, and to decide on appropriate thresholds for comparison between the two methods. It is empirically illustrated in single-disease associations that BFs have the advantage over p-values of a decreasing proportion of false positives (PFP) as study size increases. For a $\log_{10}$(BF) threshold $L\theta$ of 1.69 (R=type II error cost/type I error cost=2, π_0 = Pr(no association at SNP)=0.99), the PFP decreases from 7.38×10^{-4} (N=2,000 each cases/controls) to 3.37×10^{-4} (N=20,000) , while for p-values the PFP fluctuates near the p-value threshold α regardless of study size. Power increases with sample size are more noticeable for the Bayesian approach. For OR=1.2 and MAF 0.1, the power in studies with N=3,000 are 0.413 ($L\theta$=1.69) and 0.365 ($\alpha=5\times10^{-4}$), while at N=5,000 the respective powers rise to 0.750 and 0.716. In a preliminary overlap analysis of obesity (GIANT consortium) with OA (arcOGEN consortium), the number of signals is similar at comparable threshold levels between BFs and p-values. The two sets of results are not always overlapping and additional signals are often in already detected genes. For $L\theta$=0.91 (R=type II error cost/type I error cost=12, π_0=Pr(no association at variant)=0.99), there are 18 shared variants identified as associated, and the comparable α levels of 0.003 and 0.004 yield 15 and 28 hits, respectively. The most notable difference is that the Bayesian list contains rs13107325 (in *SLC39A8/ZIP8*), a variant previously associated with obesity-related phenotypes such as BMI, blood pressure, and HDL cholesterol, and animal studies have shown that the zinc-ZIP8-MTF1 axis regulates OA pathogenesis. This variant is only detected via p-values if we increase α to 0.005. We are pursuing replication of this finding. Extensive simulations to systematically evaluate power differences between the BF and p-value-based overlap analysis approaches are underway.

1870M

Multilevel dimensionality reduction algorithms for high-dimensional genetics data. *K. Cho[1, 2], D.R. Gagnon[2, 3], H. Wu[2, 4].* 1) Division of Aging, Brigham and Women's Hospital/Harvard Medical School, Boston, MA. USA; 2) Massachusetts Veterans Epidemiology Research and Information Center; VA Boston Healthcare System, Boston, MA, USA; 3) Boston University School of Public Health, Boston, MA, USA; 4) Computer Science and Networking, Wentworth Institute of Technology; Boston, MA, USA.

With rapid advancements in genotyping and sequencing technologies, there are more genetic data available than one can thoroughly analyze with existing tools and methods. This has become one of the analytical challenges in the initial phases of big scale studies. In whole-genome or genome-wide exploration of novel associations, without a priori knowledge of specific regions or genes of interest, researchers rely on the available techniques that perform association tests based on a way of summarization or assuming independence. In addition, traditional statistical procedures present eminent challenges in using these data, where the number of parameters p is scalably larger than number of observations n. These limitations may compromise the level of false positive results and power to conduct studies with acceptable level of confidence. One of the ways to reduce the initial efficiency burden with high dimensional genetic data is to first perform a data dimension reduction process and then evaluate the resulting panels of markers. We propose several algorithms using traditional PCA based algorithms, LASSO, and combinations of these to evaluate the impact of type one error. Using the Genetic Analysis Workshop 18 simulated data, we apply the several algorithms on null regions of chromosome 3 with regards to the diastolic blood pressure phenotype. Then we use SKAT as the evaluation tool to perform association analysis using the resulting panels from each algorithm. We then systematically compare type I error rates among different algorithms with respect to the baseline panel. Our preliminary work using 360,717 SNPs containing 1431 gene groups in unrelated samples, shows a substantial reduction of the markers through applying proposed algorithms. Our PCA based algorithm was applied to individual SNP sets and also to gene group sets. Among these, the PCA top SNP and PCA genes algorithms resulted in 106 and 23297 unique SNPs, respectively, at 85% variation. At 95% variation, 127 and 35382 resulting SNPs remained, respectively. We are also investigating PCA followed by LASSO algorithm and PCA incorporating heritability approaches using the family datasets. As the amount and dimensionality of the network of genomics data is only escalating in the coming years, a timely, robust and practical analytical pipeline through multilevel dimensionality reduction techniques provides an efficient approach for the initial screening tool.

1871T

Estimating genetic distance-dependent effects of environmental exposures by functional models. *D. Zaykin[1], O. Vsevolozhskaya[2], Q. Lu[2].* 1) National Institute of Environmental Health Sciences; 2) Michigan State University.

We present a functional method for fitting simultaneously the effects of scalar predictors, confounding factors and environmental exposures on sequenced allelic variants in a genetic region. These effects are fitted as continuous curves smoothly varying over genetic loci to help researchers shed light on which regions harbor causal variants. In a densely sequenced region, one can plot genotypic values on the y-axis versus variant position on the x-axis and 'connect the dots'. The resulting continuous genotype curve will change over variants position. By utilizing continuous curves, our method exploits linkage disequilibrium among variants in the ways that otherwise is difficult to achieve. For example, one can regress genotype curves on a set of environmental variables and obtain a subject-specific set of curves for each environmental effect that will also change over variants position. The regression on the genotype curve is produced based on a penalized spline regression and can be linked to linear mixed models which allows utilization of standard statistical software packages. We investigate advantages and disadvantages of our approach by conducting an extensive simulation study. We illustrate the approach with sequencing data from the Dallas Heart Study.

1872S

Multivariate approach for finding gene sets differentially expressed by complex phenotype. *E. Drigalenko, H.H.H. Göring.* Dept Genetics, Texas Biomedical Research Institute, San Antonio, TX.

Genes almost never act alone in a biological system; they typically work in a cascade of networks. This motivates the consideration of clusters of genes, instead of individual genes, as functional genomic units. Variations of certain phenotypes, including complex diseases and quantitative traits, may be associated with differential expressions of multiple genes instead of a single gene.

Transcriptome sequencing (RNA-seq) technology allows measuring of the expression levels of tens of thousands of genes simultaneously. Expression profiles of multiple genes are correlated and thus are appropriately modeled as mutually correlated variables in a statistical testing framework.

Here, we propose three multivariate methods for testing the association of sets of correlated genes with discrete or continuous phenotypic outcomes: (1) comparison of covariate matrices under different conditions using Box's M test; (2) joint regression of all genes in a pathway; (3) orthogonal decomposition of the variance and the regression of principal components (PC) using Kaiser's criteria for the number of essential PCs. KEGG pathway classes C, B, and A were studied separately. We use programs implemented in R. The last two methods can be used for quantitative traits.

To demonstrate the performance, we apply these methods to publicly available data from the NCBI Gene Expression Omnibus database. For this presentation, we used the genes expression (RNA-seq) data from 21 individuals (Akula et al, 2014; GSE53239) - 10 with bipolar disorder and 11 controls. As an example of gene sets, here we have used Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

We found that the proposed methods do identify differentially expressed gene pathways. In our example, the PC-based method identified the most pathways. Using it, after the Bonferroni correction for the number of pathways is a class, we found that 27 pathways are significant of total 282 in KEGG Class C, 23 of 39 in Class B, and all 6 of 6 in Class A. When examining several other example gene expression datasets, the PC-based method identified the most pathways in all instances.

1873M

Estimating and interpreting pairwise genetic correlations between hundreds of quantitative traits from population samples of thousands of individuals. *M. Pirinen[1], C. Benner[1], M.A. Rivas[2], T. Lehtimäki[3], A.J. Kangas[4,5], P. Soininen[4,5], M. Ala-Korpela[4,5,6], J.G. Eriksson[7,8,9], O.T. Raitakari[10,11], M.R. Järvelin[12,13,14], V. Salomaa[7], S. Ripatti[1,15,16].* 1) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Department of Clinical Chemistry, Fimlab Laboratories, University of Tampere School of Medicine, Tampere, Finland; 4) Computational Medicine, Institute of Health Sciences, University of Oulu and Oulu University Hospital, Oulu, Finland; 5) NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland; 6) Computational Medicine, School of Social and Community Medicine and the Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK; 7) Department of Chronic Disease Prevention, National Institute for Health and Welfare, Finland; 8) Department of General Practice and Primary Health Care, University of Helsinki, Finland; 9) Unit of General Practice, Helsinki University Central Hospital, Finland; 10) Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland; 11) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; 12) Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, UK; 13) Institute of Health Sciences, University of Oulu, Finland; 14) Biocenter Oulu, University of Oulu, Finland; 15) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 16) Hjelt Institute, University of Helsinki, Helsinki, Finland.

Several modern technologies, such as nuclear magnetic resonance and mass spectrometry platforms, produce high-dimensional phenotype data on individuals. A first step towards utilising high-dimensional phenotypes in genetic studies is to understand how their genetic components are related. Recent algorithmic advances in multivariate linear mixed models have enabled variance component estimation for pairs of traits using population samples of individuals and genome-wide panels of SNPs. However, current methods have not been tailored for situations where hundreds of traits are available on the same set of individuals. For such settings, we introduce an algorithm that efficiently decomposes pairwise phenotypic correlations into genetic and environmental components. We illustrate our approach with an application to thousands of pairs of metabolic and anthropometric traits measured on up to 14,000 Finnish individuals. For example, we estimate that the observed phenotypic correlation (-0.41) between triglyserides and HDL cholesterol decomposes into an additive genetic correlation (-0.59, s.e. 0.06) and an environmental correlation (-0.36 s.e. 0.02). We assess the interpretation of genetic correlations as correlations between locus-wise genetic effects by applying multivariate statistical models on the results of a recent genome-wide meta-analysis on lipid levels. Finally, we consider settings where prior information about genetic correlation increases statistical power to identify pleiotropic loci, i.e. loci that contribute to multiple traits.

1874T

Systemic genetics of Systemic Sclerosis through protein-protein interaction network-based analysis. *J. HAMON[1], Y. ALLANORE[2], M. MARTINEZ[1].* 1) INSERM UMR1043, Hôpital Purpan, Toulouse, France; 2) INSERM UMR1016, Hôpital Cochin, Paris, France.

Genome-wide association studies (GWAS) of complex traits have revealed reproducible genetic associations with human complex disorders, including Systemic Sclerosis. However, these studies rely on analysis at the single-nucleotide polymorphisms (SNP) level and, consequently, lack power in presence of allelic heterogeneity (multiple independent risk variants within a given gene/genomic region) and/or for detecting interacting risk variants (with low marginal effects). This complexity has generated considerable interest in multi-locus analysis techniques as, for instance, knowledge-driven methods that integrate known functional and interaction networks. These approaches vary according to the type of known information (protein-protein network/gene function pathways) that is integrated to the GWAS data; the way GWAS data in genes is used (single-SNP or SNP-set statistical significance); and the model used to test for association between a module/pathway of genes and the phenotype. Here, we present results from various study designs for network-based analysis integrated to our Systemic Sclerosis GWAS data (564 cases and 1,776 controls). We have built a robust two-stage network analysis by randomly splitting our GWAS data into a scan and a replication dataset. In the scan dataset we performed a PPI network-based approach using a dense module search strategy and using different approaches to assess GW : for instance, from the lowest single-SNP P value unadjusted (Min) or Bonferroni-adjusted (Bonf) for the number of SNPs within the gene, or by combining all single-SNP P values with the Fisher's method. We first compared the results obtained under the different strategies according to the length (i.e., number of genes) of the enriched modules and the characteristics of the genes within the modules. The top (5 and 10%) most enriched modules were further tested for enrichment analysis in the replication dataset. We finally assessed whether the genes from the replicated sub-networks were significantly clustered within KEGG pathways. Our results show that the gene-wise association P-value modeling has a large impact on the results, i.e., length (number of genes) of sub-networks, characteristics of their genes and replication rate of top most enriched sub-networks. Overall, we found low consistency across the results from the different strategies: the different strategies tend to select different sets of genes but also different KEGG pathways.

1875S

Mixed-model analysis of common variation reveals pathways explaining variance in AMD risk. *J. Hall[1], M. Pericak-Vance[2], W. Scott[2], J. Kovach[2], S. Schwartz[2], A. Agarwal[3], M. Brantley[3], J. Haines[1], W. Bush[1].* 1) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH; 2) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 3) Department of Ophthalmology and Visual Sciences, Vanderbilt University, Nashville, TN.

Age-related macular degeneration (AMD) is the leading cause of irreversible blindness in the elderly in developed countries and can affect more than 10% of individuals over the age of 80 in many populations. AMD has a large genetic component, with overall heritability estimated to be between 45% & 70%. Numerous AMD loci have been identified and implicate various molecular mechanisms and pathways in AMD pathogenesis. Eight pathways, including angiogenesis, antioxidant activity, apoptosis, complement activation, inflammatory response, nicotine metabolism, oxidative phosphorylation, and the tricarboxylic acid cycle, were selected for our study based on an extensive literature review. While these pathways have been proposed in literature, the overall extent that each pathway contributes to AMD risk is unknown. In a dataset of 1,154 AMD cases and 668 controls, we used Genome-wide Complex Trait Analysis (GCTA) to conduct mixed-model regression to estimate the proportion of variance in AMD risk explained by all SNPs in each pathway. Genes were assigned to pathways using Gene Ontology terms and SNPs were mapped to these genes and then assessed using GCTA. To extend the analysis beyond coding variation, nearby SNPs within a 50 kilobases region flanking each gene, many likely to be regulatory, were assessed, as well as more distant, putatively regulatory SNPs within 500 kilobases of each gene, based on retinal pigment epithelium (RPE) DNase I Hypersinsitivity mappings from the ENCODE project. To account for baseline known risk, we tested 19 established AMD risk SNPs and found that they contributed to 13.3% of the variation in risk in our dataset, while the remaining 659,181 SNPs contributed to 36.7%. After adjusting for these 19 risk SNPs, the complement activation and inflammatory response pathways explained a significant ($P = 6.8 \times 10^{-26}$ and 9.5×10^{-8}, respectively) proportion of additional variance in AMD risk (9.8% and 17.9%, respectively), with other pathways showing no significant effects (0.3% - 4.4%). Our results show that additional variants associated with complement activation and inflammation genes contribute to AMD risk and that these variants are likely in coding and nearby regulatory regions.

1876M

A non-parametric approach for detecting gene-gene interactions associated with age-at-onset outcomes. *M. Li[1], J.C. Gardiner[2], N. Breslau[2], J.C. Anthony[2], Q. Lu[2].* 1) Section of Biostatistics at Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR; 2) Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI.

Cox-regression-based methods have been commonly used for the analyses of survival outcomes, such as age-at-disease-onset. These methods generally assume the hazard functions are proportional among various risk groups. However, such an assumption may not be valid in genetic association studies, especially when complex interactions are involved. In addition, genetic association studies commonly adopt case-control designs. Direct use of Cox regression to case-control data may yield biased estimators and incorrect statistical inference. We propose a non-parametric approach, the weighted Nelson-Aalen (WNA) approach, for detecting genetic variants that are associated with age-dependent outcomes. The proposed approach can be directly applied to prospective cohort studies, and can be easily extended for population-based case-control studies. Moreover, it does not rely on any assumptions of the disease inheritance models, and is able to capture high-order gene-gene interactions. Through simulations, we show the proposed approach outperforms Cox-regression-based methods in various scenarios. We also conduct an empirical study of progression of nicotine dependence by applying the WNA approach to three independent datasets from the Study of Addiction: Genetics and Environment. In the initial dataset, two SNPs, rs6570989 and rs2930357, located in genes GRIK2 and CSMD1, are found to be significantly associated with the progression of nicotine dependence (ND). The joint association is further replicated in two independent datasets. Further analysis suggests that these two genes may interact and be associated with the progression of ND.

1877T

Estimation of Heritability and Association for Quantitative Traits with Repeated HbA1c Measures: Biomarker for Metformin Response. *L. Wu[1], J. Mefford[1], S.Y. Yee[2], K.M. Giacomini[2], J.S. Witte[3].* 1) Department of Epidemiology/Biostatistics University of California , San Francisco, CA; 2) Department of Bioengineering and Therapeutic Sciences,University of California , San Francisco, CA; 3) Department of Epidemiology/Biostatistics and Urology, University of California , San Francisco, CA.

Assessing heritability is important for understanding the genetic architecture of quantitative traits. Heritability explained by typed SNPs is commonly calculated using linear mixed-effects models with an estimated genetic relationship matrix (e.g., with GCTA). We adopt a linear mixed-effects model to estimate heritability and carry out association study of quantitative traits with repeated measures. We also propose a new estimate of the genetic relationship matrix whose elements are the cosine similarity of mean-centered genotype vectors of two individuals; assuming no inbreeding, this method guarantees that the diagonal elements equal one, which provides a better characterization of additive genetic relationship of the same individual. In contrast, conventional methods only guarantee the average of diagonal elements to be close to one. The marginal model implied by the linear mixed-effects model is used to lower the computational burden in association analyses. And following an initial screen, the top-ranked SNPs can be reevaluated one at a time to produce more accurate p-values of their association. In our study, we use hemoglobin A1c as a biomarker for response to anti-diabetic drugs such as metformin. We estimate the heritability of metformin response in a cohort of 500 African American diabetic patients with 1,320 repeated HbA1c measures after metformin treatment. For this data, the genetic relationship matrix obtained from GCTA has diagonal elements ranging from 0.926 to 1.511 (average = 1.004, SD 0.014); in contrast, our approach gives values all equal to 1. The off-diagonal elements of the genetic relationship matrices are similar: -0.002 (SD 0.015) from GCTA and -0.002 (SD 0.014) from our proposed method. When calculating heritability of metformin response, using the genetic relationship matrix from GCTA gives a value of 0.19 while our approach gives a value of 0.16. In conclusion, our new method can be used with repeated phenotypic measures to calculate the genetic relationship matrix in estimating heritability. The method guarantees that the diagonal elements of an estimated genetic relationship matrix equal one, offering a potentially valuable alternative to the conventional approaches in estimation of heritability and genetic association study.

1878S

Genome-Wide Association and Gene-Gene Interaction Studies to Explore Etiology of Glaucoma and Ocular Hypertension. *S. Verma[1], A. Verma[1], A. Lucas[1], J. Linneman[2], P. Peissig[2], M. Brilliant[2], C. McCarty[3], J. Haines[4], T. Vrabec[5], G. Tromp[5], J. Pathak[6], C. Chute[6], D. Crosslin[7,8], G. Jarvik[7,8], E. Larson[9], M. Hayes[9], M. Ritchie[1].* 1) Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA; 2) Marshfield Clinic, Marshfield WI; 3) Essentia Rural Health, Duluth, MN; 4) Case Western University, Cleveland, OH; 5) Geisinger Health System, Danville, PA; 6) Mayo Clinic, Rochester, MN; 7) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 8) Group Health Research Institute, Seattle WA; 9) Northwestern University, Chicago, IL.

Ocular hypertension (OHT) is a leading risk factor for glaucoma. Previous studies have identified interesting genetic and environmental factors for these two traits, which are among the leading causes of blindness worldwide. In the electronic MEdical Records and GEnomics (eMERGE) network, sites extracted phenotype information on patients with glaucoma and OHT from the EHR who were already genotyped and part of DNA biobanks. We performed GWAS and gene-gene interaction (GxG) studies for both of these traits using data imputed to 1000Genomes. A total of 3,253 and 3,154 unrelated samples of ages 40-90 across five different eMERGE site biorepositories were assembled for glaucoma and OHT studies respectively. We observed novel genome-wide significant variations on Chr1 (rs4147800) associated with glaucoma ($p= 1.04\times10^{-9}$) and on Chr13 (rs4646227) and Chr16 (rs36120466) associated with OHT ($p= 3.92\times10^{-9}$ and 4.52×10^{-9}). To examine epistasis between multiple genetic variants, considering extensive computational demands and multiple testing burden of exhaustive testing, we applied a main effect filter (SNPs with p-value less than 0.01 from GWAS) and then performed genome-wide interaction study (GWIS) using regression methods. We found several interesting interactions that remain significant after Bonferroni correction. ABCA4 shows a significant interaction with GAD2 in OHT (LRT $p= 2.71\times10^{-11}$) and also we observed significant interaction (LRT $p= 1.85\times10^{-7}$) between two genes (PON3 and PON2) in member of paraoxonase family that validates previously published associations between primary open-angle glaucoma and antioxidant enzymes affecting high-density lipoprotein (HDL). We also merged these results and performed pathway-enrichment analysis for all results with main effect $p<1\times10^{-4}$. KEGG Pathway analysis found 10 pathways that are shared for glaucoma and OHT. Notably, ABC transporter pathway contain the most significant genes associated with both Glaucoma and OHT. Similarly Serotonin 5-HT(2) receptor agonists (LTB4R and AGRP genes) linked to neuroactive ligand-receptor interactions (most significant association in OHT) are associated with both phenotypes and they have been also identified as potential agents for the treatment of both OHT and glaucoma. Our findings indicate novel SNP associations and GxG interactions for both of these traits and demonstrate the relationship between these two phenotypes at a potential molecular level with the guide of pathway analysis.

1879M

Major-Effect Loci for Lipids also Impact Phenotype Variability in the Old Order Amish. *L. Yerges-Armstrong, J. O'Connell.* Program in Personalized and Genomic Medicine, and Department of Medicine, Division of Endocrinology, Diabetes and Nutrition - University of Maryland School of Medicine, USA. Baltimore, MD 21201.

Association studies for complex traits have traditionally focused on identifying loci that shift the mean of a quantitative trait. Several recent papers describe methods for modeling genetic loci that instead impact trait variability. These phenotypic variance quantitative trait loci (or vQTL) are of particular interest as they may indicate undetected genetic or environmental interaction. With this in mind, we tested three loci for lipids segregating in the Old Order Amish (OOA) for the presence of vQTL effects. Analyses were conducted on over 1500 participants genotyped on the Illumina Human Exome BeadChip. All analyses were conducted using our mixed models analysis for pedigrees and populations (MMAP) software. Levene's F-test was implemented in MMAP to allow us to model genotype differences accounting for polygenic effects, age and sex. Both the mean and median F-tests were calculated but only the median test is presented. The first variant tested was the *APOC3* null mutation (R19X) associated with markedly lower mean triglycerides ($p=2\times10^{-23}$). In addition to the large effect on the mean, there was a significant vQTL effect ($p=4.6\times10^{-6}$). The second variant was the common, promoter variant in *CETP* (rs3764261), which was highly significant for mean differences in high-density lipoprotein cholesterol ($p=2.6\times10^{-10}$) but had only a modest vQTL effect (p=0.02). Due to a founder effect ~10% of Amish are carriers of the *APOB* 3500Q allele which is associated with ~60mg/dL higher low-density lipoprotein (LDL) cholesterol ($p=1.2\times10^{-94}$). We observed a significant vQTL ($p=1.3\times10^{-6}$) for LDL with this variant. We have previously reported on the higher coronary artery calcification (CAC) in 3500Q carriers thought to be caused by high LDL cholesterol. In 896 OOA with CAC measures we continue to see a strong vQTL for LDL cholesterol ($p=1.7\times10^{-8}$) but not for CAC (p=0.17) despite the much higher mean CAC levels in mutation carriers. To our knowledge, the vQTL effects for R19X and R3500Q are the first reported for lipids. The lack of vQTL effect for *CETP* may indicate that highly-penetrant, rare variants are more likely to be vQTLs. Interestingly, we only detected a vQTL for *APOB* with LDL and not with CAC despite the highly significant mean effect on both traits. This could be a reflection of R3500Q increasing CAC through LDL and not impacting calcification directly, but additional work with other pleiotropic variants is needed before this generalization can be made.

1880T

An integrative imputation method for multi-omic datasets. *D. LIN[1,2], J. ZHANG[2,3], J. LI[1,2], H. DENG[2,3,4], Y. WANG[1,2,3,4].* 1) Biomedical Engineering, Tulane University, New Orleans, LA,70118,USA; 2) Center of Genomics and Bioinformatics, Tulane University, New Orleans, LA, 70112, USA; 3) Department of Biostatistics and Bioinformatics, Tulane University, New Orleans, LA, 70112, USA; 4) Center for Systems Medicine, Shanghai University for Science and Technology, Shanghai, CHINA.

Despite the increasing significance of integrating diverse types of genomic and epigenetic data (e.g. mRNA, microRNA and DNA methylation), the currently available multi-omics datasets inevitably suffer missing values due to technical limitations and various constrains in an experiment. Those missing values will severely hinder the downstream analysis such as the construction of regulation patterns among genetic and epigenetic factors, and the identification of differential expressed biomarkers. Currently, imputation methods for single data type have been well studied and compared. More evidence has shown the biological interconnections among these genomic and epigenetic data. It is significant to integrate these correlated multi-omic datasets for improving the imputation accuracy of missing data. In this study, a novel imputation method was proposed to address this issue by :1) combining the estimates of missing value from individual omic datasets in a way of model stacking to obtain an optimal estimate; and 2) incorporating prior biological knowledge into the imputation across different datasets. The real Glioma cancer datasets from TCGA including 50 patients with mRNA measurements of 5939 genes, expression data of 104 microRNAs and DNA methylation measurements of 5013 genes were included in the analysis. We compared the our method with five single data imputation methods (kNN, LLS, ILLS, BPCA, and SVD) in different noise levels, sample sizes and missing rates, and the performance was evaluated by normalized root mean square error. The results indicated that our integrative method showed the lowest imputation error among all other methods across all situations. The superior performance was more prominent in cases of smaller sample size and higher noise level. We concluded that our proposed imputation method will improve the imputation accuracy by incorporating more omic datasets as well as prior biological information.

1881S

Combining allele-specific and population signal boosts power for association mapping of multiple DNA sequence-based cellular traits. *N. Kumasaka, D. Gaffney.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Association mapping of cellular traits, such as gene expression or epigenetic marking, is a powerful method for uncovering the function of genetic variation. When cellular traits are quantified using high-throughput DNA-sequencing, association signals can be detected using two orthogonal sources of information: population level differences between individuals, and allele specific differences within individuals. Here, we develop a unified probabilistic model for combining allele-specific and population level sequencing data that significantly boosts power for association detection in DNA sequencing-based cell traits, and apply this to multiple publicly available population level data sets of RNA-seq, DNase-seq and ChIP-seq data. A couple of similar approaches has already been proposed (Sun, 2012; and McVicker et al 2013). However those methods integrate allele-specific information within individuals that are heterozygous at the variant within sequenced feature (Sun 2012) or those for heterozygous at the linked putative causal variant (McVicker 2013). Our approach combines information not only from heterozygoutes but also homozygotes across multiple biallelic SNPs within a sequenced feature. It shows a significant improvement over the publicly available approaches, in particular, it significantly improves power for association detection in small sample sizes, with an approximately 7.5% increase in sensitivity at FDR 5% over the next best performing method (Sun 2012) in a sample size of 24 individuals and a 14.3% increase over simple linear regression. In addition, an internal analysis also shows that using information from both heterozygotes and homozygotes gives better power than just using heterozygotes with 4% additional power for the same data set. Our approach also explicitly models important experimental biases including mapping errors, genotype uncertainty and reference mapping bias. We believe our method is a general approach that will significantly improve association detection for any cellular phenotype in which allele-specific signature can be quantified, and will be particularly useful in cases where sample sizes are small.

1882M

Combining the association and ancestry signals through a multivariate model. *S. Eyheramendy, C. Parada.* Pontificia Universidad Catolica de Chile, Santiago, RM, Chile.

Over the past several thousand years many genetic variants have been selected, and many of these are particular to continental groups, leading to a substantial variation in the prevalence of many simple and complex diseases. The genome of individuals from recently admixed populations is a mosaic of segments from two or more ancestral populations. Thus, recently admixed populations are likely to have a larger number of genetic variants with functional effects. In recently admixed population, candidate causal risks can be search by performing the usual comparison in allele frequency between cases and controls (correcting for population structure), by performing an analysis of cases-only (admixture mapping), or by using cases-and-controls to search for unusual deviation in local ancestry. SNP and admixture association signals contain different information that can complement each other with which more powerful tests have been developed. In our study we propose to simultaneously model the two signals. We show the advantages and disadvantages of our method.

1883T

Explaining missing heritability using Gaussian Process Regression. *K.J. Sharp[1], W. Wiegerinck[1], A. Arias-Vasquez[2,3], B. Franke[2,3], C.A. Albers[2,4], H.J. Kappen[1].* 1) Biophysics, Radboud University, Nijmegen, Netherlands; 2) Human Genetics, Radboud University Medical Centre, Nijmegen, Netherlands; 3) Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, Netherlands; 4) Molecular Developmental Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University Medical Centre, Nijmegen, Netherlands.

For many traits and common human diseases, causal loci uncovered by genetic association studies account for little of the known heritable variation. 'Missing heritability' might lie in the effect of non-additive interactions between multiple loci, but this has been difficult to test using existing parametric approaches. We employed a non-parametric, Bayesian method, based on Gaussian Process Regression, for identifying associated loci in the presence of interactions of arbitrary order. On both simulated and real datasets we demonstrate that the method has considerable power to detect high-order interactions and explain missing heritability.

As a proof of principle, we analysed 46 quantitative yeast phenotypes. Whereas detected pairwise QTL-QTL interactions accounted for little of the variance (a median of 3% per trait), we found that over 70% of the total known missing heritability could be explained using common genetic variants, many without significant marginal effects. Interestingly, the availability of biological replicates significantly improved the power to identify such loci and, hence, to explain variance.

These results already represent a significant advance in approaches to understanding the missing heritability problem with potentially important implications for studies of complex, quantitative traits. Importantly, however, features of the algorithm can be exploited to permit application to datasets incorporating much larger numbers of putative QTLs. In particular, the most computationally expensive steps consist of large numbers of independent computations that conform to the SIMT parallel computation model of Graphics Processing Units (GPUs). We also describe work in progress to develop such an implementation. Using a single Nvidia Tesla M2090 GPU we already achieve two orders of magnitude improvement in wall-clock time per iteration over a serial implementation. This indicates the potential of the approach for application to human GWAS datasets.

1884S

Comparison of GWAS results from imputed SNPs and multiple anchor and partner genotyped SNPs in an isolated population, Samoa. *R.L. Minster[1], O.D. Buhule[2], N.L. Hawley[3,4], G. Sun[5], S. Viali[6], R. Deka[5], D.E. Weeks[1,2], S.T. McGarvey[7].* 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 3) Weight Control and Diabetes Research Center, the Miriam Hospital, Providence, RI, USA; 4) The Alpert Medical School, Brown University, Providence, RI, USA; 5) Department of Environmental Health, School of Medicine, University of Cincinnati, Cincinnati, OH, USA; 6) Medical Specialist Clinic and National Health Services, Government of Samoa, Apia, Samoa; 7) Department of Epidemiology, International Health Institute, School of Public Health, Brown University, Providence, RI, USA.

Imputation of genotypes using reference panels such as the 1000 Genomes Project allows for testing at markers absent from GWAS genotyping arrays. Signals have been observed using imputed genotypes that were unobserved at lab-genotyped markers. It has also been shown that supplementing global reference panels with haplotypes specific to the population under examination improves imputation quality. For isolated populations where haplotype panels have yet to be constructed, the global reference panels for imputation may therefore be of reduced value. Additionally, imputing genotypes has high computational costs, both in processing time and data storage. Howey and Cordell (2014; PMID: 24535679) proposed a method called SnipSnip that purports, in some situations, to circumvent the uncertainties and costs of imputation. SnipSnip uses pairs of typed SNPs to detect associations at untyped SNPs. We have genotypes, after stringent QC, for 895,103 markers in 3,122 study participants from Samoa, but data are not available to build a Samoa-specific haplotype panel. Thus, a method that allows us to detect associated variants absent from our genotyping array would be valuable. Here we compare three methods for the detection of association: (1) GWAS SNPs alone using mixed linear models, (2) GWAS SNPs plus SNPs imputed with SHAPEIT/IMPUTE2 from the 1000 Genomes phased reference data using PLINK and (3) GWAS SNPs alone using SnipSnip. We examined two regions of the genome in particular: (1) 5q35, for which we have a prior strong GWAS association with body mass index; and (2) 4q13, for which we have a prior marginally genomewide-significant association with fasting glucose, but a strong SnipSnip association. At 5q35 the peak p value for the GWAS was 4.93e-14; for the GWAS + imputation was 5.01e-16; and for SnipSnip was 1.85e-11. At 4q13 the peak p value for the GWAS was 5.05e-08; for the GWAS + imputation was 2.96e-08; and for SnipSnip was 1.09e-15. Our comparison of the three methods in these regions bears out the hypothesis that SnipSnip will provide improved evidence of association at some but not all loci. The signal at 4q13 was substantially increased, while the association at 5q35 was attenuated. Our comparison between these alternative analysis approaches will provide insights into their relative strengths and weaknesses in the context of a real study. This work was supported by U.S. N.I.H. grant R01HL093093 (P.I.: S. McGarvey).

1885M

Incorporating Functional Information in Tests of Excess De Novo Load. *Y. Jiang[1], Y. Han[2], S. Petrovski[2], K. Owzar[1], D. Goldstein[2], A. Allen[1,2].* 1) Biostatistics and Bioinformatics, Duke University, Durham, NC; 2) Center for Human Genome Variation, Duke University School of Medicine, Durham, NC.

A number of recent studies have investigated the role of de novo mutations in various disorders including epilepsy, autism, intellectual disability and schizophrenia. These studies attempt to implicate causal genes by looking for an excess load of de novos within a gene. Current statistical methods for assessing this excess are based on the underlying assumption that all qualifying mutations in a gene contribute equally to disease. However, it is well established that different mutations can have radically different effects on the ultimate protein product and, as a result, on disease. Here we propose a method that incorporates functional information in the excess load assessment. Specifically, we derive score statistics from a retrospective likelihood that incorporates the probability of a mutation being deleterious to gene function. We show that, under the null, the resulting test statistic is distributed as a weighted sum of Poisson random variables and we implement a saddlepoint approximation of this distribution to obtain accurate p-values. Using simulation, we show that our method outperforms current methods in terms of statistical power while maintaining validity. We apply this approach to four, currently available, de novo mutation datasets of neurodevelopmental/neuropsychiatric disorders: Autism spectrum disorder, schizophrenia, severe intellectual disability and epileptic encephalopathy. Our method identifies a new, potentially causal, autism gene [SUV420H1] that is not implicated by current methods. Further, for all genes that were implicated by existing methods, our approach also implicated them but with smaller p-values. Further, our analyses suggest the existence of shared genetic causes among the four neurodevelopmental/neuropsychiatric disorders investigated.

1886T

Beyond random effects meta-analysis: explaining why effect sizes differ between studies. *E. Eskin.* Dept Computer Sci, Univ California, Los Angeles, Los Angeles, CA.

Over the past few years, the aggregation of the results of many genome-wide association studies (GWAS) utilizing a technique referred to as "meta-analysis" has led to the discovery of thousands of variants implicated in hundreds of disease related traits. While most of the analysis of meta-analysis has focused on discovering additional genetic variants associated with the disease trait which were not statistically significant in any of the original studies due to low statistical power, meta-analysis opens up the door to the investigation of other interesting phenomena. In particular, it can be observed that the effect sizes of the association often differ among the studies included in meta-analysis. These differences are referred to as heterogeneity. The current method for taking into account these effect size differences is the random effects model which attempts to model the differences when combining the studies. The goal of the random effects model is to increase the power of identifying variants associated with the trait in the presence of heterogeneity. However, the presence of heterogeneity raises several questions which are not addressed by random effects models. Most importantly, what are the sources of the heterogeneity and how can they be discovered utilizing covariate information from the studies? Does heterogeneity suggest the presence of interactions? What is the correct interpretation of the results of a random effects meta-analysis? In this work, I describe a novel framework for meta-analysis which is based in the inference of causal graphs. In this framework, both fixed effects and random effects meta-analysis are just special cases of the framework. The framework allows for the construction of novel meta-analysis algorithms that can incorporate additional information about the studies leading to higher power and easier interpretability.

1887S

Association mapping from sequencing reads using k-mers. *A. Rahman, I. Hallgrímsdóttir, M.B. Eisen, L. Pachter.* University of California, Berkeley.

Association studies typically begin with the determination of SNP variants for a set of cases and controls. Genotypes are obtained either from arrays, or more recently by mapping whole-genome sequencing reads to a reference genome for variant discovery. Both of these approaches require prior sequencing of a reference genome for the organism on which the association study is to be performed. We present an alignment free method for association mapping using sequencing reads which is based on counting k-mers in sequencing reads, testing for associations directly between k-mers and the trait, and local assembly of the statistically significant k-mers. Both results with simulated data and an analysis using 1000 genomes provide a proof of principle for the approach. In a pairwise comparison of the Toscani in Italia (TSI) and the Yoruba in Ibadan, Nigera (YRI) populations we find that sequences identified by our method largely agree with results obtained using standard GWAS based on genotyping via SNP calling from mapped reads, yet also reveal novel sites missed due to indels and structural variation.

1888M

Haplotype eQTLs in response to trivalent influenza vaccine. *H. Xu[1,2], LM. Franco[1,4], RB. Couch[4], Y. Shen[5], JW. Belmont[1,4], Y. Guan[1,3,4].* 1) USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX; 2) Department of Biomedical Engineering, Southeast University, Nanjing, Jiangsu, China; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Department of Systems Biology, Columbia University, New York, NY.

Understanding how gene expressions respond to trivalent influenza vaccine and what genetic factors affect differential response among healthy individuals are of great importance to public health. In a previous study (Franco et al, 2013), a cohort of healthy individuals (225) were recruited, genotyped, and their blood gene-expression were assayed on the day before (designated as day 0) and multiple time points after (at the day 1, 3, and 14 respectively) the trivalent influenza vaccination. Here we reanalyze the data using a new method and report novel findings.

The novel method combines the multi-phenotype analysis (Stephens, 2013) and the haplotype association method (Xu and Guan, 2014), developed particularly for the data set. First, we jointly analyze the four time points simultaneously, treating expressions of a gene over time as correlated phenotypes (Stephens, 2013), which empowers us to detect expression quantitative loci (eQTL) responding to vaccination. Second, in addition to single SNP analysis to detect snp-eQTLs, we focus on haplotype analysis to detect hap-eQTLs, which is more powerful when there exists allelic heterogeneity (Xu and Guan, 2014).

The multi-phenotype analysis produced 425 genes with cis (defined as 1M base pairs up- and down-stream) hap-eQTLs that have Bayes factor > 10^6. These genes are enriched in alternative splicing ($p<6 \times 10^{-4}$) and splicing variants ($p<2\times 10^{-3}$). As expected, these genes are also enriched in immune response ($p<2\times 10^{-3}$), antigen presenting and processing ($p<2\times 10^{-7}$), and positive regulation of immune system process ($p<2\times 10^{-3}$, all p-values are Fisher's exact test obtained from DAVID). As a baseline for comparison, we obtained snp-eQTL for each day, and used the max Bayes factor (among four days) to declare significant cis-acting snp-eQTL (same threshold of 10^6). There are 60 genes that can only be discovered by the haplotype analysis, and 77 genes that can only be discovered by the multi-phenotype analysis.

We also discovered a single trans-acting hap-eQTL with Bayes factor of 10^{16}: haplotype variants in an interferon receptor gene affecting expression of an olfactory receptor, whose probe can be uniquely mapped on a different chromosome.

1889T

Novel gene discovery through proximity clustering of de novo mutations in rare diseases. *J. McRae, M. Hurles on behalf of the DDD study.* Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom.

Rare developmental disorders affect children and generally have a genetic basis. It can be difficult to identify pathogenic genetic variants in children with these disorders. Genes known to be associated with rare disorders frequently contain pathogenic variants. However, additional, currently undiscovered, genes would improve prioritising variants for diagnosis. One method of identifying novel genes is through recurrent de novo mutations occurring in genes across independent affected families. Often novel genes are identified as genes enriched with de novo mutations, where, given the gene length and mutation rate, the number of recurrent mutations significantly exceeds expectations. This work provides another method to identify novel genes. We examined recurrent de novo mutations from 1133 trios with developmental disorders. Genes known to be associated with developmental disorders were excluded from analysis. De novo variants were examined for their tendency to cluster within the coding sequence of genes, by comparing the proximity of observed de novo sites to a distribution from simulated de novos. The chance of selecting a site within a gene was weighted by mutation rates for the local sequence context and the consequence of the mutation. From a set of 84 genes with recurrent de novo mutations in 1133 trios, 16 genes showed significant clustering ($P < 0.05$). A meta-analysis that combined de novo mutations from the 1133 trios with de novo variants from independent studies of rare diseases found 35 out of 283 genes showed significantly clustering. When results from proximity clustering results were combined with results from gene enrichment, 12 genes were classified as novel genes in developmental disorders, adding diagnoses that would otherwise be unobtainable.

1890S

Power of paternity exclusion with DNA markers and its current use: Some corrective actions. *R. Chakraborty, M.R. Nolan.* Department of Molecular and Medical Genetics, Univ N Texas Health Sci Ctr, Fort Worth, TX.

Utility of a DNA marker for paternity testing is judged by evaluating how often a random person can be excluded as a biological parent of a child (with or without having the data on the mother of the child). This can be evaluated for every mother-child combination as well as for the locus on an average. Called probability of exclusion (*PE*), for an autosomal locus with codominant alleles, algorithms for computing expected *PE* for a multiallelic locus has been developed in terms of allele frequencies at the locus almost forty years back. Subsequently, such algorithms have been simplified giving average *PE* at a locus as a function of sum of powers of allele frequencies. While such algebraic results use the assumption of independence of alleles within the locus and codominance of alleles, alternatively average *PE* has also been approximated in terms of heterozygosity (*H*) and homozygosity ($h = 1 - H$) at the locus, given by $PE_{avg} = h^2 . [1 - 2hH^2]$. Although both methods of evaluation of PE_{avg} is based on the assumption of no mutation at the locus, there is a tradition of using PE_{avg} to evaluate paternity index (*PI*) when the probable father shows inconsistencies of transmitting the allele(s) he is obligated to pass on to the child due to meiotic mutation (e.g., $PI = m/PE_{avg}$, m is the locus mutation rate). With data on autosomal STR loci currently in use in paternity testing, we show that homozygosity-heterozygosity (*H-h*) based value of PE_{avg} can be considerably different from the one evaluated from the powers of allele frequencies (e.g., 41.0% as opposed to 46.6% for the D13S317 locus in African-Americans, and 29.9% versus 36.1% for the TPOX locus in US Hispanics). More importantly, the claim that the *H-h* based value of PE_{avg} is a valid approximation for loci having nearly equal frequencies of all alleles is shown to be grossly incorrect. Further, we show that computation of *PI* in cases where the paternity exclusion is ascribed to mutation should not be based on PE_{avg}, since the specific mother-child data is not used, and the logic is inconsistent with the computation of *PI* where allelic transmission between parents and the child are fully consistent. In summation, although we do not recommend the use of PE_{avg} to evaluate the statistical strength of paternity testing results, current discussions on computational aspects of PE_{avg} should be revised based on the present numerical validation study of this research.

1891S

Genetic ancestry associated with obesity and diabetes risks in a Mexican-American population from Houston, Texas. *H. Hu, C.D. Huff, Y. Yamamura, X. Wu, S.S. Strom.* Epidemiology, UT MD Anderson Cancer Center, Houston, TX.

The Mexican-American population is an admixed ethnic group with varying proportions of Native American, European and African ancestries. African Americans and Native Americans have higher incidence rates of obesity and diabetes compared to Europeans due to a combination of environment and genetic factors. However, the relative contribution of genetic ancestry is not clear. In this study, we focus on a well-defined Mexican-American population in Houston, Texas, which is a socioeconomically homogeneous but genetically diverse population, making it ideal for studying the genetic epidemiology of inherited diseases. We genotyped 96 ancestry informative markers in 4,817 individuals, and used ADMIXTURE to estimate the composition of genetic ancestry in each individual. The mean proportions of Native American, European and African ancestries among all individuals were 62.9%±0.3%, 30.7%±0.3%, and 6.4%±0.1%, respectively. Multivariate logistic regression models were constructed. In females, we found that African ancestry was significantly associated with obesity after controlling for confounders. Specifically, a 5% increase in African ancestry corresponded to a 12% increased risk of grade 2 obesity (BMI between 35 and 40; p=0.037) and an 18% increased risk of grade 3 obesity (BMI>40; p=0.006). In contrast, Native American ancestry was associated with increased risk of being overweight (BMI between 25 and 30; p=0.031), while its association with grade 1 obesity was positive but non-significant. In males, we did not observe any significant association between genetic ancestry and obesity (p>0.1). Our preliminary analysis on the risk of diabetes identified a positive and significant correlation between Native American ancestry and diabetes risk in females, after controlling for age, socioeconomic status, physical activity, smoking, alcohol consumption, and BMI (p=0.018). We estimated that 10% increase in Native American ancestry increases the risk of diabetes by 0.7%. We found no association between African ancestry and diabetes risk. In conclusion, our study suggests that among the Mexican-African population, African ancestry is a strong risk factor for obesity, and Native American ancestry is a modest risk factor for overweight and diabetes. Interestingly, both associations were only observed in females, indicating that the increased risk could be related to the lifestyle characteristic of females in this population.

1892M

Investigating European admixture in GERA East Asians. *Y. Banda[1], M. Kvale[1], T. Hoffmann[1,3], S. Hasselson[1], D. Ranatunga[2], L. Walter[2], C. Schaefer[2], P. Kwok[1,4], N. Risch[1,2,3].* 1) Institute Human Genetics, University California San Francisco, San Francisco, CA; 2) Kaiser Permanente Northern California Division of Research, Oakland, CA; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA; 4) Cardiovascular Research Institute, University of California, San Francisco, CA.

A number of analyses have been performed to understand East Asian genetic substructure. Most of these studies have utilized samples with 100% East Asian ancestry and not much has been done concerning European admixture in these groups. We describe the genetic structure in a sample of 7,520 subjects of East Asian ancestry from Northern California who are part of the Genetic Epidemiology Research on Aging (GERA) cohort. We observe individuals who are 100% East Asian, but also individuals with mixed East Asian-European ancestry. Subjects of Filipino ethnicity have a more continous European ancestry proportion distribution compared to other subjects, having, on average, a minimum of 5% European ancestry proportion. For these subjects we analyse expected allele frequencies given the admixture process and also calculate timing of the admixture events that gave rise to the current genetic structure of Filipinos. Analyses of some Y-chromsome loci also give further evidence of European admixture in Filipinos.

1893S

A Comparative Genomic Approach to Introgression Detection. *L. Nakhleh, K. Liu, J. Dai, K. Truong, Y. Song, M. Kohn.* Computer Science, Rice University, Houston, TX.

One outcome of interspecific hybridization and subsequent effects of evolutionary forces is introgression, which is the integration of genetic material from one species into the genome of an individual in another species. The evolution of several groups of eukaryotic species has involved hybridization, and cases of adaptation through introgression have been already established. In this work, we report on PhyloNet-HMM - a new comparative genomic framework for detecting introgression in genomes. PhyloNet-HMM combines phylogenetic networks with hidden Markov models (HMMs) to simultaneously capture the (potentially reticulate) evolutionary history of the genomes and dependencies within genomes. A novel aspect of our work is that it also accounts for incomplete lineage sorting and dependence across loci. Application of our model to variation data from chromosome 7 in the mouse (Mus musculus domesticus) genome detected a recently reported adaptive introgression event involving the rodent poison resistance gene Vkorc1, in addition to other newly detected introgressed genomic regions. Based on our analysis, it is estimated that about 9% of all sites within chromosome 7 are of introgressive origin (these cover about 13 Mbp of chromosome 7, and over 300 genes). Further, our model detected no introgression in a negative control data set. We also found that our model accurately detected introgression and other evolutionary processes from synthetic data sets simulated under the coalescent model with recombination, isolation, and migration. Our work provides a powerful framework for systematic analysis of introgression while simultaneously accounting for dependence across sites, point mutations, recombination, and ancestral polymorphism.

1894M

Rapid radiation of common Eurasian Y chromosome haplogroups occurred significantly later than the out of Africa migration. *M. Järve, International Consortium of the Estonian Centre for Genomics.* Estonian Biocentre and Department of Evolutionary Biology, University of Tartu, Tartu, Estonia.

Human genetic diversity outside Africa is low, which is commonly ascribed to a recent out of Africa bottleneck and subsequent rapid colonization of the rest of the world. Previous studies of the male-specific Y chromosome have shown that haplogroups common throughout non-African populations all coalesce to a small number of shared ancestral lineages, the branching order of which is only partly understood. Using 475 high coverage whole Y chromosome sequences, including 317 newly reported here, we selected reliable regions within the Y chromosome based on coverage analysis, mappability and sequence class. Based on these data, we refined the Y chromosome haplogroup tree, applying phylogenetic methods to establish the branching order and temporal dynamics of splits in non-African Y chromosome haplogroups. Compared to the length of the branches that separate African and non-African diversity, the internal branches distinguishing continental and sub-continental differences outside Africa are generally short, consistent with the model of a rapid initial colonization of Eurasia and Oceania. Following the split between African and non-African haplogroups [90 KYA (95% CI: 87-94 KYA)], the differentiation of South and Southeast Asian haplogroups H, S, M, and C did not begin until around 43 KYA, and haplogroups N and R, widely spread among Northeast and Northwest Eurasian populations, started to diversify significantly later [17 KYA (95% CI: 16-19 KYA) and 26 KYA (95% CI: 25-28 KYA), respectively]. Many major phylogenetic groups in different geographic regions seem to originate from the period around 50 KYA.

1895S

Ancestral Components of Admixed genomes of Chileans from Northern and Central Chile - The ChileGenomico Project Release 1. *R.A. Verdugo[1], A. DiGenova[2], L. Herrera[1], M. Acuña[1], M. Moraga[1], S. Berríos[1], E. Llop[1], D. Digman[1], M.L. Bustamante[1], C.Y. Valenzuela[1], F. Caba[3], E. Barozet[4], M. Villalón[5], S. Alvarado[5], D. Cáceres[5], A.M. Naranjo[6], K. Salgado[6], P. Portales[7], A. Moreno-Estrada[8], C.R. Gignoux[8], C.D. Bustamante[8], C. Eng[9], S. Huntsman[9], E.G. Burchard[10], N. Loira[2], A. Maass[2,11], L. Cifuentes[1].* 1) Programa de Genética Humana ICBM, Facultad de Medicina, Universidad de Chile, Santiago, RM, Chile; 2) Mathomics, Centro de Modelamiento Matemático y Centro para la Regulación del Genoma, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile; 3) Facultad de Salud y Ciencias de la Actividad Física, Universidad Sek, Santiago, Chile; 4) Departamento de Sociología, Facultad de Ciencias Sociales, Universidad de Chile, Santiago, Chile; 5) Instituto de Salud Poblacional "Escuela de Salud Pública", Universidad de Chile, Santiago, Chile; 6) Facultad de Ciencias de la Salud, Universidad de Tarapacá, Arica, Chile; 7) Corporación Municipal de Desarrollo Social, Departamento de Salud, Iquique, Chile; 8) Department of Genetics, Stanford University, Stanford, California; 9) Department of Medicine, University of California, San Francisco, California; 10) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California; 11) Departamento de Ingeniería Matemática, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile.

The Human Genome Diversity Project, the Hapmap Project, and the 1000 Genomes Project (1000G) have provided valuable surveys of Human genetic diversity worldwide. This has allowed modeling the structure of populations originating admixture from different continents. However, sub-continental components inference is limited for South America; there is currently no population southern from Lima in any of these projects. The ChileGenomico project aims to close this gap for the Chilean population by creating a detailed characterization of its genetic structure. We collected informed-consented DNA from over 3000 volunteers born in 208 districts from Arica to Punta Arenas, i.e. 4000 km of continental Chile. Our first data release includes whole-genome sequencing and genotyping with the Axiom Genome-Wide LAT 1 Array for individuals selected for putative indigenous ancestry: 9 from the north (Aymara) and 9 from the south of Chile (Mapuche). Sequencing at the National Center for Genomics and Bioinformatics of Chile gave a 3.6x average, covering 88% of the genome. We identified 4,716,457 SNPs (260,367 novel) and 544,079 INDELs (266,973 novel) using GATK. Precision (0.98), recall (0.74), false discovery rate (0.03), and accuracy (0.83) for our SNPs predictions. Genotype imputation with 66 Mexican and 60 Colombian sequenced at low coverage by 1000G increased accuracy to 98%. Global ancestry was inferred with ADMIXTURE and a panel of 30 CEU, 30 YRI, and 43 AMR unrelated individuals from 1000G. The average ancestry proportions for the Chilean sample were 20% (European), 1% (African), and 78% (Amerindian). The Amerindian percentage was higher in individuals with Aymara ancestry (82%) than those with Mapuche ancestry (74%). Two individuals had more than 97% AMR component, indicating that they were non-admixed. The Fst between the Amerindian and Europeans components (k=3) was 0.128 and between the Aymara and Mapuche components (k=5) was 0.049, i.e. 32 times larger than Fst for German vs Spanish. Local ancestry inference was done with PCAdmix using 411,674 SNPs that allowed inference in 3,259 non-overlapping genomic windows. Pseudo-Amerindian genomes were generated by removing regions of European or African origin, used to select optimal sets of Ancestry Informative Markers for the Chilean population. All results will be made public through the ChileGenomico database that integrates the GBrowse and BioMart tools with custom web reports (http://chilegenomico.uchile.cl).

1896M

High degree of admixture in an urban Brazilian population. *M.B. Melo[1], G. Ananina[1], M.A. Bezerra[2], A.S. Araujo[2], P.R.S. Cruz[1], M. Simioni[3], V.L. Gil da Silva Lopes[3], F.F. Costa[4].* 1) CBMEG, Univ Campinas, Campinas SP, Brazil; 2) Haematology and Hemotherapy Center of Pernambuco/HEMOPE, Recife, PE, Brazil; 3) Department of Medical Genetics, FCM, Univ Campinas, Campinas SP, Brazil; 4) Hematology and Hemotherapy Center/HEMOCENTRO, Univ Campinas, Campinas SP, Brazil.

The Brazilian population is the largest Latin American population rather distinct from others by their demographic and cultural history; however, it is still poorly studied genetically. Absence of a detailed genetic profile is an obstacle for many genetic studies. Here we report the results of our genome-wide study of a sample of the Brazilian population. Subjects (n=62) were recruited at two Brazilian cities: Campinas (SP) and Recife (PE). This study was approved by the Ethics Committee (FCM-UNICAMP, Campinas, SP, Brazil) in accordance to the Basic Principles of the Declaration of Helsinki. The genotyping protocol was carried out using Affymetrix® Genome-Wide Human SNP 6.0 Array (Affymetrix Inc., CA, USA) following manufacturer's recommendations. Genotyping Console v.4.1 was employed to obtain genotype calls. Publically available population data (15 populations) of HapMap and 1000 Genomes Project were used for comparisons. All SNPs were coerced to the common for all populations list and quality control filters were applied per-individual and per-marker: exclusion of related individuals and individuals with high genotype missing rate (>5%); exclusion thresholds for a SNP - MAF<5%, HWE P-value<1e-6, per SNP call rate <5%, Mendelian errors in any of the population under consideration. Basic manipulations of the data were performed with the aid of PLINK software; population structure and admixture were analyzed applying EIGENSOFT and ADMIXTURE softwares. All Brazilian samples passed inclusion control. We started with 861,474 SNPs successfully genotyped in the Brazilian cohort. The coerced list included 840,856 markers. After pruning procedure, we obtained 297,115 markers. PCA analysis was applied to the groups of 42-45 individuals (randomly picked up from the largest populations) demonstrating that Brazilian samples are gradually distributed between the populations of European and African origins. Admixture analysis identified 7 groups in our dataset revealing a highly admixed pattern for the Brazilian sample. The Brazilian population is highly admixed, challenging for genetic analysis. Studies of extended sample size are desirable for a better characterization of the degree of admixture in the whole population and in regional groups. Financial support: CNPq, Brazil (8367/2011-1, 150398/2013-1); FAPESP, Brazil (2008/57441-0).

1897S

Whole genome association and genetic admixture analysis of EEG phenotypes in a Native American community sample. *Q. Peng[1], N.J. Schork[1], K.C. Wilhelmsen[2], C.L. Ehlers[3].* 1) J. Craig Venter Institute, La Jolla, CA; 2) University of North Carolina, Chapel Hill, NC; 3) The Scripps Research Institute, La Jolla, CA.

PURPOSE. Resting EEG features are shown to be genetically influenced. EEG patterns, such as low & high voltage alpha activity, remain highly stable over most of an adult lifespan and have been associated with phenotypes such as addictions. Genetic studies of complex phenotypes such as EEG & addictions can be advantageously conducted in well-defined populations such as Native Americans (NA). This work is part of a larger study exploring risk factors for substance dependence in a NA community. We previously reported that EEG alpha power is highly heritable in this community (h2 = 0.62 frontal, 0.67 posterior scalps). We also observed that low voltage alpha (LVA) was correlated with higher fronto-parietal connectivity (FPC). The current study aims to: 1) assess ancestry admixture influence on the EEG LVA & FPC; 2) conduct GWAS on the phenotypes in this NA population (N=612). **METHODS.** EEG was transformed using fourier analyses and the alpha power determined. Wavelet analysis was used to determine the degree of phase locking between frontal (FZ) & parietal (PZ) electrodes, a measure of connectivity. Blood derived DNA was sequenced using Illumina low-coverage WGS. Ancestry & degree of admixture were estimated with a genome-wide panel of SNPs. Multiple linear regression was used to assess the relation between NA ancestry degree and EEG LVA & FPC. A variance component approach was used in the association test to control for admixture and familial relatedness. **RESULTS.** FPC showed significant correlation with NA ancestry (β=0.05±0.02, p=0.02). LVA was negatively associated with NA (β=-30.1±10.1, p=0.003). A variant downstream of *HEXIM1* was negatively associated with FPC with genome-wide significance (β=-0.07, p=4.5e-10). Nearly 20% of the NA samples were heterozygous on the variant, yet it is absent from dbSNP & 1000Genomes, suggesting its uniqueness to the NAs. Analysis of the ancestral background of the genomic region around the variant showed it resides in a region of lower NA. Of individuals carrying the variant, FPC was uncorrelated with local NA. Of those with the wild-type genotype, FPC was significantly correlated with local NA (β=0.11±0.05, p=0.004), suggesting a polygenic background effect of the NA on the phenotype. **CONCLUSIONS.** Overall a higher NA ancestry degree is associated with higher FPC & negatively associated with LVA. A likely NA-specific variant was found to be significantly negatively associated with FPC. (Supported by AA10201, DA030976).

1898M

A Genome Wide Admixture Association Study of a Sleep Disturbance Phenotype in Adults with Sickle Cell Anemia. *C. Liu[1], Z. Wang[1], L. Diaw[1], A. Oguhebe[1], C. Okungi[1], K. Vaughan[1], D. Darbari[1,2], J. Taylor[1].* 1) Genomic Medicine Section, Hematology Branch, NHLBI, NIH, Bethesda, MD; 2) Center for Cancer and Blood Diseases, Childrens National Medical Center, Washington, DC.

Population admixture is a variable which can lead to spurious associations in genome wide association studies. However, mapping by admixture linkage disequilibrium (MALD) is statistically advantageous for genome wide studies in admixed, rare disease populations. Sickle cell disease (SCD) is a Mendelian disease caused by mutation of beta-hemoglobin. SCD has systemic manifestations including hemolytic anemia, recurrent episodes of severe pain and premature mortality. Our recent work has shown that more than 70% of SCD adults have sleep disturbance defined by the Pittsburgh Sleep Quality Index (PSQI). A significant independent risk factor for sleep disturbance in SCD is more frequent hospitalizations for SCD pain crises. Prior twin and sibling studies in the general population suggest that sleep disturbance is a moderately heritable trait, although no data is available to estimate heritability in SCD. We hypothesized that sleep disturbance in SCD is a complex trait attributable to modifier genes that can be identified by MALD. To test this hypothesis, we identified ancestry informative markers (AIMs) from 3,804,602 SNPs genotyped by HapMap Phase III CEU and YRI populations. These SNPs were combined with a published admixture panel, yielding a new MALD panel of 1795 AIMs. We genotyped these AIMs in 221 HapMap subjects and 489 adults from the Bethesda Sickle Cell Cohort Study using Illumina iSelect arrays. Using our HapMap genotype data, this admixture panel has a mean Δ (difference in allele frequencies between populations) of 0.707 (SD 0.114) with an average inter-marker distance of 1.692 Mb. We then performed a MALD association study of sleep disturbance, dividing them into cases with PSQI scores >5 or controls with lower PSQI scores. We identified a region on chromosome 13p21 approaching genome wide significance (defined by a locus genome score or LGS of 5 or more) and suggestive of a sleep disturbance association (locus genome score 4.007 at rs17088390). SNPs with a case control statistics >2 span a region of approximately 4 Mb on chromosome 13, including 1 within the DACH1 gene. Ongoing work is incorporating fine mapping and modeling with known sleep disturbance variables. Overall, our results suggest that MALD may be advantageous for SCD association studies due to a high degree of population admixture and modest sample sizes due to the rarity of SCD.

1899S

Use of Long-read-sequence Aided Phasing for Inference of Ancestry Assignment in Admixed Populations. *F.L. Mendez[1], S.S. Shringarpure[1], A. Moreno[1], E.R. Martin[2], M.L. Cuccaro[2], C.D. Bustamante[1].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Center for Genetic Epidemiology and Statistical Genetics, University of Miami, Miami, FL.

Correct phase reconstruction of individual chromosomes is important for numerous genetic analyses, including inferring demographic parameters in admixture processes. Admixture is the result of interbreeding of previously differentiated populations. The chromosomes of admixed individuals are composed of segments that can be traced individually to one of the ancestral populations. The abundance and length distributions of these chromosomal segments provide crucial information on the admixture process; however, the correct inference of their length and ancestry requires phasing data from the admixed individuals. Phase reconstruction can be performed applying the rules of Mendelian segregation of variants or using statistical methods that rely on population data. Alternatively, molecular phasing (the observation of different polymorphisms in the same chromosomal sequence) provides direct evidence of phase. In this fashion, methods of molecular-phasing, like long-read sequencing, may be used to extend the range of confident phasing. We simulate long-read sequence data and explore improvement brought by molecular phasing on accuracy of ancestry assignment of chromosomal segments, in definition of segment boundaries, and in overall admixture inference. We then use genotype information from 5 trios of Latino individuals, including 10 individuals with long-read sequence information, together with three reference panels of haplotypes of European, African, and Native American ancestry to evaluate the effect of molecular phasing on inference of the admixture process.

1900M

Characterizing the Local Ancestry of Established Multiple Sclerosis Risk Loci in Hispanics. *A.H. Beecham[1,2], Z. Liu[1,2], P. Antoine-Gourraud[3], P. Manrique[1], A. Hadjixenofontos[1], I. Konidari[1], P. Whitehead[1], D. Seo[4], C. Cotsapas[5], N.A. Patsopoulos[6], P. De Jager[6], J.L. Haines[7], M.A. Pericak-Vance[1,2], G.W. Beecham[1,2], S.R. Delgado[8], J.R. Oksenberg[3], J.L. McCauley[1,2].* 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) Dr. John T. Macdonald Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA; 3) Department of Neurology, University of California at San Francisco, San Francisco, CA, USA; 4) Division of Cardiology, Miller School of Medicine, University of Miami, Miami, FL, USA; 5) Program in Medical and Population Genetics, Broad Institute of Harvard University and MIT, Cambridge, MA, USA; 6) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham & Women's Hospital, Boston, MA, USA; 7) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA; 8) Multiple Sclerosis Division, Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA.

Multiple sclerosis (MS) is an inflammatory demyelinating disorder of the central nervous system with variable prevalence across populations. European populations have a higher prevalence and have traditionally been considered more susceptible to MS than either Hispanic or African. This could be due in part to a greater genetic risk in European populations. Through genetic association studies in individuals of European descent, 110 established MS risk variants in 103 discrete loci outside of the Major Histocompatibility Complex have been identified. Our goal is to characterize the local ancestry at these 110 risk variants, first in a sample of European descent (1691 individuals affected with MS and 5542 controls), and second in a Hispanic sample (200 individuals affected with MS and 1000 population controls). We hypothesize that given the increased prevalence of disease in European populations; there will be an increased percentage of European ancestry in Hispanics affected with MS versus Hispanic population controls at these 110 risk variants. Individuals affected with MS and the controls of European descent have been genotyped using the ImmunoChip genotyping array. The Hispanic population controls have been genotyped using the Affymetrix 6.0 genotyping array. Analyses of a risk score using the 110 risk variants show a difference in risk between Hispanics affected with MS and population controls (p=3.2E-05), indicating that these risk loci are also important in Hispanics. We do not see a significant difference in the risk score between Hispanics and Europeans affected with MS (p=2.0E-01), in support of the hypothesis that the ancestry at these loci is similar amongst cases from both populations. Haplotype phasing based on the reference populations of European, African, and Asian individuals from the 1000 Genomes Project was done using SHAPEIT. With LAMP-LD/LAMP-ANC, we are currently computing local ancestry using a representative subset of possible haplotype pairs and reference populations of Europeans and Africans from the 1000 Genomes Project as well as Native Americans from the Human Genome Diversity Project. Ancestry percentages at each locus will be compared, between individuals affected with MS and controls for samples of both European and Hispanic descent in order to better characterize the local ancestry of these confirmed loci both in the population in which they were discovered and a genetically admixed population.

1901S

Linear Mixed Model-Based Admixture Mapping. *L. Brown, T. Thornton.* Biostatistics, University of Washington, Seattle, WA.

Genetic studies in recently admixed populations can provide invaluable insight into novel risk factors contributing to disease. Population admixture results in combined genomes from previously isolated ancestral populations that may have discernible allele frequency differences due to natural selection and genetic drift. Genes that underlie ethnic differences in traits and that show differential risk by ancestry can be identified using admixture mapping. Compared to studies carried out in more ethnically homogenous populations, admixture mapping has potentially greater power to detect certain genetic variants. Linear mixed models have gained traction as a tool for genome-wide association studies. Mixed model methods have been shown to protect against spurious associations in structured samples, a common pitfall in genetic association studies, by directly accounting for sources of dependence including cryptic relatedness and population stratification. We present a linear mixed model approach for admixture mapping in the presence of population structure and hidden relatedness. We implement this method using local ancestry estimates based on genome-wide SNP data. We apply the method to analyze genetic associations with white blood cell count and C-reactive protein phenotypes in the African American cohort of the Women's Health Initiative study. We demonstrate that our proposed linear mixed model method for admixture mapping provides a substantial improvement over widely used admixture mapping approaches.

1902M

Genetic evidence of archaic admixture in India. *A. Basu, D. Das, S. Das, N. Biswas.* National Institute of BioMedical Genomics, Kalyani, India.

Comparing high-coverage Denisova and Neanderthal whole-genome sequences has revealed significant admixture with all present day non-African populations. Microblade tool usage from central-India has been reported, yet no genetic-study examined archaic admixture in present day South-Asians. We report the first evidence of archaic admixture from whole genome sequence data of 4 present-day Indians. Four individuals, who are at the extremities of a two-dimensional Principal-Component plot summarizing the extant of genomic variance in 237 Indians belonging to 20 linguistically, ethnically diverse populations sampled from different geographic locations in India; were sequenced. Their population identities are Onge, Jamatia, Panniya and Birhore. All individuals showed slight excess of Denisova admixture (D-Statistic 1.6-1.99) when compared with Eurasians, with admixture evidence increasing from Jamatia(1.6) to Onge(1.99). Similar pattern was also observed when compared with Neanderthal. Our findings show evidence of archaic admixture in different present-day populations, not restricted to proximity of archaeological evidence, indicating wide spread admixture.

1903S

Pharmacogenomic patterns for Brazilian and Mexican populations. *V. Bonifaz[1], A.V. Contreras[1], C.J. Struchiner[2], R.A. Roella[3], T.K. Mazotti[3], R. Chammas[3], M.J. Gomez-Vazquez[1], L. Uribe[1], C. Rangel-Escareño[1], H.L. McLeod[4, 5], A. Hidalgo-Miranda[1], E.J. Parra[6], J.C. Fernandez-Lopez[1], G. Suarez-Kurtz[7].* 1) Instituto Nacional de Medicina Genómica, Mexico, Mexico City, Mexico; 2) Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brazil; 3) Laboratório de Oncologia Experimental—LIM24, Departamento de Radiologia e Oncologia, Faculdade de Medicina, Universidade de São Paulo, SP, Brazil; 4) DeBartolo Family Personalized Medicine Institute, Tampa, FL. U.S.A; 5) Pharmacogenetics for Every Nation Initiative, Tampa, FL; 6) Department of Anthropology, University of Toronto at Mississauga, Mississauga, ON, Canada; 7) Divisão de Farmacologia, Instituto Nacional de Câncer, Rio de Janeiro, RJ Brazil.

Latin America is a vast geographic region that is home to more than 600 million people. To a large extent, the current Latin American population reflects a complex history of admixture between people with ancestral roots in the Americas, Europe and Africa. Studies for pharmacogenomics-related traits are increasingly being performed to identify loci that affect either drug response or susceptibility to adverse drug reactions. However, the effect of a polymorphism can differ in magnitude or be absent depending on the population being assessed. In this work we present the potential impact of admixture on the distribution of genetic markers of pharmacogenetic and pharmacogenomics (PGx) relevance. Differences in admixture history may have an important impact in the distribution of allele and genotype frequencies at population level. Therefore we aim at characterizing the distribution of polymorphisms of PGx relevance of two of the most populous Latin American countries: Brazil and Mexico. The Brazilian sample includes 268 individuals stratified into census categories Branco (White), Pardo (Brown) and Preto (Black). The Mexican sample comprises 45 Native American Zapotecos and 224 self-identified Mestizo individuals located in geographically distant regions. Both samples were genotyped using Affymetrix Drug Metabolizing Enzymes and Transporters (DMET) Plus array. This platform includes 1936 variants, rare and common, in 231 genes involved in drug pharmacokinetics. Our results highlight the complex population history of these two countries. We observed differences in admixture proportions according to geography (Mexico) and census categories (Brazil). Although the overall genetic differentiation of the markers interrogated in the DMET Plus array is low, there are many loci that show high levels of genetic differentiation between the parental populations relevant for contemporary Brazilians and Mexicans. The *VKORC1* rs9923231, a major determinant of warfarin dose in CPIC guidelines, is an excellent example of the practical PGx implications of admixture history. The frequency of the rs9923231A allele, which associates with high warfarin sensitivity, varies 13-fold (52.2% frequency in Africa and 4.1% in Zapotecos) among proxy parental populations of Brazilians and Mexicans. Differences in admixture history among Latin American populations may have an important impact in allele frequency distribution and therefore also in haplotypes associated to drug metabolism.

1904M

The genetic ancestry of African, Latino, and European Americans across the United States. *K. Bryc[1], E. Durand[1], D. Reich[2,3,4], J. Mountain[1].* 1) 23andMe, Inc., Mountain View, CA; 2) Harvard Medical School, Department of Genetics, Boston, MA, USA; 3) Howard Hughes Medical Institute, Boston, MA, USA; 4) Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Within the past 500 years, North America has been the site of a dynamic mixing of people from populations that were previously separated by oceans and other geographic features. The interactions between Native Americans, European settlers, and Africans brought to the New World via the Trans-Atlantic slave trade shaped the early history of what became the United States. We studied the genetic ancestry of 5,269 self-reported African Americans, 8,633 Latinos, and 148,789 European Americans who are 23andMe customers living in the US and show that the legacy of these interactions is visible in the genetically-inferred ancestry of modern Americans.

We shed light on the unique regional differences in genetic ancestry within and across the United States. In addition to well-established variability in individual ancestry proportions, we detect systematic clines in the amount of African and European ancestry in African Americans from different states. We demonstrate that the pervasive mixed ancestry in modern Americans and the relationship between self-reported identity and genetic ancestry reflect regional social and political history. We find that levels of Native American and African ancestry in European Americans, Latinos, and African Americans are highly correlated with the population density of African Americans and Latinos in each state with one notable exception.

We provide evidence that a minor, but measurable, proportion of self-reported European Americans carry African ancestry. Likewise, we demonstrate that an appreciable fraction of European Americans carry Native American ancestry. In European Americans, we find strong differences in European subpopulation ancestry across the US, which are consistent with known major historical migrations from different regions of Europe. The genetic diversity within each identity, combined with significant overlap of ancestry profiles between identities, highlights the complexity of using self-reported or physician-inferred ancestry in a biomedical context, since the genetic ancestry of individuals within each identity is widely variable. Our results inform the geography of historical admixture in the US, have sociological implications for self-identity, and inform the use of self-reported race and ethnicity in a medical setting.

1905S

The Brazilian EPIGEN Initiative: admixture, history and epidemiology at high resolution. *F. Kehdy[1], M. Barreto[2], B. Horta[3], M.F. Lima-Costa[4], A. Horimoto[5], N. Esteban[5], W. Magalhães[1], M.R. Rodrigues[1], G.B. Souza[1], F. Soares[1], G.S Araújo[1], T.P. Leal[1], M. Machado[1], R. Moreira[1], J.M. Sanches[5], H.C. Santos[5], M. Gouveia[1], A. Pereira[5], E. Tarazona-Santos[1], The Brazilian EPIGEN Consortium.* 1) Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; 2) Universidade Federal da Bahia, Brazil; 3) Universidade Federal de Pelotas, Brazil; 4) Fundação Oswaldo Cruz, Centro de Pesquiça René Rachou, MG, Brazil; 5) Instituto do Coração, Universidade de São Paulo, Brazil.

Brazil is the largest and most populous Latin-American country, and received the largest proportion of the African slave trade along the last five centuries (approximately 4 million of slaves vs. 700 000 in the United States). We studied three Brazilian longitudinal populational cohorts from the three most populated regions of Brazil: Salvador-Bahia (n=1309), Bambui (n= 1442) and Pelotas (n=3736) from Northeast, Southeast and Southern Brazil respectively. We genotyped the Omni2.5M-Illumina for the 6487 individuals, the Omni5.0M-Illumina for 265 individuals and sequenced 30 genomes (average coverage: 42X). While Amerindian ancestry was low (5-7% at population level, with no individual with > 30% of this ancestry), the three populations showed individuals with all possible combinations of African and European ancestry. At population level, African ancestry ranged from 14-15% in Pelotas and Bambui to 51% in Salvador. The EPIGEN cohorts also exemplify how kinship and inbreeding coefficients may be differently structured in small (Bambuí), medium-size (Pelotas) and large (Salvador) admixed Latin American populations. Model-based ADMIXTURE analysis revealed an unknown cluster of African ancestry, likely Bantu or Eastern Africa-associated more common in Southern Brazil that in any other Latin America population, probably related with slave trade from regions such as Angola-Mozambique. Genetic diversity of European origin in Brazil recapitulates a considerable portion of the whole diversity of this continent, not being restricted to the Iberian Peninsula. The distribution of chromosomal segments of distinct ancestry do show evidence of very recent European admixture into the South East and Southern Brazil but not in the Salvador cohort (NE), while African admixture into admixed populations fits a model of continuous gene flow across the three cohorts. By the genome sequencing we identified between 3.6 M and 4.4 M of autosomal SNPs per each individual, and the high levels of admixture of Brazilians allowed us to identify ~1.4 M of new autosomal SNPs. The EPIGEN Initiative is also performing several GWAS and admixture mapping studies on different complex traits, including longitudinal data. Funding: Brazilian Ministry of Health/FINEP.

1906M

Inferring patterns of demography and assortative mating in the Thousand Genomes Project admixed populations from the Americas. *E.E. Kenny[1,5,6,7], C. Gignoux[2], S. Baharian[3], S. Musharoff[2], B. Maples[2], S. Shingarpure[2], A. Auton[4], C.D. Bustamante[2], S. Gravel[3], A.R. Martin[2], The 1000 Genomes Consortium.* 1) Department of Genetics, Icahn School of Medicine Mount Sinai, New York, NY; 2) Department of Genetics, School of Medicine, Stanford University, California, CA; 3) Department of Human Genetics, McGill University, Montreal, Canada; 4) Department of Genetics,Albert Einstein College of Medicine, New York, NY; 5) Institute of Personalized Medicine, Icahn School of Medicine Mount Sinai, New York, NY; 6) Center of Statistical Genetics, Icahn School of Medicine Mount Sinai, New York, NY; 7) Institute of Genomics and Multiscale Biology, Icahn School of Medicine Mount Sinai, New York, NY.

The phase 3 release of the 1000 Genomes project includes genotype and sequence data for 2,535 individuals from 26 populations around the globe. These include six populations from the Americas with mixed Native American, European and West African ancestry. We have identified admixture proportions in these six populations, which, include African Caribbean's from Barbados (ACB), African American's from south-west USA (ASW), Colombians from Medellin (CLM), Peruvians from Lima (PEL), Mexican-American from Los Angeles (MXL), and Puerto Ricans from Puerto Rico (PUR). We show presence of Native American, European and African ancestry in all six populations, in particular, we identify six ASW individuals with > 20% Native American ancestry. The European component of these individuals looks most similar to Nordic ancestry, rather than Spanish ancestry often seen in Hispanic/Latino individuals. Among the ACB, PEL, PUR, CLM and MXL populations, we find an excess of Native American and dearth of European ancestry on chromosome X compared to the autosomes, indicating a history of non-random mating in these populations. We have also inferred local ancestry tracts (LAT), identifying haplotype specific segments of ancestry across chromosomes. We assessed the accuracy of our tract calls and demonstrated accuracies >0.99, >0.98 and >0.97 in African, European and Native American tracts across all populations. By modeling the distribution of ancestral tract lengths, we inferred the timings of migration in the two populations from the America's that are new to phase 3, ACB and PEL. We estimated the PEL have had more recent admixture with European and African individuals than other Hispanic/Latino groups in the Caribbean and throughout northern South America, consistent with known migration patterns. These analyses have given us an insight into the demographic history and migration patterns among admixed populations in the 1000 Genomes Project.

1907S

Genotype and allele frequencies of *RETN* -420 C/G polymorphism in three Mexican native populations. *A. López Quintero[1,2], AG. García Zapién[1], SE. Flores Martínez[1], S. Islas Andrade[3], MC. Revilla Monsalve[3], J. Sánchez Corona[1].* 1) Centro de Investigación Biomédica de Occidente (CIBO), División de Medicina Molecular, Sierra Mojada 800, GDL, Jal; Méx; 2) Universidad de Guadalajara, Centro Universitario de Ciencias de la Salud (CUCS), Doctorado en Genética Humana, Sierra Mojada 950, GDL, Jal; Méx; 3) Unidad de Investigación en Enfermedades Metabólicas, Centro Médico Nacional Siglo XXI, IMSS, Belisario Domínguez 1000, Cd. Mex; Méx.

Introduction. The adipokine resistin plays a significant role in glucose and energy homeostasis, it is also known that resistin increases insulin resistance. The polymorphism -420 C/G in *RETN* gene is associated with increased resistin levels and type 2 Diabetes Mellitus. Objective. To determine genotypic and allelic frequencies of *RETN* -420 C/G polymorphism in individuals from three Mexican native populations: Lacandones (Lacanjá, Ocosingo, Bethel, Nahá), Yaquis and Tepehuanos. Material and methods. The present study included 419 DNA samples of individuals from three Mexican indigenous populations. The *RETN* -420 C/G polymorphism was analyzed by PCR-RFLP. After a 331 bp fragment amplification by PCR, the products were digested with the *BbsI* enzyme restriction to identify the genotypes. The digested products were separated by electrophoresis in 6 % polyacrylamide gels and identified by silver staining. Differences between genotype (GF) and allele frequencies (AF) among the three populations were assessed by using a $\chi 2$ test with the RXC program and 10000 iterations. Results. There were statistically significant differences among the Lacandones community when assessed by genotype and allele frequencies; Lacanjá vs Ocosingo, Bethel and Nahá (p=0.0003, 0.0000 and 0.0004 for GF and p=0.001, p<0.001 and p<0.001 in case of AF). The Tepehuanos population was statistically different when compared the GF (p=0.0005), and AF (p=.0006) with Bethel community. The Yaqui population displayed differences when compared with Lacanjá, and Bethel communities (p= 0.0066 and 0.0015 for GF p=0.0023 and 0.0072 in case of AF respectively). Conclusion. The genotypic and allelic distribution of the *RETN* -420 C/G polymorphism is different among the analyzed populations. Furthermore, it is not possible to consider the four Lacandones communities as one group, and this must be considered in further association studies for the Lacandones community.

1908M

Sub-continental local ancestry inference in U.S. individual. *B.K. Maples[1,2], J.K. Byrnes[3], J.M. Granka[3], K. Noto[3], S. Shringarpure[1], M.L. Carpenter[1], M.J. Barber[3], R.E. Curtis[4], N.M. Myres[4], C.A. Ball[3], K.G. Chahine[4], C.D. Bustamante[1].* 1) Genetics, Stanford University, Stanford, CA; 2) Biomedical Informatics, Stanford University, Stanford, CA; 3) AncestryDNA L.L.C., San Francisco, CA; 4) AncestryDNA L.L.C., Provo, UT.

The United States was populated through a sequence of migratory waves including immigrants from numerous distinct source populations. This "melting pot" process has led to the majority of current U.S. residents being genetically admixed. Understanding this complex genetic diversity is of great interest to the field of population genetics and accounting for it is critical for medical genetics. Numerous methods have been developed for performing local ancestry inference (LAI) in which the ancestry of each genomic locus is estimated, but the majority of these methods are only accurate at the level of continental admixture (e.g. African Americans with ancestry from Africa and Europe). Sub-continental LAI is often more difficult as neighboring populations typically have reduced differentiation. In this study we apply the LAI method RFMix that has been shown to perform well at a sub-continental level (e.g. mixtures of Northern and Southern European ancestry). RFMix is seeded with a reference panel of samples with known origins, and then iteratively learns from a larger collection of test samples. The performance of this method greatly improves with larger reference panel and test sample sizes. Here we use more than 2,000 single-origin reference samples from Ancestry.com and 1000 Genomes, along with over 100,000 research-consented customer samples with admixed origins to train the model to perform inference on individuals with admixed European ancestry. We compare the genome-wide ancestry estimates from RFMix with pedigrees. Using pedigree data as a truth set, we tune the performance of RFMix. We then compare the performance of RFMix with results from the commonly used ancestry estimation method ADMIXTURE run in supervised mode with the same initial single-origin reference panel provided to RFMix. We also use single-origin samples to create synthetic admixed samples with known local ancestry patterns to assess the accuracy of RFMix to call individual segments in admixed Europeans. Finally, we apply the highly trained version of RFMix to the National Institute on Aging's Health and Retirement Study data. We compare county-level geographic summaries of sub-continental ancestry estimates in this data to recent U.S. census data. We find strong evidence of fine-scale population structure with certain localities showing enrichment for particular ancestries (e.g. Irish ancestry in and around Boston and Scandinavian ancestry in the Midwestern states).

1909S

Drift and selection contribute to elevated susceptibility for childhood acute lymphoblastic leukemia in individuals with Native American ancestry. *P. Nakka[1], V. Perez-Andreu[2], C.R. Najera Villagran[3], F. Antillon Klussmann[3], J.J. Yang[2], S. Ramachandran[1].* 1) Brown University, Providence, RI; 2) St. Jude Children's Research Hospital, Memphis, Tennessee; 3) Unidad Nacional de Oncología Pediátrica, Guatemala City, Guatemala.

Acute Lymphoblastic Leukemia (ALL) is the most common childhood cancer and one of the leading causes of death in children and adolescents. ALL is characterized by young age at disease onset (particularly in children between 2-5 years of age), which points to a strong genetic basis of disease susceptibility. In fact, previous genome-wide association studies on multiple ancestral backgrounds have identified multiple common variants predisposing children to ALL with relatively large effect sizes. (Treviño et al. Nat Genet, 2009, Xu et al. JNCI 2013). Interestingly, ALL is more prevalent and carries a poorer prognosis in patients with high genome-wide Native American (NA) ancestry than those of any other ancestry type (Yang et al. Nat Genet 2011). This suggests that the evolutionary history of NAs has played a role in the high incidence and inferior treatment outcome of ALL in these populations. Here we analyze germline genotype data from 233 Guatemalan individuals, many of whom identify as Native American, consisting of children with ALL and their unaffected parents, genotyped with Affymetrix SNP 6.0 and Illumina Exome arrays to better understand the genomic signatures of demographic history and selection on NAs populations ,particularly at loci that contribute to genetic predisposition to ALL. Genome-wide we find that individuals in our sample from Guatemala have 83% NA ancestry on average and 14% European ancestry on average (significantly higher and lower, respectively, compared to the admixed Americans in the 1000 Genomes). Principal component analysis confirms that sampled Guatemalans capture features of NA ancestry not seen in the admixed Americans from the 1000 Genomes. Local ancestry de-convolution of the Guatemalan samples shows that Asians in the 1000 Genomes are poor proxies for NA ancestral haplotypes. Of 6 previously-identified ALL susceptibility loci, we find evidence for adaptive selection at rs4982731 (CEBPE) with a minor allele frequency difference in the top 2% between Guatemalans and Asians and an Integrated Haplotype Score (in the top 5.25% of scores indicating selection for the derived allele). The elevated allele frequencies of other ALL risk alleles in Native Americans, in contrast, may have resulted from genetic drift. Although clinical studies have observed racial differences in the frequency of ALL susceptibility variants , this study presents the first evolutionary analysis of why this is the case.

1910M

Relationship between Glaucoma and Admixture in Postmenopausal African American Women. *R. Nassir[1,3,4], L. Garcia[2], L. Qi[2], R. Kosoy[3,4], N. Fijalkowski[5], M. Haan[6], K. Singh[5], J. Robbins[4], M. Seldin[3,4].* 1) Umm Al-Quraa University, Makkah, Westren, Saudi Arabia; 2) Department of Public Health Sciences, University of California, Davis, CA, USA; 3) Department of Biochemistry and Molecular Medicine, University of California, Davis, CA, USA; 4) Department of Internal Medicine, University of California, Davis, CA, USA; 5) Department of Ophthalmology, Stanford, CA, USA; 6) Department of Epidemiology & Biostatistics, University of California, San Francisco, CA, USA.

Different ethnic groups have different propensities for Glaucoma. To further assess the relationship between ethnicity and this phenotype, we examined whether differences in continental admixture in cross-sectional analysis for both African American and Hispanic American adult women in the Women's Health Initiative were associated with this trait. The proportion of European, Sub-Saharan African and Amerindian admixture was estimated for each individual participant using a Bayesian clustering method (STRUCTURE) and 92 SNP ancestry informative markers (AIMs). In self-identified African American women (n= 11616) there was a significant association between African Admixture and Glaucoma (OR= 1.38; 95% CI= 1.02-1.86) in age-adjusted model. However, this relationship had become not statistically significant after adjusting for socioeconomic status, hypertension, diabetes and BMI (OR= 1.24; 95% CI= 0.91-1.68). In contrast, there was a significant association between diabetes and glaucoma (OR= 1.51; 95% CI= 1.37-1.67) in age-adjusted model. This association persisted after the adjustment for admixture, socioeconomic status, hypertension, and BMI (OR=1.52; 95% CI= 1.37-1.68). In addition, lower socioeconomic status was significantly associated with higher risk of glaucoma (OR=1.07; 95%CI= 1.02-1.12). But, in a fully adjusted model, which included African Admixture, diabetes, hypertension and BMI, the association between socioeconomic status and glaucoma was marginally significant (p=0.04), while hypertension and BMI were not associated with glaucoma. This study shows that, Diabetes and socioeconomic status, but not the African admixture, were found to be independently associated with glaucoma status. These finding suggest that the high frequency of glaucoma in African Americans maybe largely due to factors other than admixture such as diabetes and socioeconomic status.

1911S

Inference of the demographic history of Japan using Approximate Bayesian Computation. *C.D. Quinto[1], K.R. Veeramah[2], A.E. Woerner[3], M.F. Hammer[3].* 1) Graduate Interdisciplinary Program in Genetics, University of Arizona, Tucson, Arizona, USA; 2) Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA; 3) Arizona Research Laboratories, Division of Biotechnology, University of Arizona, Tucson, Arizona, USA.

The genetic exchange between differentiated populations, termed admixture, has increasingly been shown to be an important process in human history. The formation of Hispanic populations in the Americas is one of the best-known examples of this phenomenon. Another important but less well-known example is the origin of modern Japanese. At least two distinct incoming migrations are known to have occurred during the prehistory of Japan. The first took place at least 10,000 years ago and established the Jōmon culture, which was characterized by a semi-sedentary hunter-gatherer way of life and one of the earliest uses of ceramics. Then, around 2,300 years ago, a second migration to the archipelago brought rice agriculture and iron, and established the Yayoi culture. The mixture of the people belonging to these two cultures is believed to have formed the ancestors of the modern Japanese population. Although archaeological records provide information about the time of arrival of the Yayoi people to Japan, the dynamics of the admixture process are still unclear. Previous genetic studies, focusing on mitochondrial DNA and the Y chromosome, have supported an admixture model for the origin of the modern Japanese population. While genome-wide data have being used to investigate this question, there are currently no studies that infer the parameters describing the dynamics of the admixture process. Part of the reason for this is that explicit population genetic modeling is problematic when utilizing genome-wide arrays because of the underlying ascertainment bias in the choice of SNPs. To address these issues, we genotype 500,000 SNPs in 282 samples from populations across the Japanese archipelago and East Asia. We then attempt to correct for the ascertainment bias by using whole genome sequencing data to approximate the discovery sample used to ascertain SNPs. We utilize the SNP genotypes from the different populations to identify ancestry blocks in Japanese samples. The distribution of these blocks provides insights about the time and proportion of admixture, and this information is used in an Approximate Bayesian Computation analysis to infer other key demographic parameters such as divergence times, migration rates and population sizes.

1912M

A Fine-Scale Comparative Analysis of Population Structure, Divergence and Admixture in in Han Chinese, Japanese and Korean Populations. *S. Xu[1], Y. Wang[1], D. Lu[1], Y. Chung[2].* 1) Population Genomics, CAS-MPG Partner Institute for Computational Biology, Shanghai, Shanghai, China; 2) Integrated Research Center for Genome Polymorphism, Department of Microbiology, The Catholic University Medical College, Socho-gu Seoul 137-701, Korea.

n East Asia, human origins and dispersals remain poorly understood and debatable. As the major ethnicities of East Asia, Han Chinese, Japanese and Korean people share many similarities in characteristics, language and culture. However, the genetic relationships, divergence times and subsequent gene flow among the three populations have not been well studied or quantitatively estimated. Here, we conducted a genome-wide study using over 900,000 single nucleotide polymorphisms (SNPs) and evaluated the population structure of 182 unrelated Han Chinese, 90 Japanese and 100 Korean individuals, and compared with 663 individuals representing 8 worldwide populations. Our analysis revealed that Han Chinese, Japanese and Korean populations have distinct genetic makeup and can be well distinguished based on the genome wide data, or a panel of ancestry informative markers (AIMs) screened from genome-wide SNPs, indicating they have been isolated for substantially long time. Interestingly, population structure is perfectly corresponding to the geographical distribution of the three populations, indicating geography was an important factor resulted in population differentiation. We identified a cline of north/south admixture, which is consistent with either a scenario of isolation by distance (IBD) or that of north/south migrations or both. We theorized that both IBD effect and migrations could have resulted in such a pattern. On the other hand, our analysis revealed patterns of admixture which occurred after initial splits of populations. We further estimated gene flows among the three populations. We concluded that the genetic structure of the present-day Han Chinese, Japanese and Korean people was shaped jointly by common origin, subsequent gene flow and local adaptation. Our results advance the understanding of the genetic relationship and population history in East Asia.

1913S

Analysis of autosomal and Y-chromosomal DNA Suggests West Asian Population Derivation from Northern Middle Eastern Populations in the post-Glacial Period. *P. Zalloua[1,2], F. Utro[3], M. Haber[1], L. Parida[3], E. Matisoo-Smith3[4], D. Platt[3].* 1) Genomics Laboratory, Gradute School, Beirut, Lebanon; 2) Harvard School of Public Health, Boston, MA, USA; 3) I.B.M. T. J. Watson Research Center, Yorktown Hgts, NY; 4) University of Otago, Dunedin 9054, New Zealand.

Analysis of Y DNA J and E haplogroups in West Asians (Georgians, Armenians, Turks, Syrians, Lebanese, Jordanians, Saudi Arabians, Yemenis, Kuwaitis) suggest expansions coming primarily from the north (Turkey, Georgia, Armenia), with an early differentiation between those who headed south along the Tigris-Euphrates, versus those who headed south along the Levantine coast. We sought to resolve whether southern variations represented evolution within separate ice age refugia, or evolved from the same northern refugia as suggested by Y chromosome data by revealing population divergence times between Saudi Arabia and Yemen versus Turkey, Syria, and Armenia that predate the post-glacial expansions. We employed IRIS to compute times for grand most recent common ancestors applied between pairs of subjects drawn from Georgians, Armenians, Turks, Syrians, Lebanese, Jordanians, Saudi Arabians and Yemenis, as well as pair-wise FST's based on the estimated times. We contrasted these results with raw SNP counts and pairwise FST's obtained from those counts. We applied MDS and hierarchic clustering to identify geographically informative relationships, and observed a clear pattern of a north-to-south gradient. Within the western Middle East, our results suggest population differentiation dates consistent with post Last Glacial expansions, with subsequent population constriction into the Fertile Crescent in the presence of admixture. Our estimates show a north-to-south differentiation time of ~24,800-18,200 y.a., well within the Last Glacial Period. However, the time of J1/J2 haplogroups splits that mark this diversion are dated by BATWING well into the Last Glacial Period, around 31kya. These results place the genetic differentiation of the autosomal genome to be a bit more recent than the J1/J2 split. Expansions into Europe show a somewhat more recent record than those into Africa, with signals that show affinities with particular Middle Eastern regions, suggesting more recent trade impacts.

1914M

A method to use control data and exploit the structure of genetic ancestry space to enhance case-control studies. *C. Bodea[1], S. Ripke[2, 3], B. Neale[2, 3], B. Devlin[4], M. Daly[2, 3], K. Roeder[1, 5].* 1) Department of Statistics, Carnegie Mellon University, Pittsburgh, PA; 2) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 4) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA; 5) Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA.

In genetic studies of common and rare variants, considerable effort and expense is required to obtain a sample of control subjects matched by genetic ancestry to the case subjects. Repositories like dbGap contain genetic data from tens of thousands of potential control samples, but these data can be accessed only with considerable effort on the part of the research team. We're exploring the use of publicly available collections of control data to create a common genetic ancestry space onto which cases and controls can be mapped independently. We employ spectral clustering to construct ancestry spaces as well as to perform projections. The base space and projected controls are then used to estimate the allele frequency surface over the ancestry space. To identify small-scale frequency variation while also borrowing strength from the entire data set we employ a combination of empirical Bayesian analysis across a hierarchical clustering of the controls and, for localized ancestry regions, a Gaussian process model of the minor allele frequency. We call our approach the Universal Control Repository Network (UNICORN), because it is designed to provide optimally matched allele frequency information for any case sample at hand.

We have performed a small-scale association test on the POPRES data (Nelson et al., Am J Hum Genet, 2008) based on simulated signal of varying risk and allele frequency and found that UNICORN delivers strongly superior results over a traditional matched control setup, even when the matched controls greatly outnumber cases. We have also replicated the results of a large (5956 cases, 14927 controls) GWAS on Crohn's disease (Jostins et al., Nature, 2012) via the UNICORN framework by computing the matched allele frequency estimates from the study's control set. This result highlights that even big studies with ample control collections can benefit from UNICORN.

1915S

Population Genomics of the South American Andean Region. *J.R. Homburger[1], A. Moreno-Estrada[1], C.R. Gignoux[1], E. Sanchez-Rodriguez[2], B.A. Pons-Estel[3], E. Acevedo[4], J.M. Cucho[4], P. Miranda[5], L. Catoggio[6], M.A. García[7], G. Berbotto[8], A. Babini[9], H. Scherbarth[10], S. Toloza[11], M. Alarcon-Riquelme[2], C.D. Bustamante[1].* 1) Department of Genetics, Stanford University, Stanford, CA, USA; 2) Centre for Genomics and Oncological Research (GENYO), University of Granada, Granada, Spain; 3) Sanatorio Parque, Rosario, Argentina; 4) Hospital Nacional Guillermo Almenara Irigoyen, Lima, Peru; 5) Facultad Medicina Occidente, Universidad de Chile, Santiago de Chile, Chile; 6) Hospital Italiano de Buenos Aires, Argentina; 7) H.I.G.A. General San Martin, La Plata, Argentina; 8) Hospital Eva Peron, Granadero Baigorria, Argentina; 9) Hospital Italiano de Córdoba, Córdoba, Argentina; 10) H.I.G.A. Oscar E. Alende, Mar del Plata, Argentina; 11) Hospital Interzonal San Juan Bautista, Catamarca, Argentina.

The South American continent has experienced multiple migration and admixture events. Here, we examine the genetic history of the Andean region using 551 individuals from Colombia, Ecuador, Peru, Chile, and Argentina genotyped on Illumina SNP arrays. Combining these data with individuals from the 1000 Genomes Project and the Population Reference Panel (POPRES), we show that the admixed individuals have varying degrees of Native American and European ancestry. We use ADMIXTURE and principal component analysis (PCA) to study the genetic ancestry of the admixed South American individuals. We show that on average the Peruvian individuals have a higher amount of Native American ancestry while the Argentinian individuals had on average the highest amount of European ancestry when compared with the other admixed South American samples. We also find that Andean indigenous groups account for the largest proportion of Native American ancestry in the South American individuals. On the other hand, the largest proportion of European ancestry in admixed individuals is from Southern Europe and the Iberian Peninsula. We aim to estimate the specific timing and the subcontinental origin of ancestral components involved in South American admixture by applying ancestry-specific PCA and tract length analysis to admixed genomes.

1916M

Fast individual ancestry inference from DNA sequence data leveraging allele frequencies from multiple populations. *O. Libiger[1,3], V. Bansal[1,2].* 1) Scripps Translational Science Institute, La Jolla, CA; 2) Department of Pediatrics, University of California San Diego, La Jolla CA; 3) MD Revolution, San Diego, CA.

Estimation of individual ancestry from genetic data is useful for the analysis of disease association studies, understanding human population history and interpreting personal genomic variation. We describe a fast method for estimating the relative contribution of known reference populations to an individual's genetic ancestry. Our method utilizes allele frequencies from the reference populations and individual genotype or sequence data to obtain a maximum likelihood estimate of the global admixture proportions using the BFGS optimization algorithm. It accounts for the uncertainty in genotypes present in sequence data by using genotype likelihoods instead of genotypes. Unlike previous methods, our method does not require individual genotype data from external reference panels and can utilize allele frequencies estimated from the analysis of homogeneous as well as admixed human populations. Simulation studies and application of the method to real datasets demonstrate that our method is 8-10 times faster than ADMIXTURE and has comparable accuracy. Using data from the 1000 Genomes project, we show that our method can estimate genome-wide average ancestry of admixed individuals using exome or low-coverage sequence data. Finally, we demonstrate that our method can be used to estimate admixture proportions using pooled sequence data making it a valuable tool for controlling for population stratication in sequencing based association studies that utilize DNA pooling.

1917S

Molecular and cytogenetic analysis of inversions in human and Great Apes. *M. Miroballo, A. De Magis, F. Antonacci.* Department of Biology, University of Bari, Bari, Italy.

In the last few years the burst of whole genome techniques allowed the detection of all kinds of unbalanced structural variations, while balanced rearrangements, such as inversions, are still hardly discovered. Several studies described inversions as a very frequent interspecific rearrangement and for their high potential power in suppressing recombination they are generally considered as one of the main genomic driving forces among evolutionary mechanisms. The prominent role of inversions coupled with the lack of high-throughput techniques for their detection, force research to take advantage of prediction methods that may imply the possibility to obtain false positives that must undergo experimental validation. In the present work-in-progress we resume a previously published study (Feuk et al., 2005) about the prediction and validation of inversions in human and chimpanzee genomes extending the analysis to the rest of the Great apes. First, we investigated by PCR and FISH a total of 19 predicted inversions in a panel of 16 samples, four per each species. For those regions showing both direct and inverted orientations in one species, we enlarged the investigations to more individuals in order to evaluate their occurrence and stratification in the different geographical areas of origin. 11 out of 19 regions were confirmed as previously published: three inversions occurred in the chimpanzee lineage, six inversions occurred in the human lineage (with one region confirmed as polymorphic) and two were false positives, showing the same orientation in the two species. One of the three previously validated inverted regions in PTR (7q21.3), was found to be polymorphic among the chimpanzee subspecies: in western chimpanzees the region was found to be inverted with a 67% allele frequency, while the eastern chimpanzees uniquely showed the direct orientation. We also characterized seven predicted inversions found to be false positives, mainly due to the presence of inverted duplications at the breakpoints of the predicted rearrangements. Further analysis will include at least three steps: (i) selection and validation of additional predicted inversion regions to increase the reliability of the predictive method; (ii) sequence analysis of the inversion breakpoints to investigate the molecular mechanisms of inversions; and (iii) study of genes mapping in the inverted regions to investigate their role in the recent evolutionary history of Great apes.

1918M

Identification of pleiotropic association signals in multiple autoimmune diseases at 2q24. *J.E. Molineros, C. Sun, S.K. Nath.* Oklahoma Medical Research Foundation, Oklahoma City, OK.

Overlapping etiological factors among the autoimmune diseases (ADs) have been suspected owing to shared clinical and immunological characteristics, and familial aggregation of multiple ADs. As a follow up to our previous study of Systemic Lupus Erythematosus (SLE) association on 2q24, where we identified three functional independent variants in IFIH1, we performed association analysis on Six additional autoimmune diseases (type 1 diabetes N=823, celiac disease N=1716, vitiligo N=1089, psoriasis N=919, rheumatoid arthritis, N = 1999 and systemic sclerosis N=833). One common control dataset (N=8833) was used to facilitate comparisons of this study. Given the different SNP densities in each dataset, missing SNPs were imputed using MACH. We focused on the genomic region including genes GCG, GCA, FAP, IFIH1 and KCNH7. We performed conditional logistic regressions to identify which SNPs were statistically independent. We were able to replicate association ($P<0.05$) of our three reported IFIH1 variants (rs1990760, rs10930046 and rs13023380) in all ADs, although not all three in all of them. Surprisingly, we identified stronger association signals originating from potassium channel gene KCNH7, that after conditional analyses explained much of the association in the region for celiac disease (rs867824 P=3.0×10-7), psoriasis (rs2389739 P=6.6×10-10), rheumatoid arthritis (rs12623464 P=9.31×10-17) and type 1 diabetes (rs1352071, P=4.8×10-3), whereas the strongest signal for systemic sclerosis originated at GCA (rs17783344 P =2.8×10-10), and at FAP in the case of vitiligo (rs13422767 P=4.9×10-6). Bioinformatic analysis of these independent variants, identified eQTLs as well as transcriptional regulatory elements. Thus, together with IFIH1, additional evidence from KCNH7 and GCA points towards the involvement of 2q24 in susceptibility for multiples ADs.

1919S

Genomic Affinities Among Different Population Groups of Jammu Region of J&K State, India. *R.K. Panjaliya[1], V. Dogra[1], J. Kour[1], D. Gupta[2], P. Kumar[1], S. Gupta[1].* 1) Human Genetic Research cum Counselling, Jammu, J & K, Jammu, India; 2) University of Nottingham, United Kingdom.

Human genomic diversity is the result of differential accumulation of genetic variations in individuals and populations throughout the evolution. The identification of such distinctive characteristics in the DNA represents the basis of human identification, genetic diversity and population genetics. The state of Jammu and Kashmir, India harbors heterogenous population groups inhabiting the different geographical regions. A little work is being carried out regarding the study of genetic diversity of the people of the state. In present study, genomic diversity study in six of the prominent population groups (Brahmins, Rajputs, Bhagats, Chamar, Gujjars and Jatt Sikhs) was carried out using ten autosomal DNA markers belonging to seven Alu insertion/deletion polymorphisms namely Alu ACE, Alu APO, Alu PV-92, Alu PLAT, Alu FXIIIB, Alu D1, Alu CD4; LPL PvuII, ESR PvuII and ESR XbaI polymorphism. Blood samples were collected randomly from 600 unrelated healthy individuals after prior consent. DNA was extracted and amplified by PCR using target specific oligonucleotide primers and finally subjected to agarose gel electrophoresis. Further, for LPL PvuII, ESR XbaI and ESR PvuII, the PCR product was subjected to restriction digestion using PvuII and XbaI restriction enzymes. Allele frequencies were used to calculate average heterozygosity. All the markers except Alu CD4 were found to be highly polymorphic with high heterozygosity values in almost all the population groups of the state. It was observed that most of the genomic diversity was attributed to individuals within the population. The study is a preliminary work on the population groups of the J&K state which may help in future work on the genetic heterogeneity in other population groups of the state which may give a genetic insight and genetic basis underlying the different genetic diseases prevailing in the state.

1920M

Acetylation of RNA Polymerase II Evolved in the Early History of Animals. *C. Simonti[1], K.S. Pollard[3], S. Schroeder[3], D. He[3], B.G. Bruneau[3], M. Ott[3], J.A. Capra[1,2].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Biological Sciences, Vanderbilt University, Nashville, TN; 3) Gladstone Institutes, University of California, San Francisco, CA.

The evolution and diversification of animals from their unicellular ancestors required the development of novel functional capabilities that enabled the differentiation and coordination of different cell types. We argue that a recently discovered posttranslational modification of RNA polymerase II—acetylation of its RPB1 subunit—arose in the early history of animals and provided a substrate for developing gene regulatory mechanisms, such as polymerase pausing, essential to animal diversification. RPB1 contains a modifiable C-terminal domain (CTD) that consists of variations of a heptad repeat sequence ($Y_1S_2P_3T_4S_5P_6S_7$). Modeling the evolutionary dynamics of CTD repeat count and sequence content across diverse eukaryotes revealed an expansion of the CTD in the ancestors of Metazoa. The new CTD repeats introduced the potential for RPB1 acetylation (acRPB1) due to the appearance of repeats with lysine at position seven. This was followed by an increase in the number of lysine-containing repeats in Deuterostomia. Mouse genes enriched for acRPB1 occupancy at their promoters and genes with significant expression changes when acRPB1 is disrupted are enriched for functions essential to the development and diversification of animals, such as growth factor response, cellular adhesion, and vascular development. Genes sensitive to acRPB1 show significant enrichment for origins near the appearance of lysine-containing repeats, and p300, the acetyltransferase enzyme known to acetylate the CTD, also appeared on the same ancestral lineage. Our combined functional and evolutionary analyses show that RPB1 CTD acetylation specifically regulates processes essential to multicellularity and suggest that this regulatory mechanism was involved in the expansion of animals.

1921S

Structural Comparison and Natural Selection of Filaggrin Gene within Primates. *V. Romero, K. Hosomichi, I. Inoue.* National Institute of Genetics, Mishima, Japan.

Human Profilaggrin gene (FLG) is produced in the outer layers of the epidermis and encodes important components of the skin. Mutations in this gene are associated with a variety of skin disorders, such as, atopic dermatitis (atopy). These mutations have a high prevalence in European and Asian population. When this mutations are analysed, a higher frequency of nonsynonymous variation is found, which is indicative of positive selection and may suggest a benefit for the organism. This benefit is described in one hypothesis called “natural vaccination”. This hypothesis suggests a heterozygote advantage in which a higher number of antigens enter through skin - during childhood- and thus serve to strengthen the immune system, as an adult. High prevalence of mutations in different populations, higher nonsynonymous variations and “natural vaccination” hypothesis suggest that filaggrin is evolving under positive selection. The aim of my study is to determine the structure of FLG in Macaque, Orangutan, Gorilla and Chimpanzee, compare it with the one in human and test their sequences for evidence of positive selection. To look for this, phylogenetic analysis; maximum-likelihood tests were used and filaggrin monomers were analysed for either concerted or birth-and-death evolutionary models. Profilaggrin coding sequence consists of a repeat region formed by identical monomers. These monomers are cleaved and function independently. The structure of the repeat region has two partial monomers (not-skin related) flanking the functional ones. This structure was found to be conserved in primates, however the number of functional monomers differs in each species. Phylogenetic analysis indicated monomers of species with long divergence time cluster together and maximum-likelihood tests demonstrated positive selected codons in different branches. Evolution of filaggrin as concerted or birth-and-death has to be evaluated since results are not conclusive.

1922M

Population genomics analysis in whole genome sequencing of 152 rhesus macaques. *F. Yu[1], C. Xue[1], M. Raveendran[1], G.L. Fawcett[1], S. White[1], D. Rio Deiros[1], R.A. Harris[1], M. Dahdouli[1], W. Salerno[1], Z. Johnson[2], E. Vallender[3], R. Wiseman[4], H.M. Kubisch[5], L. Cox[6], S. Kanthaswamy[7], D.G. Smith[7], B. Ferguson[8], J. Horvath[9], D. Muzny[1], R.A. Gibbs[1], J. Rogers[1].* 1) Human Genome Sequence Ctr, Baylor College Med, Houston, TX; 2) Yerkes National Primate Research Center, Atlanta, GA; 3) New England NPRC, Southborough, MA; 4) Wisconsin NPRC, Madison, WI, 5Tulane NPRC, Covington, LA; 5) Tulane NPRC, Covington, LA; 6) Southwest NPRC, San Antonio, TX; 7) California NPRC, Davis, CA,; 8) Oregon NPRC, Beaverton, OR; 9) North Carolina Central University, Durham, NC.

Rhesus macaques (Macaca mulatta) are themost widely studied nonhuman primate species in biomedicine. Rhesus share many fundamental biological and physiological processes with humans that make them an ideal model system for vaccine research and studying molecular mechanisms of human diseases. However the patterns of the genetic variation and the evolutionary and population genomics processes that shaped rhesus genomes have not been well studied. We applied next generation whole genome sequencing for 152 unrelated individuals (144 Indian-origin, 8 Chinese-origin) using either deep (30X) or low coverage (6X) strategies. Our analysis using SNPTools identified 51.6 million SNPs. And rigorous quality control procedure was applied, the average transition transversion ratio is ~2.2. We characterized evolutionary and population genomics from five different aspects: (1) characterization of population genetics parameters; (2) detailed demographic modeling; (3) identification of loci affected by Darwinian selection; (4) characterization of the patterns of linkage disequilibrium and recombination hotspots in comparison to human and chimpanzees; and (5) functional annotation of SNPs for prioritization and hypothesis generation. On average, Indian-origin individuals have >9.5 million variants, while Chinese animals have >12 million, representing much higher diversity than humans. Current effective population size (Ne) is estimated at 65,000 for Indian-origin animals, 79,000 for Chinese-origin, substantially higher than values for humans. Analyses also reveal dramatic demographic changes over the past 10 million years. Functional annotation in coding sequences found >250,000 mis-sense variants and >4600 stop-codon-gained mutations. In addition, about 110,000 SNPs mapped to conserved ENCODE transcription factor binding motifs. We used position weight matrices from the JASPAR database to assess these SNPs and found >25,000 candidate variants that may significantly affect TF binding, and thus gene expression. We mapped rhesus SNPs to the 4% of the genome identified as conserved across 29 mammals, and found reduced SNP density and MAF, consistent with negative selection in those regions. We also applied anumber of site frequency spectrum tests and found significant new evidence for both positive and negative selection in both coding and noncoding regions. Analyses of LD and local recombination rates are in progress.

1923S

RNA-seq analysis of endogenous retroviral elements in bovine conceptuses during the period of placentation. *S. Nakagawa[1], T. Miyazawa[2], T. Gojobori[3], K. Imakawa[4].* 1) Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Kanagawa, Japan; 2) Department of Cell Biology, Institute for Viral Research, Kyoto University, Kyoto, Japan; 3) Center for Information Biology, National Institute of Genetics, Mishima, Shizuoka, Japan; 4) Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan.

In evolution of mammals, some of essential genes for placental development are known to be of retroviral origin, as syncytin-1 derived from an envelope (env) gene of an endogenous retrovirus (ERV) aids in the cell fusion of placenta in humans. Although the placenta serves the same function in all placental mammals, env-derived genes responsible for trophoblast cell fusion and maternal immune tolerance differ among species and remain largely unidentified in the bovine species. To examine env-derived genes playing a role in the bovine placental development comprehensively, we determined the transcriptomic profiles of bovine conceptuses during three crucial windows of implantation periods using a high-throughput sequencer. The sequence reads were mapped into the bovine genome, in which ERV candidates were annotated using RetroTector© (7624 and 1542 for ERV-derived and env-derived genes, respectively). The mapped reads showed that about 18 percent (284 genes) of env-derived genes in the genome were expressed during placenta formation, and about four percent (63 genes) were detected for all days examined. We verified three env-derived genes that are expressed in trophoblast cells by polymerase chain reaction. Out of these three, the sequence of env-derived gene with the longest open reading frame (named BERV-P env) was found to show high expression levels in trophoblast cell lines, and to be similar to those of syncytin-Car1 genes found in dogs and cats, despite their disparate origins. These results suggest that placentation depends on various retrovirus-derived genes that could have replaced endogenous predecessors during evolution.

1924M

How population growth affects linkage disequilibrium. *A. Rogers.* Anthropology, University of Utah, Salt Lake City, UT.

The`LD curve'' relates the linkage disequilibrium (LD) between pairs of nucleotide sites to the distance that separates them along the chromosome. It is used to map disease genes and to search for adaptive evolution. But it also responds to the history of population size. The present research describes new theoretical results about the effect of population history. When a population expands in size, the LD curve grows steeper, and this effect is especially pronounced following a bottleneck in population size. When a population shrinks, the LD curve rises but remains relatively flat. As LD converges toward a new equilibrium, its time path may not be monotonic. Following an episode of growth, for example, it declines to a low value before rising toward the new equilibrium. These changes happen at different rates for different LD statistics. They are especially slow for estimates of σ_d^2, which therefore allow inferences about ancient population history. For the human population of Europe, these results suggest a history of population growth.

1925S

The Structure of Linkage Disequilibrium in the Recently Admixed Populations. *H. Zhang, J. Jung, B. Grant.* National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Rockville, MD.

The linkage disequilibrium (LD) in human population is found to be stronger with increasing geographic distance from Africa, which reflects the Africa origin of human history. Recently admixed populations (such as African Americans and Hispanic Americans) are more likely to harbor a larger number of genetic variants, relative to their inferred ancestral populations. However, the pattern of linkage disequilibrium in these admixed populations are not well studied. Here, we conduct an analysis of linkage disequilibrium at 659,184 single nucleotide polymorphisms (SNPs) in 924 unrelated samples from 11 Hapmap3 populations and 24 samples from Karitiana population (Native American in Brazil from Human Genome Diversity Project). African Americans (ASW) derive their genomic ancestry from African and European with an average of 77.3% African and 20.0% European ancestry. Hispanic Americans (MXL) lie on a cline of an average of 45.5% European ancestry, 42.9% Native American ancestry, 4.9% East Asian ancestry and 4.4% African ancestry. The mean of SNP based haplotype heterozygosity across the whole genome in these two admixed populations is greater than that of their major inferred ancestral populations. We further use r^2 between all possible SNP pairs in various distance classes as a measure of LD and also focus on the proportion of SNP pairs with r^2 greater than 0.8. Both of these two admixed populations show intermediate LD (as measured in r^2 and the proportion of SNP pairs with r^2>0.8), compared with their two major inferred ancestral populations. The extent of LD (r^2) in African Americans (ASW) is more closer to that in African population (YRI) in the short distance classes, while the values of LD in African Americans (ASW) is more likely to be similar to the European Americans (CEU) with the increased distance classes. The amount of LD (r^2) in Hispanic Americans (MXL) shows the similar pattern, but it is much closer to European Americans (CEU) in all distance classes. The findings on the structure of LD in admixed populations are helpful to better understand the evolution of human population and the design of the genetic association studies in admixed populations.

1926M

Statistical genetic considerations for expansion of panel of DNA markers for forensic applications: Lessons learned from the panel of 29 autosomal STR loci. *M.R. Nolan, R. Chakraborty.* Department of Molecular and Medical Genetics, University of North Texas Health Science Center, Fort Worth, TX., USA.

With increasing success of using DNA markers for forensic applications, there are several recent attempts to expand the panel of autosomal short tandem repeat (STR) markers to solve more complex forensic problems and to achieve worldwide uniformity of DNA typing to aid in forensic investigations that cross international boundaries. This research deals with the relevant statistical genetic considerations that are needed to validate genotype databases of such expanded panel of markers. Recently, the US National Institute of Standards (NIST) has published genotype data for 29 autosomal STR loci typed in 1,036 unrelated individuals belonging to four populations (African-Americans - n = 342, US Caucasians - n = 361, US Hispanics - n = 236, and Asians - n = 97). Since these loci include all markers of the CODIS 13, Identifiler®15, Globalfiler®21, and PowerPlex®Fusion 22 panels, we conducted a comparative analyses of this expanded panel against the above 4 more restrictive panel of markers. Our analyses suggest that even with the increase of the number of loci, deviations from Hardy-Weinberg equilibrium and linkage equilibrium for this expanded panel did not exceed the number of deviations expected by chance alone (in each population). This is so even in the presence of 7 pairs of syntenic loci in this expanded panel (F13B and D1S1656 on 1q, TPOX and D2S441 on 2p, D5S818 and CFS1PO on 5q, SE33 and D6S1043 on 6q, vWA and D12S381 on 12p, FESFPS and Penta E on 15q, D21S11 and Penta D on 21q). This expanded panel improves the statistical power of human identification as well as kinship analyses in terms of all commonly used statistics of DNA forensics. The increased number of loci also more explicitly confirms some of the population genetic features of STR polymorphism, not clearly seen in the four more restrictive panels stated above (e.g., negative association of F_{ST} with heterozygosity at individual loci). However, while the average F_{ST} values (across the four subpopulations in the database) for the four restrictive panels are very similar (1.66 to 1.74%), addition of 7 more loci in this expanded panel increased the F_{ST} to 2.15%. This is largely due to the inclusion of the F13B locus, which has a large inter-population variation of allele frequencies and the lowest heterozygosity (42%) in the Asian population. Further expansion of panels of forensic markers should attempt to control further increase of F_{ST} for improving their power of forensic utility.

1927S

Multidrug-resistant pulmonary tuberculosis in Mexican population. Evidence of association of HLA class II and TNF-308 G/A polymorphism. *B. Silva[1], C. Saenz[1], L.A. Bracho-Vela[1], M.A. Bermúdez de León[1], S. Said-Fernández[2], J. Granados[3].* 1) CIBIN-IMSS, Monterrey, Nuevo León; 2) Fac de Medicina UANL, Nuevo León; 3) INCMN-SZ, México.

Tuberculosis (TB) remains a global emergency. It is estimated that one third of the world's population is infected with Mycobacterium tuberculosis (Mtb), and only 10% of those infected will develop clinical tuberculosis (TB), which indicates the existence of host factors regulating disease expression. We investigated the association of human leukocyte antigens (HLA) class II and tumor necrosis factor (TNF) gene with susceptibility to pulmonary TB in Mexicans with especial emphasis on their association with drug resistance. HLA-DRB1 and DQB1 gene polymorphism were analyzed in a 100 Mexican patients with Multidrug-resistant tuberculosis (MDR-TB denotes bacillary resistance to at least isoniazid and rifampicin) and 150 ethnically matched healthy controls using PCR-SSO (Luminex). Polymorphisms in the promoter region of the tumor necrosis factor (TNF) gene at positions -308 and -238 were studied using the TaqMan® allelic discrimination assays for both groups. The frequencies of HLA-DRB1*14:06 were significantly higher in patients with MDR-TB as compared to healthy controls (P=0.05). Higher frequency of TNF-308 rs1800629 G/A genotype was observed in MDR-TB cases. The results suggest that HLA-DRB1*14:06 in combination with TNF-308 rs1800629 G/A genotype influence to MDR-TB. Further studies are needed to confirm our findings using larger number of patients with MDR-TB.

1928M

A Renewal Theory Approach to IBD Sharing. *S. Carmi[1], P. Wilton[2], J. Wakeley[2], I. Pe'er[1].* 1) Department of Computer Science, Columbia University, New York, NY; 2) Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA.

Sharing of long, identical-by-descent (IBD) genomic segments is common in many human populations. Dense genotyping of large cohorts, together with sophisticated detection tools, have transformed IBD sharing into an important and popular tool in population genetics, with numerous applications ranging from demographic inference to selection detection, pedigree reconstruction, disease mapping, imputation, and phasing. Nevertheless, many key questions in the theory of IBD sharing are still open. Here, we introduce and analyze a novel theoretical framework for the IBD process, based on renewal theory. Using our renewal approach, we were able to derive, in some cases for the first time, several key quantities, including the distribution of the number and total length of shared segments. Specifically, we consider the IBD process for a pair of chromosomes under two Markovian approximations of the coalescent with recombination, SMC and SMC'. We then propose, and justify by simulations, the renewal process, in which lengths of successive shared segments are independent. We show that previously derived results for IBD sharing under SMC emerge naturally under our renewal framework at the limit of long chromosomes. Still under the same limit, we derive new results for the expected number and total length of shared segments under SMC', which is considerably more accurate for small populations, precisely where IBD sharing is abundant. We also derive expressions (for any chromosome length but in Laplace space) for the distribution of the number and total length of shared segments, verified by simulations, with explicit results for SMC. Finally, we generalize all results to populations with variable historical size. Our results provide a unified and natural analytical framework for the study of IBD sharing, with a number of potential applications from fine-scale demographic inference to pedigree reconstruction.

1929S

Using linkage disequilibrium to refine estimates of accelerating growth in human populations. *M. Reppell[1], J. Carlson[1], S. Zöllner[1,2], The BRIDGES Consortium.* 1) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) Department of Psychiatry, University of Michigan, Ann Arbor, MI.

Correctly modeling the effective size of a population is critical to making accurate inferences about mutation and migration rates, and the strength of selective pressures. In humans, several large sequencing studies have given us novel insight into a genome characterized by an abundance of extremely rare genetic variation, consistent with a history of recent massive population growth. These large sequencing studies offer us unprecedented resolution for distinguishing between models of recent growth. To improve on conventional inference methods we propose a novel likelihood based approach that incorporates pairwise r^2, a measure of linkage disequilibrium, in addition to the site frequency spectrum. We observe that over short genetic distances, pairwise r^2 is a function of the variance in ancestral tree branch lengths, and therefore contains information about ancestral population sizes lacking from the site frequency spectrum, which is a function only of the mean total ancestral branch lengths. Using simulations we show that with large samples, the inclusion of pairwise r^2 improves the accuracy of demographic inference in populations that have undergone recent growth, relative to methods relying solely on the site frequency spectrum. We quantify how increasing sample size increases the accuracy of inferences about recent demography, and magnifies the improvement our method yields versus conventional approaches. Lastly, we apply our method to regions defined as neutral in whole genome sequence data from ~4,000 European ancestry individuals sequenced as part of the BRIDGES consortium. This dataset has ideal features for our purposes; providing both a large sample and non-coding genetic regions free from evidence of ongoing selection, a mixture unavailable from exome only or functional sequencing projects. We use a Monte Carlo method to estimate the likelihood of the observed data under a range of realistic growth models, including those incorporating continuous, accelerating, faster than exponential growth. With our data we are able to simultaneously make inferences about the mutation rate, μ, and the rate of accelerating growth experienced by the European population from which our sample is drawn.

1930M

Effect of negative selection on distribution of runs of homozygosity in outbred and consanguineous human cohorts. *K. Popadin[1,2], P. Makrythanasis[1,2], S.E. Antonarakis[1,2].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) Institute of Genetics and Genomics of Geneva, Switzerland.

Offspring of consanguineous marriages are characterized by long Runs Of Homozygosity (ROHs) identical by descent. It has been shown that long ROHs are enriched in deleterious variants and thus they are subject to negative selection. In order to investigate the selection forces acting on the consanguineous human genome we analyzed the distribution of ROHs in the offspring of 60 consanguineous families and contrasted the observed pattern with the distribution of ROHs in outbred population. Using phased parental haplotypes in our in-house recombination pipeline we produced *in silico* parental gametes and fused them randomly, without any selection, generating 1000 expected offspring per each family. In order to reconstruct a null distribution of ROHs we analyzed ROHs in the genomes of expected offspring. Comparing the pattern of distribution of ROHs between expected (simulated) and observed (born) siblings in each family we observed a significant excess of long ROHs in expected siblings. This difference suggests purifying selection which eliminates unfitted fetuses (with many long ROHs) during embryogenesis. Additionally, we analyzed a distribution of ROHs across the human genome and observed that different genes are covered by ROHs in highly non-random pattern. We demonstrated that genes rarely covered by ROHs are characterized not only by the increased recombination rate but also by high haploinsufficiency score. This implies that negative selection more effectively eliminates ROHs from DNA regions with increased coefficients of dominance and thus dominant genes in homozygous state harbor higher genetic load than recessive genes in homozygous state. In order to explain the high genetic load of dominant genes in homozygous state we discuss selection coefficients of deleterious variants as well as synergistic interaction of the variants in dominant versus recessive genes. Finally, we conclude that the estimated genetic load of different ROHs may prioritize the search of genes causing recessive autosomal diseases.

1931S

Non-visual Opsin Evolution and Implications for Human Health. *A.B. Popejoy[1,2], J.H. Thomas[3], S.M. Fullerton[1,3,4].* 1) Institute for Public Health Genetics (IPHG), University of Washington, Seattle, WA; 2) Statistical Genetics Program, Department of Biostatistics, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) Department of Bioethics & Humanities, University of Washington, Seattle, WA.

Visual photoreceptor proteins (opsins) are the key biological light receptors that allow animals to turn light into vision, and their complex evolutionary history has fascinated researchers for decades. Non-visual opsins, or biological light receptors not involved in the visual pathway, may be just as important as visual opsins and are expressed throughout the brain and central nervous system in humans. Nevertheless, the biological function and evolution of non-visual opsins in humans have been largely unexplored compared to visual opsins. In this study, we performed an in-depth phylogenetic analysis of human non-visual opsins using maximum likelihood in PhyML with 1000 bootstrap replicates on protein sequences obtained through ENSEMBL. The proteins were selected for phylogenetic analysis through a custom BLAST search using OPN5, a well-established non-visual opsin, as the query and verifying that each protein sequence identified (E-value < 1e-15) contains a lysine residue in the seventh transmembrane helix chromophore binding site, a key functional element of all photoreceptor proteins. Among the 9 distinct opsin proteins we identified in humans, 4 are well established as visual opsins: 3 cone cell opsins (OPN1SW, OPN1LW, OPN1MW) and 1 rod opsin (RHO). The 5 non-visual opsins we identified are RRH, RGR, OPN3, OPN4, and OPN5. The results of our analysis suggest that non-visual opsins are highly conserved and are under similar selection pressures as visual opsins, demonstrating the importance of further exploratory analyses to uncover their phenotypic associations.

1932M

Evolution and expression of duplicated genes in the human genome. *X. Lan[1], Y. Gilad[2], J.K. Pritchard[1,3,4].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Department of Biology, Stanford University, Stanford, CA; 4) Howard Hughes Medical Institute, Stanford University, Stanford, CA.

Gene duplication is regarded as an important driving force for the evolution of new biological functions. This study seeks to understand the mechanisms driving gene duplication events, when they occurred, and how they contributed to the evolution of tissue-specific functions. We constructed a database of duplicated genes pairs in the human genome based on the pairwise sequence similarity of protein coding genes. Using the synonymous substitution rate between a pair of duplicated genes and the conservation of these genes to their orthologs in other species, we estimated the time of the duplication event of the gene pairs. We found that pairs of genes that were duplicated earlier generally have lower expression similarity among different tissues, lower Dn/Ds and increased gene length compared to genes that were duplicated more recently. The majority of gene duplications appear to be due to segmental duplications, and a relatively small fraction appears to be due to retrotransposition. Duplicate genes tend to have more tissue-specific expression than singleton genes (ie., genes with no identified duplicates in the genome). We found that duplicated genes are significantly enriched for involvement in neurological functions compared to singleton genes; however interestingly the majority of these duplication events are relatively ancient, preceding the diversification of mammals. Similarly, the majority of gene duplicates that are primarily expressed in brain and cerebellum tissue arose before the divergence of mammals. Finally we describe patterns relating to classic models of the evolution of gene duplicates, including loss-of-function, neofunctionalization and subfunctionalization.

1933S

Can phylogenomic analysis of Hemopexin repeat-containing proteins provide insights into the evolution of adaptive immunity? *L. Likins, A. Smith, J. Wyckoff.* Division of Molecular Biology and Biochemistry, School of Biological Sciences, University of Missouri Kansas City (UMKC), Kansas City, MO.

Hemopexin (HPX), a glycoprotein in the blood plasma of many vertebrate organisms, has been shown to have the highest known binding affinity for heme. Many studies have shown that the primary function of the HPX protein product, Hemopexin is to sequester unbound heme released into the plasma from the breakdown of hemoglobin. The crystal structure of Hemopexin has been solved and the molecule consists of two structurally similar N-terminus and C-terminus beta-propeller domains joined by a flexible, unstructured linker peptide. This linker region is known to be part of the heme binding site. The gene that codes for Hemopexin (HPX) has been shown to have numerous sequences identified as recognizable hemopexin-like (HX) amino acid repeats in the primary structure of the protein. Structural mapping has revealed some interesting relationships between the functional motifs and the tertiary structure of the protein that may have implications for understanding the evolution of HPX proper, as well as its historical relationship to the several other classes of proteins that contain HX repeats, including: the myriad Matrix Metalloproteinases (MMP's), Proteoglycan-4 (PRG4) and Vitronectin (VTN). All these proteins are associated with the Extracellular matrix of multicellular animals, and research has implicated most of them in immune system dynamics. Earlier studies had deduced that the HPX protein was a complex of sequences that had evolved from ancestral molecular precursors, and as such, is the most derived of the proteins containing the hemopexin domain. However, more recent phylogenetic analyses of both primary amino acid sequences, gene nucleotide sequences, and tertiary structure of these proteins has determined that hemopexin is most likely the evolutionarily ancestral molecule. We conclude that the other proteins that have identifiable hpx domain structure have adapted the 4-blade beta-propeller which has then been modified for variable functionality. Importantly, our studies show for the first time that the beta-propeller has not arisen from amino acid identity over long periods of evolutionary time; instead, the propeller structure has remained unchanged even as the components of the propellers have been labile over long periods of evolutionary history. Our initial results show that these new insights reveal key details of the early evolution of the adaptive immune system in the Vertebrata and have clear implications for Human health.

1934M

POTE: an example of gene family evolution. *F. Anaclerio[1], G. Giannuzzi[1,2], M. Ventura[1].* 1) Department of biology, University of Bari, Bari, Italy; 2) Center for Integrative Genomics (CIG), University of Lausanne, Lausanne, Switzerland.

Eukaryotic genomes are characterized by the presence of gene families, set of genes derived from duplication events of a common ancestor. We described the genomic organization and evolutionary history of a primate specific gene family, POTE, whose members are expressed both in normal tissues and in many types of cancers (testis, prostate, ovary and breast). We studied the evolution of this gene family using both *in silico* and molecular cytogenetics approaches on six species: marmoset, macaque, orangutan, gorilla, chimpanzee and human. POTE gene family evolution has been characterized by both intra and inter-chromosomal duplication events leading to the increase of gene copies from one in marmoset to 14 (located on seven chromosomes) in human. In the analyzed species, we identified four different POTE gene isoforms: isoform 1, the unique encopassing exons 12-14, showed to be the ancestral form, isoform 2 appeared in the ancestor of Catarrhini and isoforms 3 and 4 appeared after the divergence of Old World Monkeys and Great apes. Isoform 3 and 4 are described as the most expressed in cancers and they are characterized by both the presence of a Long Inverted Repeat (LIR) within the last intron and by a processed actin retrogene in the last exon. The presence of a premature stop codon within the 15th exon in isoform 3 and the duplication of exons 8-11 in isoform 4 allowed us to distinguish these two isoforms. We proposed a model for the evolution of all the four isoforms from the first POTE gene on the ancestral chromosome 8 to the most recent human specific POTEF located on chromosome 2 and characterized by the presence of two additive exons harbored in the 5' UTR and responsable for a positive regulated expression. At 2q POTE genes map at the ancestral centromeric region that lost its function due to the fusion of the two ancestral acrocentic chromosomes that originated human chromosome 2. Analyzing the increase in the number of POTE genes and their localization on different chromosomes, we showed that a burst duplication of this gene family occurred since New World Monkeys split from the common ancestor of Catarrhini, moreover, exons loss and gain and the distribution of these events along the gene let us to suppose not only the acquisition of new function for this gene family in most recent species but also the potentiality to a further evolution.

1935S

Evolutionary Triangulation: Informing Genetic Studies with Evolutionary Evidence. *M. Huang[1], B. Graham[2], L. Muglia[3], S. Williams[1,2].* 1) Department of Genetics, Dartmouth College, Hanover, NH; 2) iQBS, Dartmouth College, Hanover, NH; 3) Perinatal Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Genetic studies of disease have been successful in identifying many variants associated with clinical outcomes. However, most have performed straightforward analyses without incorporating other factors that may affect risk. One such factor is evolution, as human populations have evolved under a variety of environments that have shaped disease risk. Using the unique evolutionary trajectories in diverse populations may improve our ability to extract important genetic factors in disease. Given that a large portion of common variants differ among populations in allele frequency, as do disease prevalences, we hypothesized that patterns of disease and population genetic structure may together inform association studies. Specifically, the distribution of allelic variants that associate with certain diseases will mirror the distribution of the diseases. Moreover, variants should be detectable most easily when three or more populations are assayed simultaneously, reducing the number of differences unrelated to diseases. We use three-way comparisons of evolutionary differentiation or Evolutionary Triangulation (ET). ET selects study populations based on differences in disease rates, with two having similar rates while a third is divergent. When the same genes affect disease risk in all populations, patterns of genetic variation should parallel disease distributions. In this study, we examine genetic differentiation (F_{ST}) of variants in three distinct continental populations (Europeans, CEU; South Asian, GIH; and West African, YRI) and use CEU as the outlier. SNPs that are highly differentiated between CEU and GIH/YRI, but not differentiated between GIH and YRI are identified as ET SNPs. ET SNPs are then mapped to ET genes with a ±100 Kb window. By surveying ET genes using the online HuGE Navigator, we identify genetic associations with all diseases. ET genes are enriched for those associated with diseases appropriately distributed among the three populations, with 12.12%, 7.39% and 5.46% under different cutoffs of F_{ST} used to generate the ET genes; fewer were detected using the same number of randomly selected genes. The enrichments are statistically significant with p values of 0.022, 0.002 and less than 10^{-4} accordingly depending on the exact F_{ST} thresholds used. The results indicate that ET can filter associated variants based on evolutionary comparisons among populations, and that it can be effectively filter association results.

1936M

MtDNA and health among Taiwanese. *J. Loo, T. Wang, J.A. Trejaut, M. Lin.* Mackay Memorial Hospital.

Previous analyzing the maternal heritage of Taiwanese through mitochondrial DNA (mtDNA) unanimously agree for an ancestral origin Taiwan Indigenous people coming from Mainland Southeast Asia (MSEA) Southeast China and Indo-China. Contemporary Non Taiwan Aborigines (often referred as Taiwan Han) have a more diversified range of origin with many finding their roots in Northern Asia and others in Southern Asia. In both groups, Taiwan Aborigines and Non-Aborigines some mtDNA lineages can trace their origin back to 15,000 YBP. Using association between mtDNA and systemic functional measurements (ATP, mitochondrial membrane potential, 8OHdG, lipid peroxidation, systolic BP, diastolic BP, blood glucose, total cholesterol, TG, LDL and HDL) we have shown that several measurements were significantly associated with mtDNA variation. Since many of these measurements are used in disease diagnosis, we asked whether mtDNA sequence variation could be used as a tool to foresee disease outcomes. We used DNA samples from Taitung area, where one third of the population is Taiwan Aborigines. We also included samples from the Taipei area where most people are Taiwan Han and have a very different mtDNA polymorphism. Efforts were directed toward determining significant associations between mtDNA SNP and disease determining functional measurements. Several mtDNA haplogroup/SNPs found potentially good indicators of disease development were further investigated in relation to ethnicity in view to pre-determine the likely health status of a group, with greater attention to disadvantaged groups in the Taiwanese population.

1937S

***RUNX3* GENE POLYMORPHISMS AND HAPLOTYPES IN MEXICAN PATIENTS WITH COLORECTAL CANCER.** *A.S. Suárez Villanueva[1], M.L. Ayala Madrigal[1], J. Peregrina Sandoval[2], M.W. Centeno Flores[3], V. Maciel Gutiérrez[4], M. Gutiérrez Angulo[5].* 1) Instituto de Genética Humana. CUCS. UdG; 2) Laboratorio de Inmunobiología. CUCBA. UdG; 3) Cirugía de Colon y Recto. Hospital Civil "Juan I Menchaca"; 4) Hospital Civil "Fray Antonio Alcalde"; 5) Departamento de Clínicas. CUAltos. UdG.

Background: The human RUNX3 gene is mapped to chromosome 1p36.1, contains six exons and its overall size is approximately 67 kb (1). The RUNX3 protein is required for neurogenesis of dorsal root ganglia and in intestinal epithelial participate with WNT and TGF-β pathways (2,3). The inactivation of RUNX3 gene has been associated with colorectal cancer. In this study, we analyzed the variants rs6672420 (exon 1, c.53 A>T), rs11249206 (intron 1, T>C), rs760805 (intron 3, T>A) and rs2236852 (intron 4, G>A) in colorectal cancer patients from Mexico. Material y Methods: Genomic DNA samples were obtained from peripheral blood of 176 Mexican patients with CRC at diagnosis and 195 individuals as control group. The polymorphisms were detected by PCR-RFLP. Association was estimated by calculating the odds ratio (OR). The haplotypes and linkage disequilibrium (LD) were established in Arlequin v3.5 software. Results: The RUNX3 rs2236852 AA genotype and A allele showed association with CRC (OR = 0.39 95% CI = 0.21 - 0.73 P <0.01; OR = 0.65 95% CI = 0.49 - 0.87 P <0.01, respectively), while rs6672420, rs11249206 and rs760805 polymorphisms displayed not significative results. The SNPs rs760805 and rs2236852 showed LD with r2 = 0.70 for controls and r2=0.44 for patients. The haplotype analysis revealed the most frequent was TG, constructed by combining the wild alleles. Moreover, the analysis showed CRC risk associated to TA haplotype OR = 2.52 95% CI = 1.47 - 4.30 P<0.01). Conclusion: The AA genotype and A allele of rs2236852 polymorphism have a decreased risk and TA haplotype have a risk to CRC development in Mexican patients. 1.Tsunematsu T, Kudo Y, Iizuka S, Ogawa I, Fujita T, Kurihara H, Abiko Y, Takata T: RUNX3 has an oncogenic role in head and neck cancer. PLoS One 4:e5892, 2009. 2.Kim WJ, Kim EJ, Jeong P, Quan C, Kim J, Li QL, Yang JO, Ito Y, Bae SC: RUNX3 inactivation by point mutations and aberrant DNA methylation in bladder tumors. Cancer Res.65:9347-54, 2005. 3.Zhang Z, Wang S, Wang M, Tong N, Fu G, Zhang Z: Genetic variants in RUNX3 and risk of bladder cancer: a haplotype-based analysis. Carcinogenesis.29:1973-8, 2008.

1938M

Deep sequencing of the human MHC region reveals widespread and ancient structural variation. *A.Q. Fu[1], B. Howie[2], E.Y. Kim[3], M. Stephens[1,4], S. Wolinsky[3].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Adaptive Biotechnologies Corporation, Seattle, WA; 3) Division of Infectious Diseases, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 4) Department of Statistics, University of Chicago, Chicago, IL.

The human major histocompatibility complex (MHC) region of 3.6 Mb harbors over 200 genes, most of which have important roles in innate and adaptive immunity. Several genetic variants in this region have been associated with risk for autoimmunity and disease progression in HIV infection and AIDS. The DNA sequence of the MHC is highly heterogeneous across individuals, and much of its population variation remains to be characterized. Here, we studied the genetic diversity of the extended MHC region in 64 HIV patients who carry no CCR5-delta32 allele but at least one copy of HLA-B*57:01 or B*57:03. We performed sequence capture and targeted enrichment of the MHC and its flanking regions for a total length of 4.7 Mb followed by ultra high-throughput DNA sequencing, generating long reads of ~400 bp at 55x coverage on average. Alignment of the sequencing reads to the human reference genome (hg19) led to the identification of ~60,000 SNPs, about 10 of which are in perfect linkage disequilibrium with HLA*B*57:01/03. De novo assembly of the reads revealed sequence segments that are more similar to the genomes of apes (chimp, gorilla, or orangutan) than to the human reference genome, indeed than to the 8 available human MHC reference haplotypes. These ape-like segments map to several loci in the MHC region of the ape genomes, and their presence correlates with the occurrence of several evolutionarily related HLA-A alleles. These ape-like segments of ~1000 bp were observed in about 20 of the 64 individuals in our study, and we hypothesize that they represent ancient structural polymorphisms. We further confirmed the presence of these long structural variants in several 1000G individuals by aligning sequencing reads in the 1000 Genomes project to the consensus ape-like sequence segments.

1939S

Estimating the generation time in human evolution. *P. Moorjani[1, 4], M. Wyman[1, 4], Z. Gao[2], M. Przeworski[1,3].* 1) Dept of Biological Sciences, Columbia University, New York, NY; 2) Committee on Genetics, Genomics and Systems Biology, University of Chicago; 3) Department of Systems Biology, Columbia University; 4) contributed equally.

Across mammals, decreased generation times (i.e., mean ages of reproduction) are known to correlate with increased neutral substitution rates. This observation is believed to reflect increased mutation rates due to higher numbers of cell divisions per unit time. Perhaps reflecting the same phenomenon, whole genome sequencing studies of human pedigrees, which reflect the mutation rate at present, yield a yearly estimate that is about two-fold lower than the one obtained from phylogenetic methods, which average over millions of years. One hypothesis is that increases in the generation time in the human lineage have led to a slow down of the mutation rate per unit time. Motivated by these considerations, we introduce an approach to estimate the generation time based on the insight that distinct types of mutations arise through different mechanisms and hence have different dependencies on the generation time. Notably, transitions at CpG sites are thought to occur primarily through spontaneous deamination, implying that their mutation rate should depend primarily on absolute time and should be relatively insensitive to the generation time. In contrast, replication-driven mutations at non-CpG sites occur in higher numbers following male puberty, and yearly mutation rates should depend on generation times. We characterize these different time dependencies from human pedigree data. By relating them to the observed average pairwise diversity values at CpG and non-CpG sites, we then obtain joint estimates of the mean coalescence time in years and the generation time over evolutionary history. To investigate if the age of reproduction has changed recently, we estimate the generation time from doubleton mutations, which reflect a more recent time depth than do average pairwise differences. Our results are consistent with the ethnographic literature and suggest a slight increase in the generation time towards the present. Beyond this application, our method provides a novel, population genetics based estimator of the generation time, applicable over different time depths of human evolution.

1940M

Role of Methylenetetrahydrofolate Reductase (MTHFR) in Risk of Opioid Abuse; Association between MTHFR and SOAPP®-R & ORT test. *T.G. Onojighofia, B. Meshkin, S.V Nguyen, D. Schwarz, B. Akindele, J. Hubbard, D. Holman.* Proove Biosciences, Inc.

Background: Folate is used by many biological processes to donate methyl groups to other compounds. Its active form is L-methylfolate which is produced through a reaction catalyzed by Methylenetetrahydrofolate Reductase (MTHFR). Mutations in MTHFR have been linked to everything from heart disease to cancer, but not, until recently, addiction. Objective: The objective of this study is to understand the association between MTHFR (MTHFR C677T, Rs1801133) and two popularly accepted tests in determining risk of opioid abuse/Misuse; Screener and Opioid Assessment for Patients with Pain (SOAPP®) and opioid risk tool (ORT). ORT (Low Risk 0 - 3, Moderate Risk 4 - 7, High Risk > 8) and SOAPP® (High risk SOAPP®-R score = 22 or greater, moderate risk SOAPP®-R score = 10 to 21, low risk SOAPP®-R score < 9*) Subjects: Subjects for this study were made up of two independent groups of patients. The SOAPP®-R test was taken by 2228 chronic pain patients in a clinic in Arizona. Another 3517 chronic pain patients across 55 clinical research sites in the US took the ORT test. Methods: The study evaluated data from both groups of subjects. The subjects were genotyped with TaqMan single nucleotide polymorphisms (SNP) assays using the proprietary Proove Narcotic Risk Genetics Profile Test (Proove Biosciences Inc, Irvine, California). The subjects were stratified into low, moderate, and high risk groups, based on the results from both questionnaires (i.e. ORT & SOAPP ®-R). Results: A cross tab analysis using IBM SPSS found an association between MTHFR (MTHFR C677T, Rs1801133) and risk of opioid abuse from both ORT and SOAPP® scores (p=0.028 and p= 0.042 respectively). A multinomial logistic regression showed that MTHFR homozygous mutation is more likely to be associated with the low risk group derived from SOAPP®-R compared to high risk group (P= 0.020, OR=4.273). In addition, it was found that Homozygous mutations in MTHFR is more associated with subjects with low ORT scores (P=0.028, OR 4.929) when compared to those with high risk stratified by ORT scores. Conclusion: This study suggests that an association exist between MTHFR and these two popular tests used to determine risk of opioid abuse/misuse. Findings in this study and further studies in this direction could improve understanding about the role of MTHFR in prescription opioid medication abuse or misuse.

1941S

Y Chromosome STR Mutation Rates: the Factor of 3 Connundrum. *D.E. Platt.* IBM T. J. Watson Research Center, Yorktown Hgts, NY., USA.

Y chromosome STR mutations serve as a prime clock for phylogeographical studies. However, there is a discrepancy between mutation rate estimates of roughly a factor of 3 between pedigree derived data vs. population based approaches. In 1995, Zhivotovsky et al constructed a Wright-Fisher based multi-step STR mutation model, extending Moran's ladder model of 1975. In 2004, Zhivotovsky, et al published an STR mutation rate using population data from a range of indigenous peoples that diverged from pedigree derived results, as noted by Di Giacomo et al in 2004. In 2005, Zhivotovsky, et al published a paper indicating that diversity fluctuations could produce a diversity bottleneck, yielding an apparently low mutation rate. Interestingly, their test population dynamics model was Poisson rather than Wright-Fisher. This left the question open of whether the Wright-Fisher model is inadequate for modeling diversity fluctuations in real populations. We present an argument showing that the Zhivotovsky Wright-Fisher dynamics model and the Poisson-based model may both be viewed as limiting cases of a model with binomially fluctuating population size in each generation. We show that the generation-by-generation Zhivotovsky formulation is preserved in both of these limits. Therefore, we conclude that Wright-Fisher dynamics is adequate for predicting the level of fluctuations expected for a Poisson model. Alternatively, Xu, Peng, and Fang proposed a model for STR mutations that allowed for a length dependent rate for decreasing single step mutations which could perhaps account for limited diversity, but assumed a Markov chain equilibrium state. We extend the Xu model first to consider the non-equilibrium Markov case, and second, incorporated this mutation model in a Wright-Fisher inheritance model. We show that the Markov model is robustly convergent, with equilibration sharing the same loss of information as predicted for the Jukes-Cantor model at equilibration, and dominating Wright-Fisher convergence. We identified regimes of back mutation length dependences that may satisfy bounds imposed by phylogeographical observations of human genetic diversity. We explored how BATWING (formulated on a single step model) parameter estimation responds to data generated from the extended Xu model. We see a reduction in predicted mutation rates accounting for at least some of the discrepancies, while still seeing correlations to adequate depth to capture modern human variation.

1942M

Extending the Ewens Sampling Formula to structured populations: Recursive computation of exact probabilities of allele frequency spectra. *M. Uyenoyama, S. Kumagai, B.D. Redelings.* Dept Biol, Duke Univ, Durham, NC.

To accommodate the enormous amount of information available in a sample of entire genomes, many evolutionary analyses have relied upon summary statistics or data reduction methods that marginalize over full genealogical trees. An approach that appears to have received relatively little attention involves the Ewens Sampling Formula (ESF), which provides a closed-form expression for the probability of the allele frequency spectrum (AFS) observed in a sample of arbitrary size under the infinite-alleles model of mutation. We extend the ESF to accommodate population structure, here accommodating migration between two demes. We develop a recursion which relates a sample to its immediate ancestor sample (vertical in time). We then characterize properties of the next-sampled gene, given the AFS already observed (horizontal in time). While the AFS probabilities are fully determined by the vertical recursion, their derivation is greatly simplified by the horizontal argument. A rescaling of time or mutation rate that would permit the use in structured populations of the ESF for unstructured population does not exist, even in the case in which the sample is derived entirely from a single deme in a multideme population. For that case, we propose an approximate scaling, which interpolates between no migration and the high migration limit. This approximation provides a qualitative description of the AFS probabilities: higher migration rates increase the effective mutation rate. Under other sampling conditions and for applications in which the exact likelihood is required, our method provides the AFS probabilities through a double recursion that accommodates successively greater numbers of observed alleles and successively greater sample sizes.

1943S

Whole genome sequencing of twenty Mauritian cynomolgus macaques (Macaca fascicularis). *M. Raveendran[1], R.A. Harris[1], D. Rio Deiros[1], A.J. Ericsen[2], G.L. Fawcett[1], F. Yu[1], C. Xue[1], R. Wiseman[2], D. Muzny[1], R.A. Gibbs[1], D. O'Connor[3], J. Rogers[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Wisconsin National Primate Center, Madison, WI; 3) University of Wisconsin, Madison, WI.

Cynomolgus macaques (Macaca fascicularis), also known as long-tailed or crab eating monkeys, are used in several aspects of biomedical research including studies of AIDS, diabetes and alcohol abuse. This species is widely used in pharmacology and drug development. The natural habitat of this species extends through Vietnam, Cambodia, the Philippines and Indonesia. An isolated population of cynomolgus macaques inhabits the island of Mauritius, and is descended from a small number of founder animals transported there by humans ~500 years ago. These Mauritian cynomolgus macaques (MCM) carry a very limited number of major histocompatibility complex (MHC) alleles and haplotypes. This makes these MCM more attractive for studies of infectious disease and other research questions where MHC variation can influence results. However, little is known about variation across the genome in the MCM population. We identified Single Nucleotide Variants (SNV) through whole genome sequencing of twenty MCM samples at a deep coverage using Illumina Hi-seq platform. The Illumina reads were mapped using BWA to the Indian rhesus macaque (IRM) reference genome (rhemac2) and variant calls made using Samtools. In total we identified 29,754,615 genome positions showing differences from the reference, and 22.59 million SNVs polymorphic among the 20 MCM. 7.16 million variants are fixed differences between these 20 MCM and the rhemac2, suggesting a 0.23% sequence divergence between MCM and IRM. The average number of polymorphic variants found in each animal is 9.96 million which is much higher than in human (4.2 million per person). We identified SNVs throughout the genome, including non-synonymous coding (51,132), synonymous coding (58,022), stop altered (1,619), splice altered (1,849), 3' UTR variant (82,307) and 5' UTR variant (21,429). Non-synomous variants are found in genes associated with human diseases, such as SBF2 (Charcot-Marie-Tooth), GAK (Parkinson's disease) CACNA1 (Neuroblastoma) and CAPN10 (diabetes). Analyses of the predicted impact of amino acid substitutions using Polyphen and DAVID are in progress. These results indicate that while MCM have reduced diversity in MHC haplotypes, they nonetheless exhibit significant levels of SNV with diversity higher than found in humans. Apparently the severe bottleneck in the founding of the Mauritius population had effects on the genetic composition of these animals, but did not reduce overall SNV levels dramatically.

1944M

Forensic Phenotyping in Brazilian population: SLC24A5 and ASIP as phenotypic predictors genes of skin, eye and hair color. *C. Fridman, F.A. Lima, F.T. Gonçalves.* Dept of Legal Medicine, Ethics and Occupational He, University of São Paulo, São Paulo, São Paulo, Brazil.

Pigmentation is a very variable and complex trait in humans and it is determined by the interaction of environmental factors, age, disease, drugs, hormones, exposure to ultraviolet radiation and genetic factors, including pigmentation genes. Many of these genes and their variants have been associated with phenotypic diversity of skin, eyes and hair color in homogeneous populations. SLC24A5, TYR, MC1R, SLC45A2, ASIP, OCA2 and HERC2 genes are noteworthy for their important contribution in pigmentation process. Prediction of phenotypes by using genetic information has benefited forensic area in many countries because it has made possible to infer physical characteristics from biological samples and, thus, lead criminal investigations. The aim of this study was evaluate polymorphisms in TYR, ASIP, SLC24A5 and SLC45A2 genes in a sample of 350 individuals of admixed population from Brazil, intending to use the data in forensic genetics casework in several situations. Volunteers answered a questionnaire where they self-reported their skin, eye and hair colors, sun sensitivity and lifestyle. No significant results were observed except for SLC24A5 and ASIP. The polymorphic homozygous allele of rs1426654 and rs6058017 in SLC24A5 (OR 32.88 $p<0.0001$) and ASIP (OR 8.68 $p< 0.007$) respectively, showed strongest association with fairer skin. Besides, the polymorphic homozygous allele in SLC24A5 exhibited relation to light eye color - green (OR 9.82 - $p<0.0001$), blond hair (OR 50.14 - $p<0.0001$) and also to increased sensitivity to sun exposure (OR 7.86 - $p<< 0.0002$). Our data suggests that polymorphic allele (A) in the SLC24A5 and ASIP genes is correlated with characteristics of light pigmentation, while the ancestral allele (G) is related to darker traits. Our findings corroborate previously published data on studies in European and African populations. These associations between pigmentation genes and skin, eyes and hair color shows that it is possible to use molecular information of an individual to access its phenotypic traits and use the obtained in attempt to help forensic investigations. Additional analyzes are ongoing as part of a project that evaluates 600 samples to check possible associations of phenotypic pigmentation in the Brazilian population with the mentioned genes. Financial Support: FAPESP (2012/02043-6), LIM 40/HCFMUSP and Department of Legal Medicine, Ethics and Occupational Health - FMUSP.

1945S

Interpretation of the high allele frequency of *GJB2* c.109 G>A variant in Chinese population: a pathogenic mutation or coincidental polymorphism? *Y. Lu[1,2], J. Cheng[2], H. Yuan[2].* 1) Otolaryngology Dept., No. 150 Central Hospital of PLA, Luoyang, Henan, China; 2) Otolaryngology Dept., Chinese PLA General Hospital, Beijing, China.

Hearing loss is a common sensory disorder in humans. Genetic factor plays more and more important role in the etiology of hearing loss than the environmental factors, which are controlled gradually and effectively with the development of the economy in the past two decades. Despite the high heterogeneity of genetic defects of hearing loss, *GJB2* mutations cause more than 17% nonsyndromic hearing loss in Chinese population. The pathogenic role of a common c.109 G>A variation in *GJB2* gene keeps uncertain in hearing loss cases. In this study, we recruited 37 subjects with homozygous c.109 G>A variation in *GJB2*, and 17 subjects with both heterozygous c.109 G>A and a heterozygous mutation in *GJB2*. All the subjects were excluded hotspot mutations in *SLC26A4* gene and mtDNA 12S rRNA, which were also common causative genes in Chinese deaf population. The hearing status of the subjects included normal(7 cases), mild to moderate (both prelingual and postlingual) (21 cases), severe to profound(26 cases). DNA samples from the parents were available. Fourteen tag-SNPs were selected based on the HapMap LD data set ($r^2>0.9$, MAF>0.1) in DFNB1 locus. Haplotypes of the subjects were derived from genotyping by SNaPshot technique. The results inferred that c.109 G>A in *GJB2* gene in Chinese population originated from a founder with a common LD block rather than a recurrent hotspot. Recent functional studies have produced somewhat contradictory results regarding the pathogenicity of the c.109 G>A variant. More than 6% allele frequency of c.109 G>A were detected both in hearing loss subjects and normal hearing controls in Chinese population. Although the penetrance of mild to moderate hearing loss in homozygotes is significant higher than heterozygotes and wild type subjects, the severity of hearing loss of *GJB2* c.109 G>A subject is still unpredictable in clinical practice. It is possible that other defects, such as CNV in the upstream sequence of *GJB2*, causes functional defect in connexin 26. The defects of other genes related to hearing loss need to be investigated with next generation sequencing. In conclusion, the pathogenic role of *GJB2* c.109 G>A needs further interpretation under next generation sequencing and functional study.

1946M

Target sequencing analysis of Parkinson's disease genes in a healthy Amerindian Population from Puno-Peru. *M. Cornejo-Olivas[1, 2], I.F. Mata[3,4], M.O. Dorschner[4], M. Inca-Martinez[1], A. Medina[1,7], A.C. Shetty[5], P. Kumari[5], K. Espinoza-huertas[1], D. Veliz-Otani[1], D. Incacutipa[7], V. Marca[1], O. Ortega[1], P. Mazzetti[1,8], C. Zabetian[3,4], T. O'Connor[5,6], Latin American Research Consortium on the Genetics of Parkinson's Disease (LARGE PD).* 1) Neurogenetics Research Center, Instituto Nacional de Ciencias Neurologicas, Lima, Peru; 2) Northern Pacific Global Health Research Training Consortium, Bethesda, MD; 3) Veteran Affairs Puget Sound Health Care System, USA; 4) University of Washington, USA; 5) Institute for Genome Sciences, University of Maryland School of Medicine; 6) Program in Personalized and Genomic Medicine, University of Maryland School of Medicine; 7) Universidad Nacional del Altiplano, Puno, Peru; 8) School of Medicine, Universidad Nacional Mayor de San Marcos. Lima, Peru.

Background: Both causal and susceptibility genes for Parkinson's disease (PD) known to date were discovered in either European or Asian populations. Studies on the genetics of PD in Latin-American populations are very much needed yet all such studies face a major obstacle in that there is no systematic information available on common genetic variability in healthy Amerindians to facilitate distinguishing pathogenic mutations from benign variants. **Objective:** To characterize genetic variability and estimate the frequency of genetic variants in 15 known causal PD genes in a Peruvian Quechua native population. **Methods:** Our study was performed using DNA samples and linked demographic data from 50 unrelated Amerindian subjects recruited in Puno, Peru through the Latin American Research Consortium on the Genetics of Parkinson's Disease. Sequencing analysis, including intron/exon boundaries was performed using a panel based on modified molecular inversion probes (MIPs) that included the 15 PD genes and 29 ancestry informative markers (AIMs). The resulting variants were categorized into different groups for comparison with other populations/datasets. We compared the frequencies of the single nucleotide variants (SNVs) to those in the 1000 Genomes Project (1KG) and Exome sequencing project databases (ESP). **Results:** Analysis of the 29 AIMs indicated that the average proportion of Amerindian ancestry in our sample of Peruvian Quechuas was 90%. Within the 15 PD genes, we detected 400SNVs. Of these variants 302 were not observed in 1KG, ESP or dbSNP 137. The majority of these unique variants were rare, but 18 had a minor allele frequency ≥ 5%. We found 124 novel non-synonymous variants and 9 novel nonsense mutations. **Conclusions:** Within a sample of healthy Amerindians from Peru, we found a surprisingly large number of previously unreported variants across the coding region of known PD genes . Annotating these variants in publicly available databases will aid future studies on the genetics of PD in Latin American populations. Furthermore, our results suggest that generating exomes or whole genomes in additional Amerindian controls (to supplement publicly available data) might be of benefit to genetic studies of other diseases conducted in populations with substantial Amerindian admixture.

1947S

Prevalence and sources of genetic variation in human mitochondria. *E. Glassberg[1], A. Harpak[1], D. Calderon[2], D. Cusanovich[3], M. Caliskan[3], Y. Gilad[3], C. Ober[3], J. Pritchard[1].* 1) Biology, Stanford University, Stanford, CA; 2) Biomedical Informatics, Stanford University, Stanford, CA; 3) Human Genetics, University of Chicago, Chicago, IL.

Heteroplasmy refers to the presence of multiple distinct organellar genomes within a single individual. In humans, heteroplasmic mutations in the mitochondrial genome are associated with a variety of bioenergetic defects and maternally inherited diseases. Here, we set out to quantify the prevalence of human mitochondrial heteroplasmy at the individual and population levels. We use the observed patterns of variation to infer the contributions of stochastic genetic drift and natural selection to mitochondrial evolutionary dynamics within and between human hosts. To identify heteroplasmic sites from next-generation sequencing data, we examine the log-likelihood ratio of simple statistical models of homoplasmy and heteroplasmy at each position of the mitochondrial genome. We apply this method to 2535 individuals from the 1000 Genomes Project and to RNASeq data from 432 individuals of varying degrees of relatedness from a founder population of Hutterites in South Dakota. In both cases, we observe widespread heteroplasmy. In the Hutterites, we find that heteroplasmic sites are significantly more likely to be shared between mother-child pairs than between unrelated individuals or father-child pairs. This pattern is more pronounced among sites at which the heteroplasmic allele is present at greater than five percent frequency in the mother. However, while some heteroplasmic sites are inherited maternally, the vast majority arise *de novo* within an individual. We then develop a novel inference algorithm to estimate the intensity of genetic drift experienced by the mitochondrial population during transmission from mother to child and during expansion from the germline to somatic cells. Separating the contributions of germline drift from those of somatic drift will allow us to clearly determine the population genetic setting in which inherited and *de novo* heteroplasmic variants are lost. Further, to understand the effects of selection on heteroplasmic sites, we compare observed patterns of heteroplasmy at sites that are likely to be neutral to those at sites that are likely to be influenced by selection, especially focusing on known disease-associated sites. These analyses form a basis for disentangling the contributions of various population genetic factors to observed patters of mitochondrial genetic variation.

1948M

Maternal Age Effect and Severe Germline Bottleneck in the Inheritance of Human Mitochondrial DNA. *M. Su[1], B. Rebolledo-Jaramillo[2], N. Stoler[2], J.A. McElhoe[3], B. Dickins[4], D. Blankenberg[2], T. Korneliussen[5], F. Chiaromonte[6], R. Nielsen[5], M.M. Holland[3], I.M. Paul[7], A. Nekrutenko[2], K.D. Makova[1].* 1) Department of Biology, Penn State University, USA; 2) Department of Biochemistry and Molecular Biology, Penn State University, USA; 3) Forensic Science Program, Penn State University, USA; 4) School of Science and Technology, Nottingham Trent University, UK; 5) The Department of Integrative Biology, the University of California at Berkeley, USA; 6) Department of Statistics, Penn State University, USA; 7) Department of Pediatrics, College of Medicine, Penn State University, USA.

The manifestation of mtDNA diseases depends on the frequency of heteroplasmy (the presence of several alleles in an individual), yet its transmission across generations cannot be readily predicted due to the lack of data on the size of mtDNA bottleneck during oogenesis. For deleterious heteroplasmies, a severe bottleneck may abruptly transform a benign (low) frequency in a mother into a disease-causing (high) frequency in her child. Here we present a high-resolution study of heteroplasmy transmission conducted on blood and buccal mtDNA of 39 healthy mother-child pairs of European ancestry (a total of 156 samples, each sequenced at ~20,000x/site). On average, each indivual carried one heteroplasmy, and one in eight individuals carried a disease-causing heteroplasmy, with minor allele frequency ≥1%. We observed frequent drastic heteroplasmy frequency shifts between generations and estimated the size of the bottleneck at only ~29-35 mtDNA molecules. Strikingly, we found a positive association between the number of heteroplasmies in a child and maternal age at fertilization, likely attributable to oocyte aging. Accounting for heteroplasmies, we estimate mtDNA germline mutation rate to be 1.3×10-8 mutations/site/year - lower than in previous pedigree studies but in agreement with phylogenetic studies, thus solving a long-standing controversy and informing the use of mtDNA in dating evolutionary events. This study takes advantage of droplet digital PCR (ddPCR) to validate heteroplasmies and confirm a de novo mutation. These results have profound implications for predicting the transmission of disease-causing mtDNA variants and illuminate mitochondrial genome evolutionary dynamics.

1949S

Cilioretinal artery: is it a variant angiogenesis under the effect of PAI-1 5G allele? *I. Akalin[1], A. Ardagil Akcakaya[2], S. Basaran Yilmaz[1], Y. Dag[3], M. Guzin Altinel[3], E. Kurum[4], E. Koyun[3], S. Ari Yaylali[2], H. Bayramlar[3].* 1) Medical Genetics, Istanbul Medeniyet University, Faculty of Medicine, Istanbul, Turkey; 2) Ophthalmology, Istanbul Medeniyet University, Goztepe Training and Research Hospital, Istanbul, Turkey; 3) Ophthalmology, Istanbul Medeniyet University, Faculty of Medicine, Istanbul, Turkey; 4) Biostatistics, Istanbul Medeniyet University, Faculty of Medicine, Istanbul, Turkey.

Purpose: Cilioretinal arteries (CRA) are small accessory arteries contribute to macular blood supply. The reason why some people have CRA while the others don't has not been elucidated yet, except its higher concordance in monozygotic twins. Here, we hypothesized whether genetic tendency to thrombosis due to well-known gene polymorphisms may induce CRA as a variant angiogenesis in embryonic life. Methods: We assessed plasminogen activator inhibitor-1 (PAI-1) 4G/5G, methylenetetrahydrofolate reductase (MTHFR), factor V leiden and prothrombin gene polymorphisms on 130 patients [82 females, 48 males; Median age: 57 (18-84) with CRA and 100 (64 females, 36 males; Median age: 55 (19-90)] without CRA. Genotyping was performed by melting curve analyses using a Rotor-Gene Q Real-Time PCR system (QIAGEN, Hilden, Germany) and PAI -1 4G/5G, MTHFR (C677T and A1298C), Factor V Leiden, Prothrombin G20210A mutation analyses kits (NLM , Settala, Italy). Multiple Logistic Regression model used to predict the dependent variable (occurrence of cilioretinal artery) using a set of independent variables (gene polymorphisms) in addition to test the association significance. Results: We found PAI-1 4G/5G; MTHFR (C677T and A1298C) have significant effects on the probability of occurrence of CRA at significance level 0.05. The odds ratio (OR) corresponding to PAI-1 4G/5G was 1,984 (95% CI: 1.320-3.000, p=0.001), which implies that keeping all other variables constant and having at least one 5G allele would increase the odds of having the cilioretinal artery by 98.4%. Additionally, we observed that while keeping all other variables constant, having at least one MTHFR C677T or A1298C allele would decrease the odds of having the cilioretinal artery by approximately 38% (OR=0.618, 95% CI: 0.394-0.961, p=0.035) or 44% (OR=0.558, 95% CI: 0.354-0.871, p=0.011), respectively. No significant differences were found between occurrence of CRA and Factor V Leiden or prothrombin. Conclusion: This is the first study to suggest the existence of strong association between the presence of CRA and genetic factors, up to the literature. Here we suggest that, not only the lack of genetic predisposition to thrombosis by MTHFR gene polymorphisms, but also the PAI-1 5G allele might promote variant angiogenesis by increasing pressure on endothelial cells as a result of decreasing the blood viscosity.

1950M

Tumor Necrosis Factor-alpha Gene Polymorphism in Turkish Patients with Psoriasis. *H. Akar[1], Y. Yildiray[2], E. Koç[2], E. Çalişkan[2], A. Akar[2], Y. Tunca[1].* 1) Medical Genetics, Gulhane Military Medical Faculty, Ankara, Turkey; 2) Dermatology, Gulhane Military Medical Faculty, Ankara, Turkey.

Introduction and objectives: Psoriasis is a chronic dermatosis that includes genetic and environmental factors in the etiology. Although complex cytokine network of psoriasis has not been well enlightened yet, tumor necrosis factor-α (TNF-α) is considered as main mediator in the pathogenesis of psoriasis. In addition to this, an association between single nucleotide polymorphisms (SNPs) at positions -238, -308, -857 in the promoter region of TNF-α and psoriasis were reported previous studies in different populations. In this study, we aimed to investigate potential association of TNF-α promoter gene polymorphisms with psoriasis compared with healthy controls in Turkish population. Materials and methods: We investigated 40 patients with psoriasis vulgaris and 40 healthy controls. SNPs at positions -208, -308, -857 in the promoter region of TNF-α were analysed by competitive allele specific-polymerase chain reaction method in both groups. SNPs frequencies were compared between the two groups and significance was evaluated by chi-squired test and two-tailed Fisher's exact test. Results: At the end of the study, we found that the TNF-α -308 AA and -857 CC homozygosities were significantly increased in patients with psoriasis in comparison with healthy controls (62.5% vs. 10%, p<0.01, OR=15.000; 47.5% vs. 7.5%, p<0.01, OR=11.159). TNF-α -308 AG and -857 TC heterozygosities were significantly decreased in patients with psoriasis in comparison with healty controls (32.5% vs. 75%, p<0.01, OR=0.160; 15% vs. 47.5%, p<0.01, OR=0.195). No significant difference was observed in SNP at -238 position. Conclusion: In conclusion, our findings suggested that TNF-α -308 AA and -857 CC homozygosities were associated with psoriasis vulgaris in Turkish population and these polymorphisms might be used as biological markers for psoriasis risk prediction. In contrast to previous studies, no significant differences were observed at -238 position. Further studies with larger populations are needed to confirm associations between -308 AA and -857 CC homozygosities and risk of psoriasis development in Turkish population.

1951S

Whole genome sequencing of a gibbon parent-offspring quartet to examine mutation rate variation in apes. *D.M. Bobo[1], O. Gokcumen[2], L. Carbone[3], M.F. Hammer[4], J.D. Wall[5], K.R. Veeramah[1].* 1) Ecology & Evolution, Stony Brook University, Stony Brook, NY; 2) Department of Biological Sciences, University at Buffalo, Buffalo, NY; 3) Department of Behavioral Neuroscience, Oregon Health and Science University, Portland, OR; 4) Arizona Research Laboratories Division of Biotechnology, University of Arizona, Tucson, AZ; 5) Institute for Human Genetics, University of California San Francisco, San Francisco, CA.

Understanding the evolution of the human germ line mutation rate (μ) is vital for numerous aspects of medical, population and evolutionary genetics. Traditionally, μ is estimated using a phylogenetic approach in which genetic divergence between two species is calculated. This divergence is then calibrated using some external estimate of when the species split and an assumed generation time to give the phylogenetic mutation rate per generation. More recently, direct observation of *de novo* mutations from whole genome sequencing of parent-child trios in humans have resulted in pedigree estimates of μ per generation that are approximately half of the phylogenetic rate. This discrepancy may be explained by the 'hominoid slowdown' where μ is thought to have decreased per unit time during the transition from small ancestral primates to larger apes with a longer generation times and lower mass-specific metabolic rates. As gibbons lie intermediate of old world monkeys and great apes with regard to phylogeny, body size and generation time, they are a potentially important extant primate family for testing the robustness of the hominoid slowdown hypothesis. Therefore we generated high coverage short read second-generation sequencing data for a family quartet from Nomascus gabriellae (the yellow-cheeked gibbon) in order to estimate the *de novo* pedigree mutation rate for this species. The male sibling was sequenced using Illumina's 1TB technology and yielded ~67x mean mapped read depth. The Illumina HiSeq 2500 was used for sequencing the parents and female sibling, with genome coverage ranging from 25-33x mean read depth. A custom bioinformatics pipeline was implemented incorporating published likelihood-based methods to weight candidate germline *de novo* variants and determine the probable parent of origin. Extensive validation of *de novo* candidates across a range of confidence scores was performed using multiple methods. We then used this data to develop a machine learning algorithm to estimate μ for the two individuals in this quartet for comparison to the recent pedigree estimates in humans. We were also able to identify *de novo* copy number variation using comparisons of normalized whole genome read depth. Finally, we used a HMM method to infer inheritance states across the genome and obtain fine-scale resolution of paternal and maternal recombination events, demonstrating that gibbons show a female recombination bias similar to that observed in humans and mice.

1952M

Significant association of Pro129Thr polymorphism in the fatty acid amide hydrolase (*FAAH*) gene with body mass index in Oceanic populations. *I. Naka[1], N. Nishida[2], T. Furusawa[3], R. Kimura[4], T. Yamauchi[5], K. Natsuhara[6], M. Nakazawa[7], Y. Ataka[8], T. Ishida[9], T. Inaoka[10], Y. Matsumura[11], R. Ohtsuka[12], J. Ohashi[1].* 1) University of Tsukuba, Tsukuba, Ibaraki, Japan; 2) Research Center for Hepatitis and Immunology, International Medical Center of Japan Konodai Hospital, Ichikawa, Japan; 3) Graduate School of Asian and African Area Studies, Kyoto University, Kyoto, Japan; 4) Transdisciplinary Research Organization for Subtropics and Island Studies, University of the Ryukyus, Nakagami, Okinawa, Japan; 5) Department of Health Sciences, Hokkaido University School of Medicine, Sapporo, Hokkaido, Japan; 6) The Japanese Red Cross Akita College of Nursing, Akita, Akita, Japan; 7) Department of International Health, Kobe University Graduate School of Health Sciences, Kobe, Hyogo, Japan; 8) School of Policy Studies, Kwansei Gakuin University, Sanda, Hyogo, Japan; 9) Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo, Japan; 10) Department of Human Ecology, Faculty of Agriculture, Saga University, Saga, Saga, Japan; 11) Faculty of Health and Nutrition, Bunkyo University, Chigasaki, Kanagawa, Japan; 12) Japan Wildlife Research Center, Taito, Tokyo, Japan.

Oleoylethanolamide (OEA), an appetite-suppressing mediator, is degraded by an integral membrane enzyme encoded by the fatty acid amide hydrolase (*FAAHItalic Text*) gene. A previous study demonstrated that the activity of FAAH enzyme was influenced by a non-synonymous SNP (rs324420; Pro129Thr) in the *FAAH* gene. The FAAH activity has been suggested to be higher in subjects homozygous for rs324420-C (129Pro) than in those for rs324420-A (129Thr) due to a post-translational mechanism. The possible association of rs324420 with overweight/obesity has been assessed in various populations to date. However, the association studies have yielded contradictory results. In the present study, the associations of rs324420 with anthropometric measurements (i.e., height, weight, and body mass index [BMI]) were examined in 694 adult subjects living in Solomon Islands and Tonga. A multiple regression analysis adjusted for age, sex, and population revealed that a copy of rs324420-C allele significantly increased weight by 2.3 kg (P-value = 0.0201), whereas rs324420-C was not associated with height. The rs324420-C allele was also significantly associated with increase in BMI (P-value = 0.00627 and slope = 0.91 kg/m2). Our results suggest that the efficient degradation of OEA due to 129Pro-FAAH protein may lead to excessive food intake and significant weight gain in Oceanic populations.

1953S

Brazilian population data on 26 non-CODIS STR loci used for paternity and kinship analysis. *V.S. Sotomaior[1], T.F.C. Pereira[1,2], N. Gaburo Júnior[2], J.C.M. Magalhaes[3], M. Malaghini[2].* 1) Graduate Program in Health Sciences, PUCPR, Curitiba, Brazil; 2) Diagnósticos da América S.A., Brazil; 3) Department of Genetics, Universidade Federal do Paraná, Curitiba, Brazil.

Microsatellites or short tandem repeats (STRs) are molecular markers that are widely used in population genetics and forensic analyses because of their high degree of polymorphism and heterozygosity. However, only a few population studies have been carried out in Brazil, particularly using markers that are not included in the US criminal justice DNA database systems - Combined DNA Index System (CODIS). In this study, we investigated 26 autosomal non-CODIS STRs (D1GATA113, D1S1627, D1S1677, D2S441, D2S1776, D3S3053, D3S4529, D4S2364, D4S2482, D5S2500, D6S474, D6S1017, D8S1115, D9S1122, D9S2157, D10S1248, D10S1435, D11S4463, D12ATA63, D14S1434, D17S974, D17S1301, D18S853, D20S482, D20S1082, and D22S1045) in a sample of the Brazilian population. This sample was composed of 1,068 unrelated individuals who had undergone genetic kinship analyses between 2007 and 2013. The individuals (n = 1068) were from 5 Brazilian regions (south: 717, southeast: 163, central west: 36, northeast: 129, and north: 23). For some analyses and whenever data was available, the individual's self-declared ethnicity was taken into consideration, according to IBGE's criteria (white: 641, black: 43, brown: 174, and yellow: 19). All the loci analyzed showed great diversity, the number of alleles varied from 5 (D4S2364) to 14 (D10S1435) and the mean of alleles number in all loci was 8.6. The majority of alleles exhibited a frequency that was higher than 1%, being allele 9 at locus D4S2364 the most frequent allele (q = 0.572). No significant deviations from the Hardy-Weinberg equilibrium were observed in any locus or sample; similar findings were obtained for F_{IS} values. The heterozygosity observed for each locus varied from 0.559 (D4S2364) to 0.851 (D9S2157). All 26 markers exhibited information capacity of use in kinship tests, with PIC > 0.5. The PDM was 81%, the PEE was 99.999999%, and the ITP was 99.99999%. In order to investigate population structure, the F_{ST} values for each pair of samples were calculated according to region and ethnicity. Although differences were observed for some loci, none of the obtained values reached significance, indicating low genetic differentiation among subsamples. Therefore, it has been demonstrated that the panel of markers investigated in this study are an important tool in the analysis of complex cases during kinship investigation, allowing the use of Brazilian population allele frequencies to infer the evidential value of kinship tests in Brazil.

1954M

An estimate of the average number of recessive lethal mutations carried by humans. *Z. Gao[1], D. Waggoner[2,3], M. Stephens[2,4], C. Ober[1,2,5], M. Przeworski[6,7].* 1) Committee on Genetics, Genomics and Systems Biology; 2) Dept of Human Genetics; 3) Dept of Pediatrics; 4) Dept of Statistics; 5) Dept of Obstetrics and Gynecology, University of Chicago, Chicago, IL; 6) Dept of Biological Sciences; 7) Dept of Systems Biology, Columbia University, New York, NY.

The effects of inbreeding on human health depend critically on the number and severity of the recessive deleterious mutations carried by an individual. In humans, estimates of the burden of recessive mutations per individual are based either on comparisons between consanguineous and non-consanguineous couples, an approach that confounds socioeconomic and genetic effects, or on carrier screening for disease-causing mutations, which suffers from other biases, notably the highly incomplete catalogue of disease-causing mutations. To circumvent these limitations, we sought to estimate a lower bound of the burden by focusing on recessive lethal disorders in a founder population with almost complete Mendelian disease ascertainment and a known pedigree. By considering all autosomal recessive lethal diseases recognized in the population and simulating allele transmissions along the pedigree, we estimated that each haploid human genome carries on average approximately one autosomal recessive allele that leads to severe disorders at or after birth in homozygous condition. When compared with previous estimates, our result suggests that recessive mutations that are lethal constitute a substantial fraction of the total burden of recessive deleterious mutations in humans.

1955S

Testing the infinite sites assumption using the 1000 genomes dataset. *S. Huang, D. Yuan.* State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, China.

The infinite sites model of the neutral theory is a fundamental assumption underlying nearly all population genetic and phylogenetic studies today but has yet to be properly tested. We here examined it from two novel perspectives using the 1000 genomes dataset. First, we studied the genetic diversity patterns of different human populations using a variety of different types of SNPs, such as noncoding, stop codon, nonsyn, syn, etc. Patterns shown by noncoding SNPs are expected to be similar to those shown by known functional stop codon SNPs, if most SNPs are not neutral. In contrast, neutral SNPs should show a most different pattern from stop codon SNPs. Second, it has long been well known that most genetic variations are shared among different human groups, which has been interpreted to mean few genetic differences among the ethnic groups (Lewontin, 1972). But the possibility of saturation or independent mutations to account for this phenomenon has yet to be examined and excluded. Saturation would mean limited rather than infinite sites and fast evolving DNAs are expected to reach saturation faster. We compared the number of shared SNPs in DNAs of different evolutionary rates among different human populations to see if shared SNPs are more common in fast evolving DNAs relative to slow ones. Results from these tests and their implications for phylogenetic studies will be presented.

1956M

Inference of mutation rates using hidden relatedness. *P.F. Palamara[1,2,3], P. Wilton[4], M. Fromer[5,6], G. Kirov[7], S. McCarroll[3,6,8], P. Sklar[5,9], M. Owen[7], S. Purcell[5,6,10], M. O'Donovan[7], J. Wakeley[4], I. Pe'er[11, 12].* 1) Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA; 2) Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; 3) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Boston, MA, USA; 4) Department of Organismic and Evolutionary Biology, Harvard University, Boston, MA, USA; 5) Division of Psychiatric Genomics in the Department of Psychiatry, and Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 6) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Boston, MA, USA; 7) Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK; 8) Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA; 9) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 10) Analytic and Translational Genetics Unit, Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 11) Center for Computational Biology & Bioinformatics, Columbia University Medical Center, New York, NY, USA; 12) Department of Computer Science, Columbia University, New York, NY.

Reliably estimating the mutation rate in modern humans has several implications for our understanding of demographic history (Scally and Durbin, Nature Reviews Genetics 2012). Recent estimates of the mutation rate obtained using de novo mutations in next-generation sequencing of families, however, were found to disagree with phylogenetic mutation rates derived from fossil evidence, motivating the development of new analytical methods. We describe an approach for the inference of mutation rates based on sharing of identical-by-descent (IBD) segments in sequencing data across purportedly unrelated individuals from a population. Using coalescent theory, we derive theoretical results for the distribution of mutation events found on IBD segments longer than a specified centimorgan threshold, for arbitrary demographic settings, under the SMC and SMC' models. Leveraging the relationship between the length and the age of shared IBD haplotypes, we devise a method to estimate both genotype error rates and mutation rates. The proposed approach based on hidden relatedness offers a substantial increase in statistical power compared to family-based analysis of de-novo mutations. This gain in power occurs despite the fact that the fraction of genome shared through long (e.g. >1cM) IBD segments across purportedly unrelated individuals is usually small, since IBD regions harbor events which have occurred in the recent past, over tens to hundreds of generations. Furthermore, analysis of de-novo mutations in trio-based studies is limited to genomic regions transmitted through known pedigree relationships, while when accurately phased data is available, mutation events can be analyzed on IBD segments across the quadratically larger set of all pairs of unrelated individuals. We validate the proposed methodology using synthetic datasets for a variety of demographic scenarios, and analyze mutation rates in 1246 trio-phased unrelated individuals from a recent exome sequencing study (Fromer et al., Nature 2014) of schizophrenia patients.

1957S

Hundreds of shared 'deletions' in ancient hominins are polymorphic in modern human populations. *D. Radke[1,2], C. Lee[3], S. Sunyaev[1,2].* 1) Harvard Medical School, Boston, MA; 2) Brigham and Women's Hospital, Boston, MA; 3) The Jackson Laboratory for Genomic Medicine, Farmington, CT.

Deciphering the genetic uniqueness of modern humans in relation to distant hominins and other primates is one of the central goals of human evolutionary genomics. Recently, with the availability of high-coverage sequence data for both Neanderthal and Denisova, it is now possible to more precisely determine the particular loci responsible for modern human uniqueness. While much of the distinguishing variation may be due to single nucleotide variants, genomic structural variants may also play a crucial role. Structural variants can be a potent phenotype-shaping force, particularly for unbalanced events, such as deletions, as they can alter reading frames and remove regulatory component space. Analyzing sequence read depth across archaic genomes, we find hundreds of 'deleted' regions in Neanderthal and Denisova (including many shared deletions), which are polymorphic in modern human populations. Some shared deletions overlap genes, and shared deletions as a set have a significantly higher allele frequency in modern human populations. Because these deletions are polymorphic in modern humans, they may represent regions of modern human-specific insertion, regions lost in archaic human lineages, or deletions polymorphic in both modern and archaic populations.

1958M

The evolution of genes underlying autism and schizophrenia and its impact on disease susceptibility. *M-.J. Favé[1], J. Hussin[2], P. Awadalla[1,3].* 1) CHU Sainte-Justine, Department of Pediatrics, Faculty of Medicine, Université de Montréal, Montreal, Quebec, Canada; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 3) CARTaGENE, 3333 Queen Mary Road, Office 493, Montreal, Quebec.

It has been hypothesized that a genotypic and phenotypic continuum exists between autism and schizophrenia, with both disorders showing diametric patterns in the impairment of language, communication and social behaviour. Recent work proposes that the human susceptibility to develop these disorders may originate as a collateral consequence of the selection for cognitive traits. Over the last decade, hundreds of both rare and common variants associated with the development of these disorders have been identified: several common variants have been found by genome-wide association studies, and rare *de novo* mutations have been captured by whole-exome sequencing studies of sporadic cases, for which no family history exists. We can therefore test whether genes harbouring common and *de novo* variants have evolved under different selective regimes, and whether local genomic features influence their variation and evolution. Using a recombination map from a French Canadian population recruited by the CARTaGENE project, we found that *de novo* variants associated with these disorders are enriched in regions of low recombination rate, whereas no such biases are found for common variants. Low recombining genes harbouring *de novo* variants are predominantly involved in fundamental cellular processes, whereas the few genes (49, or 2.7%) found in high recombination regions are enriched in membrane-proteins, axonal guidance and neuronal differentiation. In order to detect if genes harbouring common and *de novo* variants associated with autism and schizophrenia evolved under different selective pressures, we measured signals of selection using dN/dS ratios on the whole gene, and using sliding windows within genes. We performed these tests along the mammal and the primate lineage to estimate the timing of any selective event. Although most genes show patterns of evolution consistent with purifying selection, some genes in which many recurrent *de novo* mutations hits have been detected, such as *ABCA13*, *LRP1* and *MYH9*, show strong signals of positive selection in at least one domain of the gene, and along the primate lineage specifically. While genes involved in neurodevelopmental disorders mainly evolved under purifying selection in low recombining regions of the genome, these results show that diversity in specific domains of key genes may have been subjected to positive selection during primate evolution.

1959S

Systematic detection of positive selection in the human-pathogen interactome. *E. Corona[1], A.J. Butte[2], C.J. Patel[1].* 1) Biomedical Informatics, Harvard University, Boston, MA; 2) Biomedical Informatics, Stanford University, Stanford, CA.

Infectious disease has affected the human genome around the world. Identifying positive selection associated with infectious disease provides information regarding pathogen exposure in distinct human populations. Human-pathogen interactions are a constantly evolving interface between host and disease. We created a human pathogen interaction database and used the integrated haplotype score (iHS) to detect recent positive selection in genes that interact with proteins from 24 different infectious organisms. The Human Genome Diversity Panel was used to identify specific populations harboring pathogen-interacting genes that have undergone positive selection. We found that genes that interact with 12 pathogen species have undergone recent positive selection. The 12 pathogens are human immunodeficiency virus 1, Yersinia pestis, human respiratory syncytial virus, human herpesvirus 8, Bacillus anthracis, measles virus, dengue virus, murid herpesvirus 1, Francisella tularensis, Zaire ebolavirus, hepatitis C virus, and influenza A virus. Human genes in the Italian, Miao, and Biaka Pygmy populations that interact with Yersinia pestis show significant signs of selection. Our results show strong positive selection in genes that interact with human immunodeficiency virus 1, providing further support for the hypothesis that ancient humans were repeatedly exposed to lentivirus pandemics. These results indicate that infectious disease created distinct genetic footprints within affected populations.

1960M

Convergent mechanisms underlying hypoxia adaptation in Drosophila and Humans. *A.R. Jha[1,2,3], D. Zhou[4,5], C.D. Brown[1,2], G.G. Haddad[4,5], M. Kreitman[1,2,3], K.P. White[1,2,3].* 1) Institute for Genomics and Systems Biology, The University of Chicago, Chicago; 2) Department of Human Genetics, The University of Chicago, Chicago, USA; 3) Department of Ecology and Evolution, The University of Chicago, Chicago, USA; 4) Department of Pediatrics, Division of Respiratory Medicine, University of California at San Diego, San Diego, CA, USA; 5) Rady Children's Hospital, San Diego, CA, USA.

The ability to withstand low oxygen (hypoxia) is a highly polygenic yet mechanistically conserved trait that has important implications for both human health and evolution. However, little is known about the diversity of genetic mechanisms involved in hypoxia-adaptation in evolving populations. We used experimental evolution and whole-genome sequencing in Drosophila melanogaster to investigate the role of natural variation in adaptation to hypoxia. Using a Generalized Linear Mixed Model we identified significant allele frequency differences between three independently evolved hypoxia-tolerant populations and normoxic controls for ~4000 single nucleotide polymorphisms. Many of these variants are clustered in 66 distinct genomic regions representing long-distance linkage in our populations. These regions are enriched for genes associated with metabolic processes and contain genes that are differentially expressed between hypoxia-tolerant and normoxic populations. Additional genes associated with open tracheal system development and notch signaling pathways also showed evidence of directional selection. Knocking down the gene expression of a handful of candidate genes showed striking enhancement in survival in severe hypoxia, demonstrating their functional relevance in hypoxia adaptation. Using whole genome genotyping data from three high-altitude human populations, namely— Sherpas, Tibetans, and Ethiopians , we show that the human orthologs of the genes under selection in flies are also under positive selection in all three high-altitude human populations. Therefore, comparative genomics approaches, such as the one we have taken here, can be powerful in revealing genes and pathways underlying evolutionarily ancient traits that have conserved functions for millions of years.

1961S

Evaluating the impact of recent human demography on the frequency spectra using numerical solution of time-inhomogeneous diffusion equation. *E. Koch[1], J. Novembre[2].* 1) Department of Ecology and Evolution Unversity of Chicago, Chicago, IL; 2) Department of Human Genetics Unversity of Chicago, Chicago, IL.

Differences in recent demographic history appear to be an important driver of observed levels of genetic diversity among human populations. Recent attention has particularly centered on how populations that went through the out-of-Africa bottleneck have lower heterozygosity and polymorphic sites that are proportionally more likely to be nonsynonymous or predicted to be damaging. These results have suggested differences in the frequency spectrum of deleterious variation are also caused by varying population demographic histories. To investigate these phenomena in more detail, we perform numerical solutions to time-inhomogeneous diffusion equations for the allele frequency spectrum under the Poisson Random Field Model. This allows us to efficiently examine how the frequency spectra has evolved through time under a large number of possible human demographies and distributions of selective effects. We also are able to easily stratify variation observed today by the age at which the variation was generated. Using these tools, we demonstrate the ability of natural selection and demography to produce observed patterns and evaluate the relative impacts of population bottlenecks, recent growth rates, and changing efficacy of selection on the abundance of different variant types. The results emphasize how human frequency spectra are far from equilibrium and make more clear how frequencies are affected by major human demographic events at different timescales. For instance, in out-of-Africa populations the impacts of the bottleneck on the frequency spectra are still being realized, even as more recent growth events lead to an overlaid influx of rare variants. We quantify these effects and discuss their importance for interpretation of human genetic variation patterns. .

1962M

The Genetic Architecture of Skin Pigmentation in the Southern African ≠Khomani San. *A.R. Martin[1], J.M. Granka[2], C.R. Gignoux[1], M. Lin[3], C. Uren[4], M. Möller[4], C.J. Werely[4], J.M. Kidd[5], M.W. Feldman[2], E.G. Hoal[4], C.D. Bustamante[1], B.M. Henn[1,3].* 1) Genetics Department, Stanford University, Stanford, CA; 2) Department of Biological Sciences, Stanford University, Stanford, CA, 94305; 3) Department of Ecology and Evolution, SUNY Stony Brook, NY 11794; 4) Division of Molecular Biology and Human Genetics, Stellenbosch University, Tygerberg, South Africa; 5) Department of Human Genetics, University of Michigan, Ann Arbor MI.

Skin pigmentation is one of the most recognizably diverse phenotypes in humans across the globe, but its highly genetic basis has been primarily studied in northern European, Asian, and African American populations. The Eurasian pigmentation alleles are among the most differentiated variants in the genome, suggesting strong selection for light skin pigmentation. Light skin pigmentation is also observed in the far southern latitudes of Africa among KhoeSan hunter-gatherers of the Kalahari Desert. The KhoeSan hunter-gatherers are among the oldest human populations, believed to have diverged from other populations 100,000 years ago, and maintain extraordinary levels of genetic diversity. It is unknown whether light skin pigmentation represents convergent evolution or the ancestral human phenotype. We have collected ethnographic information, pigmentation phenotypes, and genotype data from 136 individuals in the ≠Khomani San from the Kalahari. To understand the genetic basis for light skin pigmentation, we have also exome sequenced 83 ≠Khomani San individuals to high coverage, generating one of the largest indigenous African exome datasets sequenced outside of the 1000 Genomes Project. In this study, ≠Khomani individuals have 11.5% admixture with Europeans and 10.9% admixture with Bantu speakers on average. European ancestry significantly lightens skin and explains 13.3% of the variance in pigmentation, and Bantu ancestry significantly darkens skin and explains 16.1% of the variance in pigmentation on average. We estimate that pigmentation is highly heritable ($h2 = 0.887 \pm 0.188$ standard error) and find that most of the heritability can be explained by 50 known pigmentation genes (0.527 ± 0.310 or 64.1% on average). After controlling for admixture with European and Bantu-speaking populations, a linear mixed model GWAS approach does not identify variants significantly associated with pigmentation. However, pigmentation genes are among the most globally differentiated between the ≠Khomani San and European or Bantu individuals, and aggregating differentiation with association data improves power to detect variants influencing selected traits. We identify highly differentiated variants between the ≠Khomani and both European and Bantu populations in multiple canonical pigmentation genes, including OCA2 and MITF. Our results highlight the strength of diverse population studies to explain phenotypic variation impacted by human evolutionary history.

1963S

Association study confirms that two OCA2 polymorphisms are involved in normal skin pigmentation variation in East Asian populations. *E. Parra, K. Eaton, P. Kavanagh, M. Edwards, S. Krithika.* Dept Anthropology, Univ Toronto, Toronto, ON, Canada.

The last decade has witnessed dramatic advances in our understanding of the genetic architecture of normal skin pigmentation variation in European populations. However, evidence is much more limited for East Asian populations. Recently, we carried out a study aimed at identifying putative signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. Based on the list of genes that show putative signatures of selection in East Asia, we prioritized a number of polymorphisms based on 1/ allele frequency information (e.g. differences in frequency between East Asian and non-East Asian populations) 2/ potential functional effects (e.g. Polyphen, SIFT and CADD scores) and 3/ conservation (e.g. GERP++ scores). The panel of SNPs selected includes 3 markers in the LYST gene (rs3754234, rs7522053 and rs4659610), one marker in the MLPH gene (rs2292881), 2 markers in the OPRM1 gene (rs1799971 and rs6917661), one marker in the EGFR gene (rs2227983), 4 markers in the BNC2 gene (rs9406647, rs3739714, rs10756778 and rs10962591), one marker in the TH gene (rs4930046), 3 markers in the OCA2 gene (rs1800414, rs74653330 and rs7497270), one marker in the TRPM1 gene (rs3809578) and 2 markers in the MC1R gene (rs33932559 and rs885479). We evaluated the association of these polymorphisms with skin pigmentation measured quantitatively using a DSM II colorimeter in a sample comprising 452 individuals of East Asian ancestry. Two previously described nonsynonymous polymorphisms within the OCA2 gene, rs1800414 (His615Arg) and rs74653330 (Ala481Thr) were strongly associated with melanin levels in this sample. Under an additive model, the common rs1800414 G allele, coding for Arginine, is associated with a decrease of 0.9 units in melanin levels. The rs74653330 A allele, coding for Threonine, is present at low frequency in East Asia (around 3% in our sample) and has a stronger effect on melanin levels than rs1800414 (decrease of 1.3 melanin units). No significant associations with skin pigmentation were observed for any of the other variants.

1964M

Positive Selection on Loci Associated with Drug and Alcohol Dependence. *B. Sadler[1], G. Haller[1,3], A. Goate[1,2,3].* 1) Department of Psychiatry, Washington University, St. Louis, MO, USA; 2) Department of Genetics, Washington University, St. Louis, MO, USA; 3) Department of Neurology, Washington University, St. Louis, MO, USA.

Much of the evolution of human behavior remains a mystery, including how certain disadvantageous behaviors are so prevalent. Nicotine addiction is one such phenotype. Several loci have been implicated in nicotine related phenotypes including the nicotinic receptor gene clusters (*CHRNs*) on chromosomes 8 and 15, and the nicotine metabolizing gene CYP2A6 on chromosome 19. Here we use 1000 Genomes sequence data from 3 populations (Africans, Asians and Europeans) to examine whether natural selection has occurred at these loci. We used Tajima's D and the integrated haplotype score (iHS) to test for evidence of natural selection. Our results provide evidence for strong selection in the nicotinic receptor gene cluster on chromosome 8, previously found to be significantly associated with both nicotine and cocaine dependence as well as evidence of weaker, but still detectable, selection acting on the region containing the *CHRNA5* nicotinic receptor gene on chromosome 15, that is genome wide significant for risk for nicotine dependence. To examine the possibility that this selection is related to memory and learning, we utilized genetic data from the Collaborative Studies on the Genetics of Alcoholism (COGA) to test variants within these regions with three tests of memory and learning, the Wechsler Adult Intelligence Scale (WAIS) Block Design, WAIS Digit Symbol and WAIS Information tests. Of the 17 SNPs genotyped in COGA in this region, we find one significantly associated with WAIS digit symbol test results. This test captures aspects of reaction time and memory, suggesting that a phenotype relating to memory and learning may have been the driving force behind selection at these loci.

1965S

Neanderthal Origin of the Haplotypes Carrying the Functional Variant Val92Met in the *MC1R* in Modern Humans. *Q. Ding[1], Y. Hu[1], S. Xu[2], C. Wang[1], H. Li[1], R. Zhang[1], S. Yan[1], J. Wang[1], L. Jin[1,2].* 1) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China; 2) CAS-MPG Partner Institute for Computational Biology, Shanghai Institute for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), Shanghai, China.

Skin color is one of the most visible and important phenotypes of modern humans. Melanocyte-stimulating hormone and its receptor played an important role in regulating skin color. Here we present evidence of Neanderthal introgression encompassing the melanocyte-stimulating hormone receptor gene *MC1R*. The haplotypes from Neanderthal introgression diverged with the Altai Neanderthal 103.3 KYA, which postdates the anatomically modern human - Neanderthal divergence. We further discovered that all of the putative Neanderthal introgressive haplotypes carry the Val92Met variant, a loss-of-function variant in *MC1R* that is associated with multiple dermatological traits including skin color and photoaging. Frequency of this Neanderthal introgression is low in Europeans (~5%), moderate in continental East Asians (~30%), and high in Taiwanese aborigines (60-70%). Since the putative Neanderthal introgressive haplotypes carry a loss-of-function variant that could alter the function of *MC1R* and is associated with multiple traits related to skin color, we speculate that this Neanderthal introgression, together with the previously reported Neanderthal introgression at *HYAL2*, may have played an important role in the local adaptation of modern Eurasians to sunlight intensity.

1966M

Altitude adaptation in Tibet caused by introgression of Denisovan-like DNA. *E. Huerta-Sanchez[1,2,3], X. Jin[2,4], A. Asan[2,5,6], Z. Bianba[7], B. Peter[1], N. Vinckenbosch[1], Y. Liang[2,5,6], M. He[2,8], M. Somel[9], P. Ni[2], B. Wang[2], X. Ou[2], H. Huasang[2], J. Luosang[2], Z.X. Ping Cuo[10], K. Li[10], G. Gao[11], Y. Yin[2], W. Wang[2], X. Zhang[2,12,13], X. Xu[2], H. Yang[2,14,15], Y. Li[2], J. Wang[2,15], J. Wang[2,14,16,17,18], R. Nielsen[1,2,19,20].* 1) University of California at Berkeley, Berkeley, CA; 2) BGI-Shenzhen, Shenzen, China; 3) University of California at Merced, Merced, CA; 4) School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, China; 5) Binhai genomics institute, BGI-Tianjin, Tianjin, 300308, China; 6) Tianjin Translational Genomics Center, BGI-Tianjin, Tianjin, 300308, China; 7) The People's Hospital of Lhasa, Lhasa, 850000, China; 8) Bioinformatics and Computational Biology Program, Iowa State University; 9) Department of Biological Sciences, Middle East Technical University, Ankara, Turkey; 10) The No.2 people's hospital of Tibet Autonomous Region,850000, China; 11) The hospital of XiShuangBanNa Dai Nationalities, Autonomous Jinghong 666100, Yunnan, China; 12) The Guangdong Enterprise Key Laboratory of Human Disease Genomics, BGI-Shenzhen,Shenzhen,China; 13) Shenzhen Key Laboratory of Transomics Biotechnologies, BGI-Shenzhen,Shenzhen,China; 14) Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21589, Saudi Arabia; 15) James D. Watson Institute of Genome Science, Hangzhou, China; 16) Department of Biology, University of Copenhagen, Ole MaaløesVej 5, 2200 Copenhagen, Denmark; 17) Macau University of Science and Technology, AvenidaWai long, Taipa, Macau 999078, China; 18) Department of Medicine, University of Hong Kong, Hong Kong; 19) Department of Statistics, University of California, Berkeley, CA; 20) Department of Biology, University of Copenhagen, Copenhagen, Denmark.

As modern humans migrated out of Africa, they encountered many different environmental conditions including temperature extremes, new pathogens, and high altitude. These diverse environments have likely acted as agents of natural selection and led to local adaptations. One of the most illustrious examples in humans is the adaptation of Tibetans to the hypoxic environment of the high-altitude Tibetan plateau. A hypoxia pathway gene, EPAS1, was previously identified as having the most extreme signature of positive selection in Tibetans, and was shown to be associated with differences in hemoglobin concentration at high altitude. Re-sequencing the region around EPAS1 in 40 Tibetan and 40 Han individuals, we find that this gene has a highly unusual haplotype structure that can only be convincingly explained by introgression of DNA from Denisovans or Denisovan-related individuals into humans. Scanning a larger set of worldwide populations, we find that the selected haplotype is only found in Denisovans and in Tibetans, and at very low frequency among Han Chinese. Furthermore, the length of the haplotype, and the fact that it is not found in any other populations, makes it unlikely that the Tibetan/Denisovan haplotype sharing was caused by incomplete ancestral lineage sorting rather than introgression. Our findings illustrate that admixture with other hominin species has provided genetic variation that helped humans adapt to new environments.

1967S

Whole genome sequencing to uncover adaptation to high altitude in the Andes. *M. Muzzio[1,2], K. Slivinski[3], M.C Yee[4], T. Cooke[5], C.D Bustamante[5], G. Bailliet[1], C.M. Bravi[1,2], E.E. Kenny[3,4,6,7,8].* 1) Consejo Nacional de Investigaciones Cientificas y Tecnologicas, La Plata, Buenos Aires, Argentina; 2) Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, Argentina; 3) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, NY; 4) Dinneny Lab. Carnegie Institution of Washington. Department of Plant Biology, CA; 5) Stanford University School of Medicine, CA; 6) Department of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, NY; 7) The Center for Statistical Genetics, Icahn School of Medicine at Mount Sinai, NY; 8) The Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, NY.

There is interest in human adaptation to a diversity of environments, including finding the genetic basis to phenotypes favorable to pressures such as hypoxia. We have preliminary Illumina Exome Array data on a set of 43 individuals from high altitude villages in the Andes from the Humahuaca area, Argentina (~2500 meters above sea level) and 11 individuals from a neighboring lowland population, Tartagal, Argentina (less than 500 meters above sea level), all with over 90% Native American ancestry estimated using the Admixture software. Currently, we are sequencing full genomes of 10 individuals from each of these populations, in search for new population-specific variants. We will use the population branch statistics (PBS) to identify highly differentiated genomic regions between the highlanders (Andean) and lowlanders (Chaqueños). We will discuss the results of our scan in light of related work on the adaptation of Tibetans, Ethiopians, and other Andean populations to hypoxia.

1968M

***IFNL3*/*IFNL4* region shows evidence for recent positive selection specific to Asian populations.** *G.L. Wojcik, C.D. Bustamante.* Department of Genetics, Stanford School of Medicine, Stanford, CA.

Hepatitis C virus (HCV) is a global health burden, chronically infecting 130-150 million people and causing 350,000-500,000 deaths per year from HCV-related liver disease. Twenty-five years after the discovery of HCV, there is no vaccine and treatment remains ineffective in a large proportion of individuals. Heterogeneity in clinical outcomes such as spontaneous clearance of the virus, as well as sustained virologic response (SVR) after treatment, has been observed between individuals of different genetic ancestry. Previous genetic studies have pinpointed a single nucleotide polymorphism (SNP) in the interferon-γ 3 and 4 (*IFNL3*/*IFNL4*) region (rs12979860) as being strongly associated with clinical outcome. While the derived and favorable allele of rs12979860 (C) is present globally, its frequency is greatly differentiated by continent with the lowest in African populations (34-49%), and the highest in Asian populations (89-96%). To determine if these differences are due to selective pressures, data from the phase 3 release of the 1000 Genomes Project (TGP) was analyzed for population-specific signatures of selection. Derived allele frequency (DAF), F_{st}, nucleotide diversity (π), and haplotype structure were examined and compared in populations from Europe, Africa, Asia, and the Americas. A 5 kilobase (kb) region around *IFNL3*/*IFNL4* showed decreased nucleotide diversity, high DAF, and increased haplotype homozygosity in Asian populations. This pattern is not found in Native American populations, suggesting recent positive selection specific to Asia. Historical selective pressures from HCV, or likely a related ancestral virus, may have driven the favorable rs12979860 allele to near fixation. However, Asia currently has disproportionately high HCV-related morbidity and mortality despite this adaptation, suggesting further evolution of the virus. Differences in clinical outcomes within Asian populations may therefore be also due to non-*IFNL3*/*IFNL4* genetic variation. Further studies are needed to identify additional genetic associations that will better our knowledge of how HCV interacts with the human immune system.

1969S

A genome-wide natural selection scan using 1000 high-coverage, Alzheimer's-specific whole-genome sequences. *M. Ebbert[1], H. Smith[1], T. Dawson[1], S. Grossman[2], M. Norton[3], J. Tschanz[3], R. Munger[4], C. Corcoran[5], P. Ridge[1], J. Kauwe[1], ADNI.* 1) Department of Biology, Brigham Young University, Provo, UT; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Department of Family Consumer and Human Development, Utah State University, Logan, Utah; 4) Department of Nutrition, Dietetics, and Food Sciences, Utah State University, Logan, Utah; 5) Department of Mathematics and Statistics, Utah State University, Logan, Utah.

Natural selection studies have impacted genetic research and our understanding of human adaptations, including malaria resistance, skin pigmentation, and others. More recently, Grossman et al. discovered adaptations to bacterial response and specific human phenotypes by performing a genome-wide selection scan using the 1000 Genomes data—identifying specific adaptive mutations without foreknown, adaptive phenotypic traits. This scanning approach successfully reversed the study type from a hypothesis-driven to a hypothesis-generating study. While the genome-wide scan was successful, there are potential limitations: (1) the 1000 Genomes data has only 179 whole-genome sequences; (2) the sequences were low coverage (2-6x average coverage); and (3) genotypes for the 1000 Genomes data may be inaccurate due to low coverage and because they were not genotyped using modern 'joint-calling' algorithms. We are performing an updated analysis including 1000+ Alzheimer's-specific, whole-genome sequences with 37x average coverage. Our data set includes 152 Alzheimer's disease (AD) cases and 211 'super controls'. The 'super controls' are APOE ε4 positive individuals aged 75+ that do not exhibit AD symptoms. Using our large, high-coverage data set, we will explore whether larger sample size and deeper coverage reveals previously undiscovered loci under selection. We will also explore whether using an AD-specific data set will enhance selection signals related to AD under the premise that AD-related loci are known to be under selection. As such, AD may be the result of a conflicting pleiotropic effect of an otherwise beneficial genotype. After joint calling all samples using GATK's HaplotypeCaller, we will perform a genome-wide natural selection scan using the Composite of Multiple Signals (CMS) algorithm on our data set to identify specific loci under selection. These results will be compared to Grossman et al.'s previous results to determine whether any new loci show evidence of selection and whether any previously identified regions were eliminated (potential false positives). Previous and newly identified loci will be examined for potential AD implications based on known disease associations and functional annotations. Top candidates will be tested using an association test. Natural selection studies reveal important genetic artifacts for observed phenotypes. Many AD-related genes are under selection and there are likely other undiscovered AD-related genes.

1970M

Identification of functional signals of recent selection in the Sea Island Gullah African Americans. *C.D. Langefeld[1], S. Sajuthi[1], J. Divers[1], Y. Huang[2], U. Nayak[2], W.M. Chen[2], K.J. Hunt[3], D.L. Kamen[3], G.S. Gilkeson[3], J.K. Fernandes[3], IJ. Spruill[3], W.T. Garvey[4], M.M. Sale[2], P.S. Ramos[3].* 1) Biostatistical Sciences and Center for Public Health Genomics, Wake Forest Medicine, Winston-Salem, NC; 2) Department of Medicine and Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 3) Department of Medicine, Medical University of South Carolina, Charleston, SC; 4) Department of Nutrition Sciences and Birmingham VA Medical Center, University of Alabama, Birmingham, AL.

Many common diseases are more prevalent in specific ethnic groups. Given the growing number of disease-associated loci in regions that show evidence of selection, identification of alleles under selection may provide insight into disease susceptibility. Relative to other African-Americans (AA), the Gullah population has lower European admixture and higher ancestral homogeneity from the Sierra Leone area in Far-West Africa. We sought to capitalize upon the relative closeness between the Gullah and Sierra Leoneans to identify regions that differentiate both populations and may be under recent population-specific selective pressures. We integrated these signals with functional annotation to identify biologically relevant variation. Using 277 Gullah and 400 Sierra Leonean samples, we computed a linear regression model of the HapMap YRI principal component (PC2) as a quantitative outcome. We adjusted for European admixture via the CEU component (PC1) as a covariate. We computed the cross population extended haplotype homozygosity test (XP-EHH), as a formal test of selection. In total, 679,513 SNPs with MAF>5% were used in this analysis. SNPs with strong evidence of association with PC2 and strong evidence for selection from the XP-EHH test were prioritized, as measured by the smallest Euclidean distance (L2-norm) between both tests. The top 1,000 variants with combined evidence for differences in PC2 and extended haplotype homozygosity were annotated and prioritized based on the potential impact of amino acid changes and regulatory functions using RegulomeDB and HaploReg. Some of the regions that best differentiate the Gullah from the Sierra Leonean and show evidence for selection harbor missense SNPs (HLA, RAI14, SEC31B and PRR14L genes). Other regions harbor multiple SNPs with high regulatory scores based on the simultaneous presence of eQTLs, transcription factor binding and DNase sites, including those of the MTRR, SEPP1, HLA, RAPGEF5, AUTS2, TNKS, HBG1, and L2HGDH genes. Enhancer enrichment analysis of the top 1,000 SNPs revealed a 10.8 fold significant enrichment of strongest enhancers in H1 human embryonic stem cells. In summary, we identified several functional variants that differentiate the Gullah from the Sierra Leoneans, suggesting that recent selection may be operating at these loci. Identification of functional regions that might be under selection in the Gullah has the potential to elucidate disease risks in AA.

1971S

Dissecting Genetic Architectures of Human Zinc Transporter Genes and Searching Footprints of Natural Selection in Global Populations. *J. Li, C. Zhang, S. Xu.* CAS-MPG Partner Institute for Computational Biology, Shanghai, China.

Zinc transporters play very important roles in all eukaryotes by maintaining the rational zinc concentration in the cells. The genetic diversities of those transmembrane proteins are expected to affect the biologic functions of both zinc transporters themselves and the other zinc-related enzymes. Although there are many works about the huge diversity of Zn content in soils or crops around globe-scales, the diversity of Zinc transporter genes (ZTGs) in worldwide human populations have not been well studied. Here, we investigated the global genetic diversity of 24 human ZTGs, including 10 SLC30A and 14 SLC39A family genes basing on full sequencing data of 1000 genomes project. Intriguingly, we found some of ZGTs are very evolutionarily conserved in all human populations, such as SLC30A6 with significant low GA-FST (0.015), while some other ZGTs exhibited the extremely high differentiations among populations, such as SLC30A9(GA-FST, 0.284), SLC30A3 (GA-FST, 0.154), etc. Moreover, ZTGs harbored higher differentiated SNPs than random genes, suggesting that large genetic differentiations among population exhibited in ZTGs. More surprisingly, SLC30A9 was conformed to be selected both in East Asians and Africans but the selective pressures were different in the continental groups. In SLC30A9, the frequencies of two different alleles of a non-synonymous SNP, rs1047626, are almost fixed in Africans and East Asians, respectively, i.e. 96.4% A in CHB, 92% G in YRI. Therefore, there are two different functional haplotypes exhibited dominated abundance in Africans and East Asians, respectively. Furthermore, according to previous studies of global distributions Zn contents in soils or crops, strong correlation was observed with the haplotype frequencies of SLC30A9 among different human populations. We speculate that population was forced to utilize advantageous functional haplotype to adapt to the local zinc state or diets and the genetic differentiation of ZTGs could contribute to population heterogeneity in zinc transporting capability in different living environment so that the balance of zinc concentration in serum or cell can be kept properly. Our investigations should facilitate to further functional studies of ZTGs and medical studies on worldwide and regional nutrient problem and zinc-related diseases.

1972M

Genomic Patterns of Natural Selection on Toll-Like Receptors and N-Glycosylation Genes in Humans and Toll Genes in Drosophila: A Comparative Study on the Evolution of Innate Immunity Genes from Invertebrates to Vertebrates. *S. Mukherjee[1], D. Ganguli[2], P.P. Majumder[2].* 1) BioMedical Genomics Centre, Kolkata, West Bengal, India; 2) National Institute of Biomedical Genomics, Kalyani, West Bengal, India.

Background: The innate immunity genes are the first line of host defence, being developed before the separation of invertebrates and vertebrates. There is considerable interest to decipher the effect of natural selection on these genes across diverse human populations, and also across invertebrates and vertebrates, to account for the change in pathogen load and diversity. Among them the most important are the cell surface Toll-Like Receptors (TLRs) that engage in direct host-pathogen interactions in humans and other mammals. They are the homologs of Toll genes identified in Drosophila. Furthermore, genes involved in the N-Glycosylation pathway that shapes the cell-surface glycome moiety regulates the innate immune response through host-pathogen interactions and are hence potential candidates for natural selection. Objectives: To study the differential natural selection patterns on the Toll-Like Receptors and the N-Glycosylation genes across human populations and characterizing the TLR homologous Toll genes in multiple Drosophila genomes. Methods: We have generated DNA sequence data on the important cell surface TLR genes in pre and post agricultural populations of India (n=266) and analysed them in conjunction to the data available in the 1000 Genomes Project (n=1092). The TLR homologs in Drosophila Toll genes are identified and genome-level data are downloaded from the Drosophila Reference Gene Panel for further analysis. For studying the N-Glycosylation genes across Indian populations, we have analysed genome-wide data from diverse ethnicities and performed data analyses for SNPs in these genes. Results: For the TLR genes in humans, we observed an excess of rare variants and a large number of low frequency haplotypes, non-synonymous changes are few (dN/dS<1) and tests of neutrality (Tajima's D, FuLi's D*,F*,Fu's Fs) are found to be negative. No evidence of extended haplotype homozygosity is observed. In the Drosophila Toll genes, SNVs with MAF<0.05 are predominant. The results are quite similar to that observed in humans. The N-Glycosylation genes in humans show significantly lower heterozygosity values compared to neutrality. Population differentiation indices (FST) are lower for all the genes across multiple populations. Inference: The results obtained suggest possible role of purifying selection operating on key innate immunity genes across both invertebrates and vertebrates, thus focusing on their immense importance in the survival of the host.

1973S

Evolutionary history of pigmentation candidate gene diversity in a Melanesian population. *H. Norton, E. Werren.* Department of Anthropology, University of Cincinnati, Cincinnati, OH.

Pigmentation of the skin, hair, and eyes are complex phenotypic traits determined by multiple loci. Human skin pigmentation is a trait that is believed to have evolved under strong natural selection in response to varying levels of ultra-violet radiation (UVR) intensity. Lighter skin color has evolved multiple times in human evolutionary history, but it is unclear if the darker skin color observed in many high UVR populations is also the result of evolutionary convergence (suggesting that population-specific mutations may have been favored by positive selection) or if instead ancestral variants associated with darker skin color have been maintained in high-UVR populations via purifying selection. To begin to address this question we compare DNA sequence variation from multiple pigmentation candidate genes in a Melanesian population to variation observed in European, East Asian, and African populations sequenced in the 1000 Genomes Project. Summaries of the site frequency spectrum, including Tajima's D (TD), for three genes, *ASIP*, *OCA2*, and *TYRP1*, do not indicate that any of these genes were targeted by positive selection in the Melanesian population (*ASIP* TD = 0.037, *OCA2* TD = -0.85, *TYRP1* TD = -0.55). With the exception of a single novel haplotype in the *OCA2* locus observed at a frequency of ~10% there is little evidence that Melanesians exhibit any high frequency population-specific haplotypes at these loci, suggesting that if an independent adaptation to high UVR conditions occurred in Melanesians then other pigmentation loci are responsible. However, there is also little evidence that Melanesians are similar to Africans at these loci, which one might expect if Melanesians share ancestral haplotypes with other high UVR populations: pairwise F_{ST} estimates between Melanesians and Africans for the pigmentation loci examined here range from 0.043-0.443, and the majority of Melanesian haplotypes are common haplotypes shared between Africans, Europeans, and East Asians. We explore these patterns of sequence variation and inter-population divergence at pigmentation loci in the context of evolutionary models for pigmentation change in the human species and with consideration to Melanesian population history.

1974M

Inference of the strength of purifying selection based on haplotype patterns. *D. Ortega Del Vecchyo[1], K.E. Lohmueller[1,2], J. Novembre[3].* 1) Interdepartmental Program in Bioinformatics, University of California, Los Angeles, CA; 2) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA; 3) Department of Human Genetics, University of Chicago, IL.

The strength of purifying selection is a central factor underlying levels of genetic diversity in a population and is important to characterize to understand the expected genetic architecture of disease traits. Recent sequencing studies with large sample sizes have revealed a much higher proportion of non-synonymous variants among rare versus common variants in human populations. This finding suggests that natural selection is acting against such variants to keep them at low frequencies in the population. To estimate the strength of purifying selection, we have developed a method that uses the lengths of pairwise haplotype identity among rare-variant-carrying haplotypes. Unlike previous approaches, our method conditions on the present-day frequency of the allele and is based on the intuition that alleles under purifying selection are on average younger than neutral alleles and, therefore should have higher average levels of haplotype identity among variant carriers. To obtain the probability distribution on the lengths of pairwise haplotype identity, one needs to perform two integrations: one over all possible allele frequency trajectories and another one over all pairwise coalescent times given a certain allele frequency trajectory. The integration over the space of possible allele frequency trajectories is done using a fast importance-sampling algorithm while the integration over the coalescent times is done using an analytical solution. Using the probability of the lengths of the haplotypes under different selective coefficients, we can calculate the likelihood for a selective coefficient for a single variant or set of variants. We use simulations to test how accurately the method estimates the selective coefficient under different demographic scenarios, such as a constant population size and a realistic model of European population growth. Variants with the same selective coefficient are harder to differentiate from neutral variants in scenarios of recent population growth. These methods will be applied to a set of 202 drug target genes sequenced in 14,002 individuals (Nelson et al, 2012, Science) to identify which genes are most likely to harbor damaging variants that may predispose to disease.

1975S

Highlighting strongly differentiated regions using three high coverage genomes each from a set of worldwide human populations. *L. Pagani[1,2,3], T. Kivisild[1].* 1) Division of Biological Anthropology, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom; 2) The Wellcome Trust Sanger Institute, CB10 1SA, Hinxton, UK; 3) Molecular Anthropology Lab, Department of Biological Geological and Environmental Sciences, University of Bologna, Italy.

Following the steady reduction in sequencing costs, several international projects will shortly make available sets of 2-4 high coverage genomes each from hundreds of worldwide human populations. While these resources allow for refining the demographic histories of the studied populations, little can be done to detect signatures of differentiation, possibly driven by natural selection, on these populations. The selection scan methods available to date indeed focus on various genomic components (SNPs, Haplotypes, LD blocks), yet relying on genome frequencies rather than on the full sequence information. Here we show how the top 1% of genic regions analysed using only three genomes each from two populations (CEU and YRI) contains as many as 25% of the top 5% FST candidates obtained using 160 low coverage individuals from the 1000 Genomes Project. The three genomes from each chosen population are combined in three pairs, and FST based on average pairwise differences is calculated between populations. The average FST is computed on a sliding window of 10000 or 50000 bp across all the pop1-pop2 sets of genomic pairs. The top 1% windows showing the highest differentiation were selected and inspected for their gene content. Of the 1785 genes identified by the FST scan based on the 160 low coverage individuals (taken as the gold standard), 98 were found among the 439 genes included in the top 1% 50000bp windows of the YRI-CEU pairs. This 2.4-fold enrichment was found significant with a chi-squared test (p=10-19). The empirical ranking nature of the gold standard did not allow a formal assessment of the false positive rate of our newly developed method. However, the overlap between the top genes retrieved using the 10000 and 50000 bp windows showed a significant enrichment in high ranking FST signals. In summary the proposed approach based on three genomes per population is capable of retrieving at least 25% of the genes under putative natural selection found from traditional methods. Ongoing power assessment will also inform on the optimal number of high coverage genomes per population required to further reduce the false positive rate. These promising results, given the limitations imposed by the small sample sizes, make our method suitable to be applied on newly sequenced populations (expected to be released on Mid June 2014, during the SMBE conference).

1976M

Identification of functional variants under positive selection in the Gullah African American population of South Carolina. *P.S. Ramos[1], S. Sajuthi[2], J. Divers[2], Y. Huang[3], U. Nayak[3], W.-M. Chen[3], K.J. Hunt[1], D.L. Kamen[1], G.S. Gilkeson[1], J.K. Fernandes[1], I.J. Spruill[1], W.T. Garvey[4], M.M. Sale[3], C.D. Langefeld[2].* 1) Department of Medicine, Medical University of South Carolina, Charleston, SC; 2) Department of Biostatistical Sciences and Center for Public Health Genomics, Wake Forest Medicine, Winston-Salem, NC; 3) Department of Medicine and Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 4) Department of Nutrition Sciences and Birmingham VA Medical Center, University of Alabama, Birmingham, AL.

Relative to other African-Americans (AA), the Gullah population has lower European admixture and higher ancestral homogeneity from the Sierra Leone (SL) area in Far-West Africa. The shorter genetic distance between the Gullahs and SL suggests that population genetic signals, such as regions under recent selection, may be more easily detected in the Gullahs than in other AA populations. Since both protein-coding and regulatory variation have important roles in recent human adaptation, the goal of this study was to integrate evidence for natural selection with functional annotation for the identification of biologically relevant signals. We computed the cross population extended haplotype homozygosity test (XP-EHH) to identify alleles with higher than expected frequency relative to their haplotype length in the Gullah (n=277) relative to SL (n=400), to HapMap Phase II YRI (n=203), and CEU (n=165). In total 679,513 SNPs with MAF>5% met standard GWAS QC criteria. Variants that met suggestive significance (|XP-EHH|>4, P<E-04) were annotated and prioritized based on the potential impact of amino acid changes and regulatory functions using RegulomeDB and HaploReg. Nearly the same number of loci showed suggestive evidence for selection between the Gullah and YRI (0.15% of all SNPs), and Gullah and SL (0.14%), although only 106 SNPs in 12 regions showed evidence for selection in both comparisons. Fewer loci showed evidence for selection between Gullah and CEU (0.06%). Enhancer enrichment analysis revealed a significant enrichment of strongest enhancers in H1 human embryonic stem cells. Several regions harbor missense SNPs, including those showing evidence of selection between the Gullah and the YRI (CENPO, FSHR, CHUK, PKD2L1, USP31), Gullah and SL (PRR14L), and Gullah and CEU (ADPRHL2, LCT). Other regions harbor multiple SNPs with high regulatory scores based on the simultaneous presence of eQTLs, transcription factor binding and DNase sites, including those showing evidence of selection between Gullah and YRI (e.g. CCR2, ADCY2, HLA, CD36), Gullah and SL (e.g. CCR2, ADCY2, TNKS), and Gullah and CEU (e.g. NARS2). These results reveal several novel regions with evidence for selection and concomitant high functional potential in the Gullah AA population. Given that many common diseases are more prevalent in specific ethnic groups, identification of functional regions under selection in the Gullah has the potential to elucidate disease risks in AA.

1977S

A model for gene expression level evolution to identify expression conservation, divergence, and diversity. *R. Rohlfs, R. Nielsen.* Integrative Biology, University of California, Berkeley, Berkeley, CA.

As RNA-Seq becomes more available, we see more extensive comparative expression datasets, particularly datasets with multiple individuals sampled per species. These new data facilitate transcriptome-scale analyses of expression variance between and within species. We have developed a phylogenetic model of expression level evolution which can be used to specifically investigate the ratio of within to between species expression variance. We have applied this method to a phylogeny of 15 mammals, mostly primates. We have identified genes with high expression divergence between species as putative targets for expression level adaptation. Conversely, we have identified genes with high expression diversity within species and conservation between species, which likely have plastic responses to environmental inputs. Additionally, we identified genes on the human and catarrhine lineages which show a shift in expression level, putatively due to lineage-specific expression level adaptation. This analysis shows the flexibility and utility of our model in facilitating comparative expression analyses to illuminate the biological meaning of expression levels in humans and other organisms.

1978M

The Clade Fitness Proxy haplotype score: delineating the progression of soft selective sweeps at the haplotype level. *R. Ronen[1], G. Tesler[2], S. Zakov[3], N.A. Rosenberg[4], V. Bafna[3].* 1) Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, CA; 2) Department of Mathematics, University of California, San Diego, La Jolla, CA; 3) Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA; 4) Department of Biology, Stanford University, Stanford, CA.

The dynamics of selective sweeps have long been of interest in population and evolutionary genetics. Of particular interest is the problem of selection on standing variation, when the beneficial allele already segregates in the population at a non-negligible frequency at the time that selection begins. In this case, multiple haplotypes all carrying the beneficial allele will drift relative to one another, while outcompeting other haplotypes. Moreover, the beneficial allele may fix without any of the carrying haplotypes reaching fixation. Consequently, we observe a soft selective sweep, or a weakening of the classical signature of a selective sweep. To identify regions undergoing soft sweeps, we develop a new coalescent-based statistic dubbed the Clade Fitness Proxy (CFP) score, which scores individual haplotypes based on the frequencies of alleles they carry. We develop the statistical properties of the CFP score in the framework of a neutral coalescent model. We use theoretical calculations and simulated data to demonstrate its utility both as a general test of selection and as a proxy for the relative fitness of individual haplotypes. Under a soft sweep, we observe a bimodal distribution of CFP scores, where 'fit' haplotypes have high CFP score and 'unfit' haplotypes have low CFP score. This bimodality can be used in a general test of selection. Perhaps more interestingly, it suggests that CFP scores can serve as a proxy for the relative fitness of individual haplotypes. Thus, the CFP score may be useful for elucidating the state of a selective sweep within a population sample (i.e., at the level of individual haplotypes). Once a sweep has completed and the beneficial mutation has reached fixation (all haplotypes are 'fit'), the in-sample distribution of CFP scores is no longer bimodal. While at this stage the relative fitness of haplotypes is not of interest, as it is by definition equal, the distribution of CFP scores is still highly informative about the occurrence of a sweep in the population's history for many generations thereafter. We expect that the CFP score will be a powerful tool to study regions undergoing soft selective sweeps, and that it will contribute to uncovering the structure of relative fitness within population samples.

1979S

Searching for soft selective sweeps in worldwide human populations. *Z.A. Szpiech[1], R.H. Hernandez[1,2,3].* 1) University of California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 3) Institute for Quantitative Biosciences (QB3), University of California, San Francisco, San Francisco, CA.

There is ample debate about the strength and mode of natural selection that has occurred in recent human evolution. This is particularly so for classical hard sweeps, during which an adaptive allele quickly drags a single haplotype to high frequency. An alternative model of adaptation involves soft sweeps, whereby multiple haplotypes are brought to high frequency (i.e. when a previously segregating neutral or slightly deleterious allele becomes adaptive in a new environment). Existing haplotype-based tests—such as the integrated haplotype score (iHS) that scans for positive selection by tracking the decay of haplotype homozygosity—work under the assumption that a positively selected region will be dominated by a single haplotype. However, iHS is expected to lose power under a soft sweep. Here we develop a statistic, inspired by iHS and recent work in Drosophila population genetics, designed to detect recent soft sweeps by tracking the decay of homozygosity of multiple haplotypes away from a core locus. We evaluate our statistic with rigorous simulations under multiple realistic models of human demography. We find that it has high power to detect both hard and soft sweeps and has improved power compared to iHS. In particular, for a fixed selection coefficient, our simulations suggest that we have greatest power to detect soft sweeps in African populations, which have been understudied to date. We apply this statistic and iHS to a large human genotype dataset of 1,728 unrelated individuals spanning 20 worldwide populations from the 1000 Genomes Project. A large number of regions identified by our statistic are not identified by iHS, in particular in African populations. This suggests a possibly important role of soft sweeps in recent human evolution.

1980M

Natural selection at the melanocortin-3 receptor gene loci. *I. Yoshiuchi.* Dept Diabetes Mellitus and Medicine, Yoshiuchi Medical Diabetes Institute, Kamakura, Kanagawa, Japan.

Obesity is significantly associated with type 2 diabetes mellitus, metabolic syndrome, hypertension, stroke, and cardiovascular diseases. The worldwide prevalence of obesity is increasing steadily. Obesity is highly heritable disease that causes serious health problems. During the traditional cycles of feast and famine, natural selection of obesity-related genes would be significant because these genes control body weight and fat levels. Human adaptation to environmental changes in food supply, lifestyle, and geography may have influenced the selection of genes associated with the metabolism of glucose, lipids, carbohydrates, and energy. The melanocortin-3 receptor (MC3R) gene is one of obesity-associated genes, and MC3R mutations have been shown to be associated with obesity. MC3R-deficient mice showed increased fat mass. Here, We aimed to uncover evidence of selection at the MC3R gene loci. We performed a three-step method to detect selection at the MC3R gene loci using the HapMap population data. We used Wright's F statistics as a measure of population differentiation, the long-range haplotype test to test extended haplotypes, and the integrated haplotype score test to detect selection at the MC3R gene loci. We observed natural selection at the MC3R gene loci by the integrated haplotype score test in the African population. This finding provides evidence of natural selection at the MC3R gene loci. Further discoveries are warranted on the adaptive evolution of obesity-associated genes.

1981S

Identification of putative high altitude adaptation determinants in Tibetan whole genome sequences. *J. Downie[1], H. Hu[2], T. Simonson[3], D. Witherspoon[1], G. Glusman[4], J. Roach[4], G. Cavalleri[5], M. Brunkow[4], A. Cole[5], M. McCormack[5], N. Petousi[6], P. Lorenzo[7], R. Gelinas[4], J. Prchal[7], P. Robbins[6], C. Huff[2], L. Jorde[1].* 1) Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112 USA; 2) Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030 USA; 3) Division of Physiology, Department of Medicine, University of California San Diego, La Jolla, California, USA; 4) Institute for Systems Biology, Family Genomics Group, Seattle, WA 98109 USA; 5) Royal College of Surgeons in Ireland, Molecular and Cellular Therapeutics, St Stephen's Green, Dublin 2, Ireland; 6) University of Oxford, Department of Physiology, Anatomy and Genetics, Oxford, OX1 3PT, United Kingdom; 7) Division of Hematology, University of Utah School of Medicine, Salt Lake City, UT 84112 USA.

Despite its extreme high-altitude environment, the Tibetan Plateau has been inhabited by humans for thousands of years. Tibetans possess heritable adaptations that allow them to better survive in hypoxic conditions. A number of genes that could account for Tibetans' ability to live at high altitude have been found, including *EGLN1* and *EPAS1*, using SNP arrays and low-coverage exome-sequencing. However, there has been little success thus far in finding specific variants responsible for the adaptive traits seen in Tibetans. To address this, we performed Complete Genomics whole-genome sequencing of 17 Tibetans from the United Kingdom (12) and Utah (5). The Tibetan ancestry of these genomes was confirmed by comparing them with genotypes of other Tibetan and Eurasian samples using *EIGENSOFT*. We phased the 17 Tibetan genomes using *SHAPEIT*. Next, the integrated haplotype score (iHS) and cross-population extended haplotype homozygosity (XP-EHH) tests, using 1000 Genomes CHB+CHS as a comparison, were employed to detect selective sweeps. We calculated the number of iHS values ≥ |2.0| and the max XP-EHH value for 200kb non-overlapping windows. We compiled a candidate gene list based on their function in the Gene Ontology and PANTHER to search for adaptive genes. Intersecting the candidate list with genes in the top 1% of iHS regions returned 8 putatively advantageous genes (*MTOR*, *PIK3R3*, *EGLN1*, *EPAS1*, *VHL*, *RAF1*, *UBE2D2, TPTE*) while the top 1% of XP-EHH windows also contained 8 genes (*EGLN1*, *KLRK1*, *CALM1*, *EPAS1*, *EP300*, *TLR2*, *NRF1*, *LOXL2*). We searched for novel, nonsynonymous variants in these genes that might be responsible for high-altitude adaptation but found no such variants at appreciable frequency. Further, described nonsynonymous variants in these genes were found at similar frequency as the Han-Chinese population. In conclusion, we identified both novel and previously described positively selected regions of the genome in Tibetans but found no novel nonsynonymous variants in these regions.

1982M

Positive selection in smallpox associated genes among Mesoamericans. *O.A. Garcia[1], K. Arslanian[2], D. Whorf[1], M. Shriver[3], L.G. Moore[4], T. Brutsaert[5], A.W. Bigham[1].* 1) Department of Anthropology, University of Michigan, Ann Arbor, MI; 2) Department of Anthropology, Yale University, New Haven, CT; 3) Department of Anthropology, Penn State University, University Park, PA; 4) Department of Obstetrics and Gynecology, University of Colorado, Aurora, CO; 5) Department of Exercise Science, Syracuse University, Syracuse, NY.

During the colonization of Mesoamerica, one of the major causes of death was the introduction of novel infectious diseases. Among the most lethal infectious diseases was smallpox. Therefore, studying signatures of natural selection in genes related to smallpox infection and immune response not only provides a window to our evolutionary past but is also a particularly attractive strategy to identify host factors for modern infectious disease. To characterize host risk factors within Mesoamerican populations, we interrogated 906,600 SNPs assayed using the Affymetrix 6.0 genotyping array for signatures of natural selection in 231 immune response genes. Our populations included: Mesoamerican: 25 Maya and 14 Nahua, Mixtec, and Tlapanec speakers from Mexico, Andean: 25 Aymara from Bolivia, and 24 Quechua from Peru. Additionally, we used available data from 60 Europeans of northern European ancestry and 90 East Asians from China and Japan. We applied three statistical tests to identify signatures of natural selection: locus specific branch length (LSBL), the natural log of the ratio of heterozygosities (lnRH), and Tajima's D. Furthermore, we analyzed partial and hard sweeps with two haplotype texts: integrated haplotype score (iHS) and cross population extended haplotype homozygosity (XP-EHH). We determined statistical significance based on an empirical distribution. Among our strongest results for positive selection were *CD74*, *ZAP-70*, and *IKZF1* that were significant in all the statistical tests at the 5% and 1% level for Mesoamericans between East Asians and European Americans comparisons. Furthermore, they were statistically significant in comparison to the Andean populations. *CD74* is major histocompatibility complex class II (MHC II) invariant chain. Studies have shown *CD74*'s protein to function as a receptor for cytokine MIF, a critical immune response factor. *ZAP-70* is an integral part of the T-cell signaling pathway thereby regulating adaptive immune response. Several studies have shown *CD74* and *ZAP-70* expression to be correlated. *IKZF1* has mostly been studied as in autoimmune disorders as part of the pathway regulating haematopoiesis. The results of this study will aid future studies by pinpointing candidate genes for infectious disease susceptibility and resistance in Mesoamerican populations.

1983S

Selection and reduced population size cannot explain higher amounts of Neanderthal ancestry in East Asian than European human populations. *B. Kim[1], K. Lohmueller[1,2].* 1) Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA; 2) Interdepartmental Program in Bioinformatics, University of California Los Angeles, Los Angeles, CA.

Understanding the Neanderthal ancestry of modern humans may provide crucial insights into the evolution of different human populations. It is believed that Neanderthals admixed with European and Asian populations to a much greater degree than with African populations. Additionally, recent studies show a higher frequency of Neanderthal alleles in East Asians relative to Europeans. Several hypotheses to explain this difference have been proposed. One hypothesis posits that there was a single admixture event in the population ancestral to modern Europeans and East Asians and that many of the Neanderthal alleles were weakly deleterious in modern humans. Because East Asians have historically had smaller population sizes than Europeans, purifying selection may have been less effective at removing the Neanderthal alleles from East Asian populations, leading to the observed higher proportion of Neanderthal ancestry in East Asians. Here we test this hypothesis using forward-in-time population genetic simulations. These simulations include plausible models of European and East Asian population history which have been estimated from data as well as models of the fitness effects of Neanderthal alleles in humans that include different dominance scenarios and a distribution of selection coefficients. Starting with the same amount of Neanderthal ancestry in both populations, we find that the differences in population size between European and East Asians combined with purifying selection cannot lead to the observed increase in the amount of Neanderthal ancestry in East Asian populations. Furthermore, when starting with the same initial amount of Neanderthal ancestry in both populations, realistic population size changes alone are insufficient to decrease or increase the Neanderthal ancestry in one population relative to the other. The observed data must be explained by some other process, such as additional waves of Neanderthal admixture into East Asian populations.

1984M

Identifying incomplete selective sweep using Sequential Markov Coalescent. *K. Liao[1], W. Hsieh[2], C. Tang[1].* 1) Department of Computer Science, National Tsing-Hua University, Hsinchu City, Hsinchu, Taiwan; 2) Institute of Statistics, National Tsing-Hua University, Hsinchu City, Hsinchu, Taiwan.

Positive selection leaves signatures in genomes of the population and identifying such signatures provides insight into the genetic basis of how human is different from other species. Among these signatures, long-haplotypes are useful for identifying incomplete selective sweep, but it is usually difficult to compute the likelihood of statistics based on long-haplotype under the neutrality assumption. Hence, the statistical significance must be evaluated by empirical distribution derived from a large number of simulations. Sequentially Markov Coalescent (SMC) is a promising framework of haplotype-based analysis and it has the potential to be extended to various population genetics model. We implemented a long-haplotype based test for incomplete selection under SMC framework. The likelihood of a sample is computed with Product of Approximate Conditionals (PAC) and the statistical significance can be evaluated by the likelihood-ratio test. In addition, the strength of selection is estimated through maximizing PAC-likelihood function. The procedures will be evaluated with simulated data and compared with other methods based long-haplotype signatures.

1985S

Asian diversity project: a survey of population structure and local adaptations in Asian populations. *X. Liu[1,2], D. Lu[3], W.Y. Saw[1], T.H. Ong[1], C. Simmons[4], P. Suriyaphol[5], S. Tongisma[6], B.P. Hoh[7], N. Kato[8], Y.Y. Teo[1,9].* 1) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 2) NUS Graduate School, National University of Singapore, Singapore; 3) Max Planck Independent Research Group on Population Genomics, Chinese Academy of Sciences and Max Planck Society Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences,, Shanghai, China; 4) Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Viet Nam; 5) Division of Bioinformatics and Data Management for Research, Mahidol University, Bangkok, Thailand; 6) Genome Institute, National Center for Genetic Engineering and Biotechnology, Pathumtani, Thailand; 7) Insitute of Medical Molecular Biotechnology (IMMB), Faculty of Medicine, Universiti Teknologi MARA (UiTM) Malaysia, Sg Buloh, Selangor, Malaysia; 8) Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan; 9) Department of Statistics and Applied Probability, National University of Singapore, Singapore.

As the largest continent on Earth, Asia hosts more than 60% of the human populations in the world. Great genetic diversity exists in the Asian populations. The HUGO Pan-Asian SNP consortium provided a valuable genetic resource of Asian populations and performed a thorough survey of genetic diversity and population history of Asian populations. However, the sparse coverage of SNPs made the analysis of natural adaption difficult to perform. In this study, we collected dense genotyping data from 46 populations across Asia. More than 4093 individuals from East Asia, Central Asia, Southeast Asia and South Asia were genotyped on various genotyping platforms. Principal components analysis (PCA) and admixture analysis were performed to elucidate the population structure in ADP populations. It was revealed that geographic played an important role in shaping the population structure of Asian populations; and the ADP populations were further grouped into East Asian, Central Asian, Southeast Asian and South Asian subgroups. We performed a genome wide scan of positive selection signals in the ADP populations using iHS, XP-EHH and haploPS. A total of 669 candidate selection regions were detected across the 46 ADP populations. A PCA analysis on the selection signals were performed to investigate the degree of sharing of the selection signals in the 46 populations. It was found that clustering of populations by selection signals resembles the clustering inferred from population structure analysis. East and Southeast Asian groups share the largest number of selection signals; and the South Asian group possesses distinct selection signals from the rest of the Asian populations. For selection signals shared by multiple populations, we studied the origin of the selection, *ie.* either the selection originated from a single mutation in the common ancestor followed by subsequent gene flow, or it was the result of convergent evolution, where the selection emerged separately from multiple mutation events. The origins of positive selection signals were investigated by calculating the haplotype similarity index. The haplotype similarity index identified 36 selection regions under convergent evolution, and most of them involve aboriginal populations from Southeast Asia.

1986M

The pleiotropic effects of *EDARV370A* in an admixed Uyghur population. *Q. Peng[1], J. Li[1], J. Tan[2,3], Y. Yang[2,3], Y. Guan[4], L. Zhang[4], Y. Jiao[4], P. Sabeti[5,6], L. Jin[1,2,3], S. Wang[1].* 1) CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China; 2) MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai, China; 3) CMC Institute of Health Sciences, Taizhou, Jiangsu Province, China; 4) Department of Biochemistry, Preclinical Medicine College, Xinjiang Medical University, Urumqi, Xinjiang, China; 5) The Broad Institute of Harvard and MIT, Cambridge, USA; 6) Center for Systems Biology, Department of Organismic and Evolutionary Biology.

An adaptive variant of the human Ectodysplasin receptor, *EDARV370A*, showed one of the strongest signals of recent positive selection from genome-wide scans. In transgenic mice and in humans, it is found that *EDARV370A* affects ectodermal related phenotypes, including hair thickness and shape, active sweat gland density, and teeth formation. However, previous human studies were all based on East Asian populations, in which the frequencies of ancestral allele 370V are low. It is inconclusive whether the genetic model of *EDARV370A* is additive or dominant. The lack of power was due to the low presence of 370V homozygotes, which made it impractical to explore a large spectrum of potentially affected ectodermal related phenotypes. In this study, we took advantage of an admixed population between East Asian and European - the Uyghur, to investigate the pleiotropic nature and the genetic model of *EDARV370A*. By examining a series of ectodermal related phenotypes and the *EDARV370A* genotype in 294 Uyghur samples, we replicated the previous association findings in incisors shoveling (P= 5.76×10^{-12}) and hair straightness (P=3.37×10^{-03}), and further confirmed the association is following an additive genetic model. We also found *EDARV370A* associated with novel phenotypes including higher total sweat gland density (P=0.03) and triangular earlobes (P=2.05×10^{-04}). By revealing more pleiotropic effects of *EDARV370A* and confirming its genetic model, our study provides a more complete picture for the adaptive evolution of *EDARV370A* in human history.

1987S

A hidden Markov framework to estimate the timing of selection for hard sweeps. *J. Smith[1], M. Stephens[2], M. Przeworski[3], G. Coop[4], J. Novembre[2].* 1) Department of Ecology and Evolution, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Department of Biological Sciences, Columbia University, New York, NY; 4) Department of Evolution and Ecology, University of California--Davis, Davis, CA.

Dispersal across the globe has resulted in humans occupying a wide range of ecological habitats. Natural selection seems to have played a role in this process, as current methods have identified a number of well supported loci that have undergone a recent selective sweep. In some cases, comparing estimates for the timing of selection with events in the historical/archeological record can provide a more clear picture of the ecological context driving adaptation in a population. For example, an overlap between cultural shifts towards dairy food production with the timing of selection on the lactase persistence allele has helped evaluate a possible cause for the observed selective sweep. As a result, there is substantial interest in methods to infer the age of a positively selected allele. A key principle for allele age estimation is that due to recombination and mutation, the signature of a selective sweep decays at a constant rate per generation. Current methods to estimate the age of selective sweeps either rely on a heuristic estimate of the length of the selected haplotype or employ a simulation-based framework to identify the distribution of ages that produce the observed summary statistics of the complete data. In practice the confidence intervals for these estimates are large. Here, we provide methods for inferring the ancestral haplotype of the selected allele and the recombination breakpoints off of this haplotype in order to provide more refined estimates of allele age. We do so using a hidden Markov model framework which allows us to integrate over uncertainty in recombination breakpoints. This framework uniquely uses both the present day length distribution of the ancestral haplotype and the number of derived mutations to estimate the number of generations since the sweep occurred. The joint use of haplotype lengths and derived mutations increases the total number of observed events and provides more narrow confidence intervals for the age estimate. Using this joint estimator on simulated data, 95% quantiles for estimates of sweep ages from 400 to 500 generations are within 35 generations of the true value. Whereas estimates based on derived mutations or haplotype lengths alone provide 95% quantiles ~70 generations from the true value. Future applications will revisit the timing of selection for lactase persistence in Northern Europeans, skin pigmentation alleles in Europe and Asia, and malaria resistance at the G6PD locus in Africa. .

1988M

Genome wide survey of positive selection signals in African Americans since admixture. *H. Wang[1], Y. Choi[2], X. Wang[3], B. Tayo[4], u. Broeckel[5], C. Hanis[6], S. Kardia[7], S. Redline[8], R. Cooper[4], H. Tang[2], X. Zhu[1].* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Departments of Preventive Medicine, Biomedical Informatics, and Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY; 4) Department of Public Health Science, Loyola University Medical Center, Maywood, IL; 5) Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI; 6) 6 Department of Epidemiology, Human Genetics and Environmental Sciences, University of Texas Health Science Center at Houston, Houston, TX; 7) Department of Epidemiology, University of Michigan, Ann Arbor, MI; 8) Department of Medicine, Harvard Medical School, Boston, MA, USA.

In an admixed population such as African Americans, over or deficient ancestry in a local genomic region may suggest natural selection. We scanned three large African American cohorts of 20,153 individuals but failed to identify any genome-wide significant over or deficient signals. We showed that the failure to identify any significant selection signals can be attributed to the estimated variance of the test, which consists of two components: variance due to sampling error and variance due to genetic random drift. The proportion of variance due to genetic random drift increases when sample size increases. Thus, a test based on examining local ancestry excess is not efficient and its power will not increase when increasing sample size. We also showed that the high correlations of local ancestries between different cohorts are due to the historical recombination and genetic random drift. Assuming African-Americans have been admixed for 8 to 12 generations, we estimated the effective population size as between 32,000 to 48,000.

1989S

Differential purifying and positive selection across genes stratified by X chromosome inactivation status. *A. Slavney[1], F. Gao[2], A. Clark[1,2], A. Keinan[2].* 1) Genetics, Genomics & Development, Cornell University, Ithaca, NY; 2) Biological Statistics & Computational Biology, Cornell University, Ithaca, NY.

In eutherian mammals, dosage compensation between XX females and XY males occurs through ChrX inactivation (XCI). XCI randomly silences transcription from one of the two X chromosomes in each female cell, but at least 25% of human ChrX loci escape XCI to varying degrees, creating three broad gene groups: genes that i) are consistently inactivated, ii) escape XCI in some indivuduals, and iii) consistently escape XCI. Evolutionary mechanisms underlying XCI status are poorly understood, but the predominant model posits that XCI evolved in response to ChrY degeneration. In this model, XCI escapers encode products that are dosage-sensitive and/or female-biased, and thus highly conserved. We therefore tested the hypothesis that mutations in group ii) and iii) genes are more likely to be deleterious than those in group i) genes

To evaluate the effects of mutations in each XCI group, we first used European American single nucleotide variant (SNV) data from the NHLBI Exome Sequencing Project to compare synonymous (S) and non-synonymous (NS) SNV frequency distributions and estimated the individual burden of private mutations (IBPM) for all groups. These analyses showed a greater NS:S excess at low SNV frequencies and greater IBPM estimates in groups ii) and iii) than in i), suggesting stronger conservation in XCI escapers. We then incorporated divergences from chimpanzee to estimate statistics derived from the McDonald-Kreitman test, including the fractions of strongly (d) and weakly deleterious (b) sites, for each group. We found significantly higher d and lower b values in groups ii) and iii) than in group i), which support stronger conservation in escapers. However, we were surprised to find that group ii) showed evidence of experiencing stronger purifying selection than group iii) as per the IBPM, d and b values.

In conclusion, we report that polymorphism and divergence data support a model of XCI escape driven by ancient and ongoing evolutionary conservation, but that heterogeneous escapers show stronger purifying selection than consistent escapers. These results suggest a potentially important role for XCI escapers in phenotypes that affect fitness, but also demonstrate a need for further investigation into the true extent of XCI profile heterogeneity among tissues and individuals.

1990M

Inference of Neandertal gene expression from modern samples. *E.E. Quillen, M.A.A. Almeida, J.M. Peralta, J.E. Curran, M.P. Johnson, H.H.H. Goring, R. Duggirala, S. Williams-Blangero, J. Blangero.* Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX.

The sequencing of multiple Neandertal specimens has opened up the possibility of estimating Neandertal phenotypes based on the effect of "Neandertal alleles" present at low frequency in modern human populations. To capture a portion of the functional effects of Neandertal variants, we consider whole genome sequence and mRNA expression profiles obtained using an Illumina Sentrix Human Whole Genome BeadChip microarray from lymphocytes drawn from 921 participants in the San Antonio Family Study (SAFS). 56,294 single nucleotide variants (SNVs) were identified as monomorphic in the three Vindja Neandertal genomes, but polymorphic in the SAFS. Of these SNVs, 2,950 are located upstream of genes represented by high-quality, heritable probes on the microarray in a putatively cis-regulatory position. 81 SNV-probe pairs were found to be significantly associated ($p < 1E-5$) in a variance components-based association analysis performed in SOLAR. Additionally, the effect of genome-wide Neandertal ancestry on transcript expression levels was considered. The full set of SNVs were used to generate a covariance matrix derived from LDAK which calculates kinship coefficients while correcting for linkage disequilibrium. A variance component model was fit for each of the 2,869 probes and the variance explained by the Neandertal ancestry covariance matrix was determined. Neadertal ancestry had significant ($p < 0.05$) explanatory power for 61 probes. To contextualize these results, all significant probes were considered in a gene-set enrichment analysis of biological domains defined from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Enrichment was determined by an empirical p-value derived from comparing the test set of significantly associated mRNA transcripts to 1000 randomly selected sets of equal numbers of transcripts represented on the array. These results indicate an enrichment of related to the growth, development, and innervation of the epithelium and immune-related pathways among probes associated with Neandertal ancestry. With these results and additional data which will become available as more Neandertal sequencing occurs, we can infer a portion of a hypothetical transcriptome of our hominid relatives and better understand how Neandertals and modern human populations differ.

1991S

Khoisan hunter-gatherers have been the largest population throughout most of modern human demographic history. *HL. Kim[1,2], A. Ratan[2], GH. Perry[3], A. Montenegro[4,5], W. Miller[2], SC. Schuster[1,2].* 1) Singapore Centre on Environmental Life Sciences Engineering, Nanyang Technological University, Singapore; 2) Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, PA, USA; 3) Department of Anthropology, Pennsylvania State University, PA, USA; 4) Department of Geography, Ohio State University, OH, USA; 5) Campus do Litoral Paulista, Unesp - Univ Estadual Paulista, Brazil.

We sequenced the complete genome sequences of five Khoisan hunter-gatherers from the Kalahari Desert and one Bantu-speaking agriculturalist individual also from southern Africa, with a high accuracy. Compared the 420K SNP genotyping dataset from 490 worldwide individuals, admixture analyses showed that three of our Khoisan genomes from the Ju/'hoansi group (northern Khoisan) have no or minimal admixture from non-Khoisan populations, allowing us to assess the early demographic history of the human species. Population genomic analyses for our complete genome sequences along with those from eight non-Khoisan humans were performed to infer their effective population sizes and demonstrated that the Ju/'hoansi population have maintained their large effective population size and been the people most isolated from all the other human populations, since the earliest population split between the Khoisan and other populations ~100-150 thousand years ago (kya). In contrast, all other human populations, including the ancestral Bantu-speaking agriculturists (currently the largest population within Africa in terms of census size), have experienced severe bottlenecks and lost more than half of their genetic diversity from ~120 to 30 kya. According to paleoclimate records and models, west-central Africa became drier, while southern Africa experienced increases in precipitation, ~80-100 kya. We hypothesize that these climate differences might be related to the divergent ancestral population history within African human populations.

1992M

Insights from low-coverage whole Y chromosome sequencing of 1,244 individuals. *Y. Xue for The 1000 Genomes Project 1000Y Group.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

The 1000 Genomes Project Phase 3 has sequenced 1,244 males belonging to 26 populations from Africa, South and East Asia, Europe and the Americas. In addition to these low coverage (~4-6x) sequences, the project also includes over two hundred males sequenced to high coverage by Complete Genomics and several sets of SNP genotype data used for validation. The group generated a union set of 80,895 Y-SNP, 2,830 Y-MNP and 6,076 short indel calls by combining results from seven different callers. Approximately 6,000 Y-STRs and ~1,000 large structural variants were also called, using two callers for each class. Validation suggests that the Y-SNP, Y-STR and large deletion and duplication calls have very high quality, but that the Y-MNPs and indels do not. Using 59,666 stable high confidence Y-SNPs, we have constructed a phylogenetic tree, to which the more complex classes of variant can be added. The tree recapitulates and extends the established phylogeny. It confirms a very rapid Paleolithic expansion (in number) of Y lineages post-dating the movement out of Africa, and Neolithic or later expansions of independent Y lineages in Africa, Europe, East Asia and South Asia. We observe different patterns in different continental regions, suggesting that this male expansion was extremely rapid in Europe, rapid in Africa, and less rapid in South and East Asia. These data thus provide powerful new insights into male evolutionary history and promise further insights into Y-chromosomal mutation and selection processes.

1993S

Exploring the Y-Chromosome Variation of Modern Panamanians. *A. Achilli[1], V. Battaglia[2], V. Grugni[2], U.A. Perego[1,3], H. Lancioni[1], M. Tribaldos[4], A. Olivieri[2], I. Cardinali[1], E. Rizzi[5], A. Raveane[2], M.R. Capodiferro[1], S.R. Woodward[3,6], J.M. Pascale[4], R. Cooke[7], N. Myres[3,6], A. Torroni[2], J. Motta[4], O. Semino[2].* 1) Dept. of Chemistry, Biology and Biotechnology, University of Perugia, Perugia, Italy; 2) Dept. of Biology and Biotechnology, University of Pavia, Pavia, Italy; 3) Sorenson Molecular Genealogy Foundation, Salt Lake City, Utah, USA; 4) Gorgas Memorial Institute for Health Studies, Panama City, Panama; 5) Institute of Biomedical Technologie, National Research Council (ITB-CNR), Segrate, Milano, Italy; 6) Ancestry, Provo, Utah, USA; 7) Smithsonian Tropical Research Institute, Panama City, Panama.

The Isthmus of Panama - the narrow neck of land connecting the northern and southern American landmasses - was a forced corridor for the Paleo-Indian expansion that originated from Beringia ~15-17,000 years ago. Archaeological findings suggest that some descendants of the earliest migrants remained on the isthmus, while accounts from early European explorers witnessed the presence of two main indigenous groups (the Cueva and the Coclé) in pre-Columbian times - populations that have since disappeared due to disease, warfare, and enslavement following the Spanish conquest. Today's indigenous groups total about 5.3% of the Panamanian population, and are mainly represented by the Ngöbé, Buglé, Kuna, Emberá, and Wounan tribes, which traditionally appear to have settled in Panama from surrounding regions after the autochthonous natives were decimated. However, there is no evidence that the ancestral indigenous gene pool was completely replaced. If this was the case, the populations of modern Panama should have retained at least a fraction of the native pre-Columbian gene pool, possibly at a variable extent, given the differential degree of geographical and genetic isolation of the different Panamanian communities during the past five centuries. A recent study of the mtDNA history of the modern Panamanian population (Perego et al., 2012), based on a sample of 1565 individuals with Native American maternal ancestry, concludes that (1) the first settlement of Panama occurred quite rapidly after the initial colonization of the American continent, (2) based on complete sequence analyses, the founder ages of the most common lineages point to an ancient expansion supporting the antiquity of the Pacific coastal route, 3) the mitochondrial gene pool exemplifies the link between pre- Columbian and modern Panamanian populations (in fact, 83% of modern Panamanians clusters into native pan-American lineages). It appears that the Spanish conquistadores and additional more recent European demographic influences did not contribute significantly to today's genetic composition of Panama, at least with regard to the maternal side. In this study, we have now tested the same scenario from the paternal side by employing the analysis of the Y-chromosome variation in modern Panamanians.

1994M

Population specific patterns of novel haplotype groups at the *PAH* locus. *G. Povysil[1], S. Wieser[2], S. Hochreiter[1], J. Zschocke[2].* 1) Institute of Bioinformatics, Johannes Kepler University Linz, Linz, Austria; 2) Division of Human Genetics, Medical University Innsbruck, Innsbruck, Austria.

The phenylalanine hydroxylase (*PAH*) gene is of particular interest for population genetic studies because the distribution patterns of well-defined phenylketonuria (PKU [MIM 261600]) mutations can be linked with distinct SNP haplotypes for the assessment of ancient migration. Through family segregation analysis and molecular haplotyping with long-range PCR in PKU patients (232 *PAH* mutant alleles) and controls (157 *PAH* normal alleles) from various European countries we identified five major haplotypes in the distal 15 kb region of the *PAH* gene. Haplotypes differ by 3-16 specific SNPs each and have been quite stable over the last millennia. The 29 common European PKU mutations can be linked to specific haplotypes with little evidence of recombination in the *PAH* gene. The results were compared to available sequencing data of Africans, East Asians, Europeans, and Admixed Americans from the 1000 Genomes Project. Additional data from chimpanzee, orangutan, and macaque, as well as high coverage sequences of Neandertal and Denisova, were used in conjunction with these data to establish a possible evolutionary tree of haplotype emergence. There are five major distal *PAH* haplotypes that can be found in all continental populations, but at different relative frequencies. For Europeans and Asians they make up more than 98% of all *PAH* alleles. While Europeans have comparably high frequencies for all of them, the most common haplotype in Asians amounts to almost 78% of *PAH* alleles. In contrast, Africans have many very rare haplotypes that can only be found in Africans, or Africans and Admixed Americans. The ancestral haplotype that matches the sequences of chimpanzee, orangutan, and macaque, has only been found in Africans and one Admixed American individual. The haplotype that matches the Neandertal and Denisova sequence can be found in Africans, Admixed Americans and one Asian individual. Interestingly, additional variations on the Asian haplotype are more similar to the Neandertal than to other present day individuals with this haplotype. The combination of disease mutations and common gene variants with molecular haplotyping and available genetic data from different countries allows a unique insight in the genetic history of human populations.

1995S

High-coverage sequencing of diverse Y chromosomes via in-solution capture. *A.F. Adams[1], G.D. Poznik[1], R.E.W. Berl[2], N.A. Hammond[1], F.L. Mendez[1], O.E. Cornejo[2], M. Napel[1], M. Schertler[1], P.A. Underhill[1], M.W. Feldman[1], C.D. Bustamante[1].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) School of Biological Sciences, Washington State University, Pullman, WA.

As the longest stretch of non-recombining DNA in the human genome, the Y chromosome provides unique insight into the demographic and evolutionary history of modern humans. However, most large-scale studies of human Y chromosome diversity have been confined to a small number of STRs and SNVs. We conduct high-throughput sequencing of Y chromosomes to study male-line demography and evolution.

Having previously determined which regions of the Y chromosome are amenable to short-read sequencing, we designed a set of NimbleGen biotinylated DNA probes to target these 10.4 Mb. We then performed in-solution, hybridization-based capture, using 105 individuals from the 1000 Genomes Project to optimize and validate our capture protocol. Because our approach yields higher coverage than that obtained by the 1000 Genomes Project, we are able to increase resolution at the tips of the phylogeny—the portion most informative of recent demography, such as admixture within the Americas. Following protocol optimization, we performed Y-chromosome capture on 34 saliva-extracted samples from three southern Ethiopian minority ethnic groups. These data provide an unprecedented look at the population structure of southern Ethiopia.

Our capture protocol can increase coverage of the Y-chromosome target regions by two orders of magnitude, as compared to whole-genome shotgun sequencing. We achieve >30× coverage of the targeted region for each of 24 samples multiplexed in a single HiSeq lane. We report some of the first high-coverage sequencing of the hgA and hgA0 lineages and uncover deep novel substructure within the poorly characterized E2 haplogroup. We are now working to apply this capture method to a large panel of individuals from a wide range of populations in order to gain a greater understanding of human Y chromosomal diversity.

1996M

The Kalash isolate from Pakistan. *Q. Ayub[1], L. Pagani[1,2], M. Mezzavilla[1,3], C. Tyler-Smith[1].* 1) The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom; 2) Division of Biological Anthropology, University of Cambridge, Cambridge, United Kingdom; 3) Institute for Maternal and Child Health — IRCCS "BurloGarofolo" — Trieste, University of Trieste, Trieste, Italy.

The Kalash represent an enigmatic isolated population that has been living for centuries in the Hindu Kush mountain ranges of present-day Pakistan. Previous uni-parental (Y and mitochondrial) DNA markers provided no support for their claimed Greek descent following invasion of this region by Alexander III of Macedon, and analysis of autosomal loci provide evidence of a strong genetic bottle-neck. To understand their origins and demography further, we genotyped 23 unrelated Kalash samples on the IlluminaHumanOmni2.5 BeadChip and sequenced a male individual at high coverage on an Illumina Hi-Seq 2000. Comparisons with neighboring populations confirmed results based on genotyping 650,000 common single-nucleotide polymorphisms in the Kalash samples from the Centre Etude Polymorphism Humain (CEPH) Human Genome Diversity Project (HGDP) Cell Line Panel. However, we observed no evidence for admixture as suggested recently by Hellenthal et al. The mean time of divergence between Kalash and other populations currently residing in this region, that also speak Indo-European languages, was estimated to be 11.8 (10.6 -12.6) KYA. Since the split the Kalash have experienced little, or no, gene flow from their geographic neighbors and have maintained a low long-term effective population size (2,247-2,780). They could represent some of the earliest migrants into the Indian sub-continent.

1997S

Identifiability and efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. *A. Bhaskar[1,2], Y.X.R. Wang[3], Y.S. Song[1,2,3,4].* 1) Simons Institute for the Theory of Computing, University of California, Berkeley, Berkeley, CA; 2) Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA; 3) Department of Statistics, University of California, Berkeley, Berkeley, CA; 4) Department of Integrative Biology, University of California, Berkeley, Berkeley, CA.

Several recent large-sample human genetics studies have found a massive excess of rare variants compared to predictions of previously inferred demographic models of human history. A widely cited explanation is that such polymorphism patterns are indicative of explosive and accelerating population growth in recent human history. Using the site frequency spectrum (SFS), a summary of genetic variation in a set of sequences that counts the segregating sites as a function of the mutant allele frequency, we develop an efficient method for inferring recent population demography that can scale to samples involving tens or hundreds of thousands of individuals. Using analytic results for the expected SFS under the coalescent and by leveraging the technique of automatic differentiation, we develop a very efficient algorithm to infer piecewise-exponential models of the historical effective population size from the distribution of sample allele frequencies. Our method is orders of magnitude faster than previous demographic inference methods based on the frequency spectrum and can also accurately estimate locus-specific mutation rates. We show that our method can accurately infer multiple recent epochs of rapid exponential growth, a signal which is difficult to pick up with small sample sizes. We apply our method to a recent large-sample exome-sequencing dataset of 11,000 European individuals and find evidence of rapid recent exponential population growth of 1.5% per generation during the last 370 generations. We also study the statistical identifiability aspect of this inference problem. It has been recently shown that very different population demographies can generate the same SFS for arbitrarily large sample sizes. Although in principle this non-identifiability issue poses a thorny challenge to statistical inference, the population size functions involved in these counterexamples are arguably not biologically realistic. We revisit this problem and show that the SFS of even moderate-sized samples uniquely determines the population demography when the population size is *piecewise-defined* with each piece belonging to some family of biologically-motivated functions. In the cases of piecewise-constant, piecewise-exponential, and piecewise-generalized-exponential models, which are often assumed in population genomic inferences, we provide explicit values for the sample sizes that are sufficient for identifying the demographic model from the SFS.

1998M

Does genetic hypermutability contribute to the current prevalence of connexin 26 deafness? *D.C. Braun[1], E.A. Craft[1], B.K. Herold[1], K.S. Arnos[1], M. Tekin[2,3], A. Pandya[4].* 1) Department of Science, Technology, and Mathematics, Gallaudet University, Washington, DC; 2) John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 3) Department of Pediatrics, Ankara University School of Medicine, Ankara, Turkey; 4) Department of Human Genetics, Virginia Commonwealth University, Richmond, VA.

Approximately 35% of congenital deafness in North America results from recessive mutations in the *GJB2* (connexin 26) gene, making it the most prevalent cause for hereditary deafness [MIM 220290] in this population. There are more than 100 known *GJB2* mutations, placing it in the top 1% of mutational diversities reported by the Human Gene Mutation Database. Another remarkable characteristic of *GJB2* is that some mutations have ethnic associations: 35delG in Caucasians, 235delC in Asians, and 167delT in Ashkenazi Jews. These associations would normally suggest single origins in founders. Some researchers have proposed that *GJB2* is hypermutable. For example, the 35delG mutation occurs within a homopolymeric run of six guanines and could have arisen repeatedly due to strand slippage. *GJB2* also possesses sequence motifs associated with hypermutability in humans or other organisms. The hypermutability hypothesis is important because it could explain the prevalence of *GJB2* deafness. To test the hypothesis of generalized hypermutability in *GJB2*, we analyzed the historical mutation rate of *GJB2* by directly sequencing an 8 kilobyte region of genomic DNA, containing *GJB2*, from participant samples from two different repositories: Virginia Commonwealth University, which holds the largest repository of genetic deafness in North America, and Ankara University in Turkey. First, we directly measured this substitution rate using orangutan and chimpanzee outgroups and their known evolutionary distances. Second, we reconstructed the molecular genealogy using BEAST software, which provided inference of the substitution rates. These analyses generated values that fell within average human genome-wide mutational rates reported elsewhere. Next, we examined whether 35delG occurred on different haplotypic backgrounds. Consistent with hypermutability, we discovered two backgrounds, which we named haplogroups A and B. All North American participants with 35delG carried haplogroup A, whereas Turkish samples carried haplogroup A or B. We conclude that the 8 kb locus containing *GJB2* is not generally hypermutable. However, our discovery of two 8 kb haplogroups for 35delG confirms that it has multiple origins and therefore, strand-slippage hypermutability within the hexaguanine run may drive the prevalence of 35delG mutations. Our research is significant because it directly addresses the causes for the prevalence of *GJB2* deafness, which is not yet understood.

1999S

Identity by descent segments within and across worldwide populations from sequence data. *S.R. Browning[1], B.L. Browning[1,2].* 1) Department of Biostatistics, University of Washington, Seattle, WA; 2) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA.

Segments of identity by descent (IBD) shared by individuals within and across populations provide information on key aspects of demographic history, such as effective population sizes and migration rates.

Sequence data present opportunities and challenges for IBD analysis. Sequence data are more informative than SNP array data, improving power to accurately detect smaller IBD segments and hence obtain higher levels of information about demographic history. On the other hand, low-coverage sequence data have high rates of error, whereas SNP array data are usually extremely accurate.

We recently developed two IBD segment detection methods: Refined IBD and IBDseq. Refined IBD is a haplotype-frequency-based method designed for SNP array data, while IBDseq is an allele-frequency-based method designed for low-coverage sequence data. Both methods were developed in the context of samples from a homogeneous population. When using frequency-based methods in a heterogeneous setting we expect increased rates of false-positive IBD within sub-populations.

We use 1000 Genomes Project data and simulated data to investigate the performance of the IBDseq and Refined IBD methods when analyzing sequence data from world-wide populations. We find that the allele-frequency-based IBDseq method suffers from increased rates of false positive detected IBD segments due to population heterogeneity, whereas the haplotype-frequency-based Refined IBD approach is much less affected. We develop a strategy using multiple runs of Refined IBD and a process of filling small gaps between adjacent detected segments in order to recover near-complete large IBD segments while having high power to detect short segments. Our approach enables powerful IBD detection in the 1000 Genomes project data.

2000M

The Population Genomic Landscape of Human Genetic Structure, Admixture History and Local Adaptation in Peninsular Malaysia. *L. Deng[1], B. Hoh[2], D. Lu[1], R. Fu[1], M. Phipps[3], S. Li[4], A. Nur-Shafawati[5], W. Hatin[6], E. Ismail[7], S. Mokhtar[2], L. Jin[4], B. Zilfalil[5], C. Marshall[8], S. Scherer[8,9], F. Al-Mulla[10], S. Xu[1].* 1) Max Planck Independent Research Group on Population Genomics, Chinese Academy of Sciences and Max Planck Society (CAS-MPG) Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shangh; 2) Institute of Medical Molecular Biotechnology, Faculty of Medicine, Universiti Teknologi MARA, Sungai Buloh Campus, Jalan Hospital, 47000, Sungai Buloh, Selangor, Malaysia; 3) Jeffrey Cheah School of Medicine and Health Sciences, Monash University (Sunway Campus), Selangor 46150, Malaysia; 4) Ministry of Education (MOE) Key Laboratory of Contemporary Anthropology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China; 5) Department of Pediatrics, School of Medical Sciences, Universiti Sains Malaysia, Kelantan 16150, Malaysia; 6) Human Genome Center, School of Medical Sciences, Universiti Sains Malaysia, Kelantan 16150, Malaysia; 7) School of Biosciences & Biotechnology, Faculty of Science & Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia; 8) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 9) McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, Canada; 10) Department of Pathology, Faculty of Medicine, Kuwait University, Safat 13110, Kuwait.

Peninsular Malaysia is a strategic region which might have played an important role in the initial peopling and subsequent human migrations in Asia. However, the genetic diversity and history of human populations—especially indigenous populations—inhabiting this area remain poorly understood. Here, we conducted a genome-wide study using over 900,000 single nucleotide polymorphisms (SNPs) in four major Malaysian ethnic groups (MEGs; Malay, Proto-Malay, Senoi and Negrito), and made comparisons of 17 world-wide populations. Our data revealed that Peninsular Malaysia has greater genetic diversity corresponding to its role as a contact zone of both early and recent human migrations in Asia. However, each single Orang Asli (indigenous) group was less diverse with a smaller effective population size (N_e) than a European or an East Asian population, indicating a substantial isolation of some duration for these groups. All four MEGs were genetically more similar to Asian populations than to other continental groups, and the divergence time between MEGs and East Asian populations (12,000—6,000 years ago) was also much shorter than that between East Asians and Europeans. Thus, Malaysian Orang Asli groups, despite their significantly different features, may share a common origin with the other Asian groups. Nevertheless, we identified traces of recent gene flow from non-Asians to MEGs. Finally, natural selection signatures were detected in a batch of genes associated with immune response, human height, skin pigmentation, hair and facial morphology and blood pressure in MEGs. Notable examples include *SYN3* which is associated with human height in all Orang Asli groups, a height-related gene (*PNPT1*) and two blood pressure-related genes (*CDH13* and *PAX5*) in Negritos. We conclude that a long isolation period, subsequent gene flow and local adaptations have jointly shaped the genetic architectures of MEGs, and this study provides insight into the peopling and human migration history in Southeast Asia.

2001S

Fine-Scale Genetic Structure in the Open Population of Western France. *C. Dina[1,2,3,4], F. Simonet[1, 2,3], P. Olivier[1,2,3], J. Violleau[1,2,3,4], F. Gros[1,2,3], S. Lecointe[1,2,3,4], S. Küry[5], S. Bézieau[5], H. Le Marec[1,2,3,4], J-J. Schott[1,2,3,4], M. Karakachoff[1,2,3,6], R. Redon[1,2,3,4].* 1) institut du Thorax, Nantes, France; 2) CNRS, UMR 6291, Nantes, France;; 3) Université de Nantes, Nantes, France;; 4) CHU Nantes, l'institut du thorax, Service de Cardiologie, Nantes, France;; 5) CHU Nantes, Service de Génétique Médicale, 44093 Nantes CEDEX 1, France; 6) Institute of Clinical Physiology, National Research Council, Pisa, Italy;.

The 'Common Variant - Common Disease' hypothesis was only partly verified through effective discovery of statistical associations. This empirical observation led the research community to reconsider the involvement of rare genetic variation in predisposition to common disease. Rare alleles of recent origin are likely to cluster geographically in communities with limited migration rates, such as the French rural populations before the 20th Century. Should these rare alleles strongly increase disease risk, one would expect outbreaks of disease prevalence and enhanced power to establish the variant-phenotype relationship. Recently, we demonstrated genetic structure at the level of Brittany and Anjou (Karakachoff, 2014). In two new datasets, we propose to confirm this observation at a much finer scale. The first dataset (VACARME) includes 190 individuals which have their four grand-parents born within a distance of less than 15 kilometers. The second dataset includes 170 individuals born in the same area. Geographical location of individuals used to perform spatial analysis was defined according to the declared place of birth (or that of their grand-parents). The birthplaces are restricted in a small area spanning 200x200 kilometers in Western France. Individuals were genotyped on Affymetrix Axiom genome-wide CEU-1 array plates (1st dataset) and CytoChip 370k (2nd dataset). In the preliminary Principal Components Analysis, the first two principal components revealed an important correlation with latitude ($p<1e-16$) and longitude ($p<1.e-06$) with a better correlation for grand-parent's place of birth. Moreover, part of the distinction is possibly attributable the Loire river acting as a barrier. These preliminary results show for the first time, to our knowledge, that population stratification can be observed even at a very fine geographical level in populations usually considered as open and panmictic. This strongly suggests that recent rare alleles are likely to cluster geographically even in these populations. Therefore, identification of rare variants inducing disease susceptibility can benefit from a strategy focusing on small geographic units. Moreover, we will apply spatially explicit methods like spatial-PCA in order to better discriminate between clinal structure and presence of genetics clusters. We are now investigating through simulations which allele frequency spectrum can be expected to cluster geographically given this differentiation pattern.

2002M

Shared Identity by Descent segments within current Italian population reveals new details about recent population history. *G. Fiorito[1,2], C. Di Gaetano[1,2], F. Rosa[1], S. Guarrera[1,2], B. Pardini[1], A. Piazza[1,2], G. Matullo[1,2].* 1) Human Genetics Foundation, Turin, TORINO, Italy; 2) Department of Medical Sciences, University of Turin, Turin, Italy.

The inference of Identity by Descent (IBD) shared segments were recently enabled by high-resolution genomic data from large cohorts and novel algorithms for IBD detection. This approach permits to examine more in detail the genetic structure of a population as well as to get information about recent demographic events such as bottlenecks and migrations. This study aims to characterize the genetic variability within the Italian population. We present analytical results on the relationship between IBD sharing across 301 unrelated Italian individuals genotyped for about 2.5 million Single Nucleotide Polymorphisms (SNPs). Each sample has well-defined geographical origins (four grandparents coming from the same geographical region). Due to the well-known common ancestral origin of the Italian population we focused our attention on long-range and relatively recent shared IBD segments. By using Principal Component Analysis (PCA) and ancestry estimation, we ascertain Sardinia as the genetic outlier within Italy. Moreover a certain degree of differentiation is still detectable within Aosta Valley population. For each of the 11 subpopulation, we find a significant highest number of shared IBD segments within vs. between population, suggesting isolation by distance. Samples sharing the highest number of internal IBD blocks are Sardinian as expected, followed by those living in Aosta Valley, Tuscany and Sicily. We also evaluate the relationship between shared IBD segments and geographical distance. Contrary to what is expected, the decay of IBD with distance is not steeper for longer (recent) blocks. Such result suggests a constant exchange due to several migratory waves within Italy and/or to the considerable high number of population that have lived in Italy. We finally demonstrate that regions of increased IBD sharing are enriched for structural variation and loci implicated in natural selection and we highlighted the relationship between shared IBD haplotypes and demographic events occurred both in Sardinia and in the Italian peninsula. In conclusion, our results suggest that the study of shared IBD segments between populations is a useful method to detect novel details about relatively recent population history.

2003S

Reconstruction of ancestral human haplotypes using genetic and genealogical data. *J.M. Granka[1], R.E. Curtis[2], K. Noto[1], Y. Wang[1], J.K. Byrnes[1], M.J. Barber[1], N.M. Myres[2], C.A. Ball[1], K.G. Chahine[2].* 1) AncestryDNA, San Francisco, CA; 2) AncestryDNA, Provo, UT.

The genomes of individuals who lived long ago may persist in modern populations in the form of genomic segments broken down by recombination and inherited by their descendants. We develop a novel computational method to reconstruct the chromosomal haplotypes of human ancestors given genetic data from a sufficient number of their present-day descendants. After genealogical information is used to identify descendants of an ancestor (and therefore also of their partner), we use phased genome-wide single nucleotide polymorphism (SNP) data to find regions of the genome that are identical-by-descent (IBD) among them. We develop a novel stitching algorithm to reconstruct up to four chromosomal haplotypes of an ancestor and their partner given the descendant IBD segments and haplotypes. The method aims to remove spurious IBD segments caused by false inference, inaccurate genealogical information, or multiple common ancestors. Short regions of the genome with missing data can also be imputed given flanking reconstructed haplotypes. Lastly, given descendants of other individuals related to the ancestral couple, we show that it is sometimes possible to tease apart the personal identity of each of the reconstructed chromosomal haplotypes (i.e., which are the ancestor's, and which are their partner's). Through simulations, we calculate the amount of the genome that can theoretically be reconstructed given the number of generations back to the ancestor and the number of actual and sampled descendants. Given sufficient data, we can reliably reconstruct the haplotypes of in silico ancestors with high precision and recall; performance is sensitive to genealogical tree quality and accuracy of inferred IBD. We apply our method to phased genome-wide SNP data, obtained from the AncestryDNA customer database, from several hundred individuals descended from an 18th century couple. In genomic regions with many inferred IBD segments, we can reconstruct the haplotypes of the couple, in some cases assigning each haplotype to a specific member of the pair. In regions of the genome with fewer segments, we are less able to discover all haplotypes with certainty. Finally, we demonstrate that given these reconstructed haplotypes, we can infer a given ancestor's ancestry and select physical features. Our study highlights the feasibility of reconstructing the genomes of human ancestors and has immediate applications in population genetics, medical genetics, and genealogy research.

2004M

Identity by descent between humans, Denisovans, and Neandertals. *S. Hochreiter, G. Povysil.* Institute of Bioinformatics, Johannes Kepler University Linz, Linz, Austria.

We analyze the sharing of very short identity by descent (IBD) segments between humans, Neandertals, and Denisovans to gain new insights into their demographic history. Short IBD segments convey information about events far back in time because the shorter IBD segments are, the older they are assumed to be. The identification of short IBD segments becomes possible through next generation sequencing (NGS), which offers high variant density and reports variants of all frequencies. Only recently HapFABIA has been proposed as the first method for detecting very short IBD segments in NGS data. HapFABIA utilizes rare variants to identify IBD segments with a low false discovery rate. We applied HapFABIA to the 1000 Genomes Project whole genome sequencing data to identify IBD segments which are shared within and between populations. Some IBD segments are shared with the reconstructed ancestral genome of humans and other primates. These segments are tagged by rare variants, consequently some rare variants have to be very old. Other IBD segments are also old since they are shared with Neandertals or Denisovans, which explains their shorter lengths. The Denisova genome most prominently matched IBD segments that are shared by Asians. Many of these segments were found exclusively in Asians and they are longer than segments shared between other continental populations and the Denisova genome. Therefore, we could confirm an introgression from Deniosvans into ancestors of Asians after their migration out of Africa. While Neandertal-matching IBD segments are most often shared by Asians, Europeans share more than other populations, too. Again, many of the Neandertal-matching IBD segments are found exclusively in Asians, whereas Neandertal-matching IBD segments that are shared by Europeans are often found in other populations, too. Neandertal-matching IBD segments that are shared by Asians or Europeans are longer than those observed in Africans. This hints at a gene flow from Neandertals into ancestors of Asians and Europeans after they left Africa. Interestingly, many Neandertal- or Denisova-matching IBD segments are predominantly observed in Africans - some of them even exclusively. IBD segments shared between Africans and Neandertals or Denisovans are strikingly short, therefore we assume that they are very old. This may indicate that these segments stem from ancestors of humans, Neandertals, and Denisovans and have survived in Africans.

2005S

Inferring demographic history from whole genome using Approximate Bayesian Computation. *F. Jay, F. Austerlitz.* Laboratoire Eco-Anthropologie et Ethnobiologie, Muséum National d'Histoire Naturelle, Paris, Paris, France.

Reconstructing past demography from neutral genetic data is essential as it both improves our knowledge about human history and provides an accurate neutral model against which selective hypotheses can be tested. Approximate Bayesian Computation (ABC) has proven to be useful for inferring demography from microsatellite or SNP data. This approach consists in simulating genetic data under a large range of complex demographic scenarios and realistic biological processes. Simulations are then compared to observed data using a set of informative summary statistics. Whole-genome data are expected to be extremely rich in information about past demography but, because simulations were, until recently, computationally too costly, ABC methods have not been thoroughly tested on such very long sequences. Dense polymorphism data contain extra information that is not available from unlinked site polymorphisms, and will, therefore, hopefully improve the reconstruction of demographic history. They allow computing specific statistics, such as the decay of linkage disequilibrium with distance, the distribution of length of haplotypes shared between two or more individuals, or the "allele frequency identity-by-state" as described by Theunert et al. (2012). The power given by some of these statistics to infer demographic parameters has been investigated. However, studies were done independently on single classes of statistics, and not always under the approximate Bayesian framework.

Here, we examine how combining these "dense data statistics", with "classical statistics" (e.g. pairwise differences, heterozygosity) in an ABC framework improves the inference of demographic history. Furthermore, we describe how sequencing errors that are usually more frequent in full sequences than SNP data impact the summary statistics and the ABC inference. To diminish these effects we propose to either (i) filter data drastically and prune summary statistics that are highly sensitive to errors, or (ii) model errors within our ABC and incorporate them into simulations. We benchmark these different approaches for simple demographic scenarios, and focus more specifically on population expansion events that happened in recent human history.

Theunert C et al (2012) Inferring the history of population size change from genome-wide SNP data. MBE, 29, 3653-3667.

2006M

Genetic Structure of North-Indian Punjabi Population Based on Autosomal Microsatellite Loci. *M. Kaur, B. Badaruddoza.* Deptt. of Human Genetics, Guru Nanak Dev University, Amritsar, India.

The population of Punjab, India, possesses an exclusive genetic profile, primarily due to the many migratory events in this region which caused an extensive range of genetic diversity. Hence, the present study is an attempt to find out the genetic similarity and phylogenetic position of north-west Punjabi population with respect to past history of admixture of foreign populations, especially, Caucasoid Populations. In this study, six microsatellite markers: THO1, TPOX, CSF1PO, vWA, D7S820 and FGA have been analyzed among 516 samples from five endogamous population groups, Jat Sikhs, Mazhbi Sikhs, Brahmins, Ramdasias and Muslims of north-west border districts of Punjab. The number of alleles ranged from 8 to 12 at six STR loci. The exact test probabilities for HWE suggested some significant departures in certain loci and population groups. In general, the average observed heterozygosity was lower than expected heterozygosity in six STR markers among the five population groups. The average sub-ethnic genomic differentiation (Fst) among five population groups of northwest Punjab was 0.0335. The CSF1PO showed highest sub-ethnic differentiation (0.0649), whereas, the lowest Fst has been observed for FGA locus (0.013). The percentage of genomic diversity attributable to different populations relative to the total genomic diversity (Gst) varied between 6.1% for CSF1PO locus and 0.9% for D7S820. When all the loci were jointly considered, 3.0% of the total genomic diversity was attributable to the five population groups. The maximum gene flow was observed in FGA (Nm=18.77%), which was followed by THO1 (15.92%). The lowest amount of gene flow was observed in CSF1PO (3.6%). However, in general with all loci the gene flow was observed to be 7.2% among these studied population groups. To understand the extent of sub-structuring among five northwest Punjabi population. Structure analysis was also performed with different values of K. The log probability values and the membership proportion of each group showed clear sub-structure among the population groups. Overall, five Punjabi speaking population groups of northwest Punjab are regionally well differentiated and exhibit strong genetic affinity based on their origin, settlement and their shared ethno-historic background.

2007S

Demographic inference of human population by diffusion model with nonparametric regression. *Y. Kawai, Y. Sato, Y. Yamaguchi-Kabata, N. Nariai, S. Sugimoto, T. Mimori, K. Kojima, M. Nagasaki.* Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan.

The demographic inference of human population is important not only for archeological and evolutionary studies but also for guiding sampling design for medical genetic study. Recent increase in the amount of human genetic data combined with the population genetic method contributed to the understandings of the demographic history of modern human. Both theoretical and experimental studies have shown that some populations have experienced the rapid population growth resulting in more low-frequency alleles than expected under the neutral and constant population size model. Thus, large samples, which include substantial amount of low-frequency alleles, are necessary for the demographic inference of rapidly growing population. The coalescence-based inference methods are commonly used for demographic inference with such large samples. With the aid of progress in computational power, a forward-time simulation under Wright-Fisher diffusion model is also applicable for observed allele frequency spectrum. We developed a method for demographic inference based on a diffusion approach, in which variable population size is estimated by nonparametric regression. By applying this method to the complete sequences of mitochondrial DNA from Japanese and Basque populations, we observed the recent population growth in both populations but with different growth rates and different time periods of growth.

2008M

Decoding ancient Bulgarian DNA with semiconductor-based sequencing. *Y. Ku[1], M.L. Carpenter[2], M. Sikora[2], H. Schroeder[3], C.C. Lee[1], C. Davies[1], M.T.P. Gilbert[3], C.D. Bustamante[2], G.D. Meredith[1].* 1) Thermo Fisher Scientific, South San Francisco, CA. USA; 2) Stanford University, Stanford, CA, USA; 3) Centre for GeoGenetics, Copenhagen, Denmark.

With the development of Next Generation Sequencing Technology (NGS), the field of hominin paleogenetics has transformed significantly from studying specific DNA markers to revealing whole genome information. However, ancient DNA of interest is usually highly fragmented so an NGS library protocol optimized to capture short DNA fragments (40bp to 200bp) was developed. The improved workflow includes the use of column-based DNA purification and concentration and automated gel-based size-selection. This workflow permitted production of "shotgun" genomic libraries from very limited input DNA (6ng to 39ng). Methods that permit the use of such low input, degraded DNA enable the partitioning of exceedingly rare samples into multiple analytical workflows. For example, to establish highest confidence SNP calls from ancient genomes it is best to sequence the sample on multiple orthogonal platforms. To pilot this approach, DNA extracted from 4 human tooth specimens from Bulgaria (ancient Thracians) that date to the Bronze and Iron Ages (1500-400 yr BCE) were sequenced on a semiconductor-based platform. Per individual sample, 259 million to 312 million sequence reads were produced. Deep sequencing (467 Mb) on one of the samples (P192-1) yielded detection of ~400,000 SNPs. Using principal component analysis that included more than 1,300 modern Europeans, this large number of SNPs indicated clustering of P192-1 (from an ancient farming community) closely with modern Sardinians; this result resolves ambiguity in the ancestral lineage of this individual that was recently reported based on ~10-fold fewer SNP calls and supports the hypothesis that the ancient Thracians and Sardinians share a heritage that dates to the initial spread of farming into Europe about 7,000 years ago. Data from two orthogonal sequencing platforms for these ancient Bulgarian samples demonstrated very similar base-substitution profiles with C>T and G>A variants accounting for ~75-80% of all SNPs called in both datasets. With such orthogonal validation, we expect to be able to reduce the false positive rate and generate a "truth" list of SNPs that will enhance our understanding of ancient population genomics and migrations. In summary, we have demonstrated a library preparation and semiconductor-based NGS workflow that is applicable for processing contaminated and degraded samples and can be used for ancient DNA research. For Research Use Only. Not for use in diagnostic procedures.

2009S

Exploring Detailed Demographic Histories of Human Populations Using SNP Frequency Spectrums. *X. Liu, Y.-X. Fu.* Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX.

Inferring human demographic history using genetic information can shed light on important prehistoric evolutionary events such as population bottleneck, expansion, migration, and admixture, among others. It is also the foundation of many population genetics analyses, as demographic history is one of the most important forces shaping the polymorphic pattern of DNA sequences. We developed a novel model-free method called stairway plot, which infers detailed population size changes over time using SNP frequency spectrums. This method can be applied to low-coverage sequence data, pooled sequence data and even reference-free sequence data for species whose reference genome are not yet available. Another advantage of this method is the ability to handle whole-genome sequences of hundreds of individuals. Using extensive simulation we compared our method to Li and Durbin's method based on the pairwise sequentially Markovian coalescent (PSMC) framework and the results show that our method outperformed the PSMC method for inferring recent population size changes. We applied our method to the genomes of nine non-admixed populations (CEU, GBR, TSI, FIN, CHB, CHS, JPT, YRI and LWK) from the 1000 Genomes Project, and showed a detailed pattern of human population fluctuations from 10 to 500 thousand years ago (kya). The results supported many mainstream viewpoints on the demographic histories of human populations, and at the meantime also produced several interesting observations worth further and more careful investigations.

2010M

Exome sequencing of 3,000 individuals reveals differences in recent demographic history between East Asian and European populations. *K.E. Lohmueller[1], M. He[2,3], Y. Li[3], B. Kim[1], L. Sun[4], X. Zhang[4], X. Jin[3], K. Kristiansen[3,5], T. Hansen[6,7], J. Wang[3], O. Pedersen[7,8,9], E. Huerta-Sanchez[10], R. Nielsen[5,10].* 1) Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA; 2) Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA; 3) BGI-Shenzhen, Shenzhen, China; 4) Department of Dermatology, First Affiliated Hospital, Anhui Medical University, Hefei, China; 5) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 6) Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark; 7) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark; 8) Faculty of Health Sciences, Aarhus University, Aarhus, Denmark; 9) Institute of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark Denmark; 10) Integrative Biology, University of California, Berkeley, Berkeley, CA.

Studies of genetic variation in thousands of individuals have found evidence for extreme population growth within the last 10,000 years in European and African American populations. The magnitude of recent growth in other continental populations, such as East Asians, has received comparatively little attention. In order to learn more about recent population history in East Asia, here we analyze high-coverage exome sequencing data from 1,449 Han Chinese individuals sampled from the Anhui province of China and 1,449 Danish individuals. We estimated recent demographic history using the site frequency spectrum. We find that the current effective size of the Han is approximately 4-fold larger than that estimated in the Danish population. Thus, while previous studies of common variants suggest historically smaller effective sizes in East Asian populations relative to European populations, our estimates of recent effective population sizes show the opposite pattern and trend in the same direction as the census population sizes. Next, we characterize the relationship between our estimates of the current effective population sizes and the census sizes. The ratio of the census size (over the last 200 years) to the recent effective size is significantly higher in the Han population than in the Danish population ($P<2\times10^{-4}$). This difference can be explained by greater variance in reproductive success in the Han population as compared to the Danish population. Alternatively, this result could be due to greater migration into the Danish population than the Han population. While it is appreciated that effective sizes of human populations are smaller than the census sizes, here we demonstrate that the magnitude of this difference varies across populations, even after accounting for population size changes. Finally, we examine patterns of deleterious variants in the Han and Danish populations. We find that the proportion of private variants that are nonsynonymous is higher in the Han sample (67.6%) than in the Danish sample (64.6%; $P<10^{-10}$), consistent with recent population growth increasing the input of weakly deleterious mutations into the population that selection has not had sufficient time to remove. Our study provides the first analysis of recent population history and exploration of neutral and deleterious rare variants in an East Asian population.

2011S

Analysis of Genetic Diversity Representation of the 1000 Genomes in Worldwide Human Populations. *D. Lu, S. Xu.* Partner institute for Computational Biology, Shanghai, Shanghai, China.

The 1000 Genomes Project (1KG) aims to provide a deep characterization of human genome sequence variation, by design was expected to aims to provide a comprehensive resource on human genetic variation. With an effort of sequencing 2,500 individuals, 1KG is expected to cover the majority of the human genetic diversities worldwide. However, it would be interesting to evaluate to what extent the 1KG data represent the genetic diversity of human populations in each region, which will give insight into the power of 1KG and also give guidance to regional efforts for further sequencing project and study design. In this study, using analysis of population structure based on genome-wide single nucleotide polymorphisms (SNPs) data, we examined and evaluated the coverage of genetic diversity of 1KG samples with the available genome-wide data from 3,831 individuals representing 140 worldwide population samples. We demonstrated that the 1KG does not have sufficient coverage of human genetic diversity in Asia, especially in Southeast Asia. We thus suggest a better coverage of Southeast Asian populations be considered in 1KG or a regional effort be initialized to provide a more comprehensive characterization of the human genetic diversity in Asia, which is important for both evolutionary and medical studies in the future.

2012M

Visualizing the Geographic Distribution of Genetic Variants. *J.H. Marcus, J. Novembre.* Department of Human Genetics, University of Chicago, Chicago, IL.

One of the core features of any genetic variant, beyond its potential phenotypic effects or its frequency, is its geographic distribution. The geographic distribution of a genetic variant can shed light on where the variant first arose, in what populations it survived and spread within, and in turn help us learn about historical patterns of migration and natural selection. Collectively the geographic distribution of genetic variants can help to explain how populations have been related through time (e.g. levels of gene flow and divergence). For variants with large effects, it can also help us understand the geographic distribution of spatially-varying phenotypes. For these reasons, visual inspection of geographic maps for genetic variants is common practice in genetic studies. Here we develop a series of reusable interactive visualizations for illuminating the geographic distribution of genetic variants. We specifically address several non-trivial challenges of this type of visualization; in particular, how to represent non-uniform levels of uncertainty in allele frequencies due to variable sample sizes; how to represent results from data with >10,000 individuals in which allele frequencies can vary over 4 orders of magnitude; how to display data for regions of the globe with dense sampling of populations; and how to quickly access frequency data from large samples. To meet these challenges, we implement a flexible REST API for allowing for easy access to allele frequency and sample size data from large scale public genomic datasets. Built upon this API we develop a web-based browser, entitled the Geography of Genetic Variants (GGV) browser for visualizing the geographic distribution of genetic variants. The GGV browser rapidly provides maps of derived allele frequencies in populations distributed across the globe. The GGV browser builds upon past tools such as the HGDP Selection browser by allowing for more interactive features, new representations of rare variation, as well as incorporating uncertainty in allele frequency estimation. As ancillaries, we also develop a research visualization toolkit that includes a method for displaying high Fst outlier SNPs from the joint site frequency spectrum and an interactive version of commonly used PCA figures. We hope the GGV browser will be a valuable research and education tool for exploring population genetics data.

2013S

A Novel Likelihood Ratio Test Framework for Sex-Biased Demography and its Application to Human and Dog Genomic Data. *S. Musharoff[1], S. Shringarpure[1], T. Cooke[1], A. Adams[1], O. Cornejo[2], C.D. Bustamante[1], S. Ramachandran[3].* 1) Genetics, Stanford Univ Sch Medicine, Stanford, CA; 2) School of Biological Science, Washington State University, Pullman WA; 3) Ecology and Evolutionary Biology, Brown University, Providence, RI.

Sex-bias is defined as an unequal number of breeding males and females in a population. This can be caused by variance in reproductive success, demographic events involving unequal numbers of males and females, and/or differential selection at sex-linked genomic loci. A commonly used estimator of the proportion of females is based on the test statistic Q where Q is the ratio of current neutral genetic diversity estimated from the X chromosome to that estimated from the autosomes. This is problematic if the population changed in size: because X chromosomal diversity recovers from size changes at a different rate than autosomal diversity due to unequal effective population sizes, this estimator of the proportion of females based on current diversity will be biased. To this end we present a novel likelihood ratio test framework for sex-bias in a single population based on the Poisson random field model. We use the program dadi to estimate demographic parameters jointly from autosomal and X chromosomal data and test first for a persistent sex¬ bias and then for a sex-biased demographic event. When applied to simulated data, our test has more power to detect sex-bias from unlinked or partially linked sites than the commonly used test statistic Q for a range of demographic scenarios. Encouragingly, our test is well powered for events relevant to human history including recent rapid expansion whereas the test statistic Q is not. We recover the true proportion of females with our framework whereas the estimator based on Q underestimates the magnitude of sex-bias. To test for sex-bias in humans we applied our methodology to genome-wide data from three Thousand Genomes populations and find evidence for female sex-bias in Europeans, Asians, and Africans after modeling recent demographic events including a bottleneck and recent growth. Our method takes autosomal and X chromosomal site frequency spectra as input and can be applied to organisms with sex chromosomes. To investigate sex-bias in dogs, we analyzed a novel genome-by-sequencing (GBS) dataset comprised of genome-wide putatively neutral regions. After applying our methodology to breed dogs and to village dogs, we find that both groups show evidence of female bias with the breed dogs showing a more extreme bias than village dogs. These findings argue for the importance of modeling demography when assessing sex-bias in populations and highlight the role of sex-bias in the history of humans and of dogs.

2014M

Inferring the effective founder size of a spatially expanding population. *B.M. Peter, M. Slatkin.* University of California, Berkeley, Berkeley, CA.

The gradual loss of diversity associated with range expansions is a well known pattern that can be explained with a serial founder model. Using a branching process model, we show that this loss in diversity can be mainly attributed to the difference in offspring variance between individuals at and away from the expansion front. We use this model to define an effective founder size as a measure of the decrease of genetic diversity with distance. We show that many classical properties of the effective population size, such as the two-sex formula and the harmonic mean for time-varying models extend naturally to expanding populations. We demonstrate that the predictions from the branching process model fit very well with Wright-Fisher forward simulations and backwards simulations under a modified Kingman coalescent, and further show that estimates of the effective founder size are robust to possibly confounding factors such as migration between subpopulations.

2015S

Finding the oasis of humanity in Neanderthal deserts. *B. Vernot, JM. Akey.* Department of Genome Sciences, University of Washington, Seattle, WA.

As anatomically modern humans dispersed out of Africa, they encountered Neanderthals in Eurasia and low levels of hybridization occurred such that approximately 2% of each non-African's genome is inherited from Neanderthal ancestors. Recently, we developed an approach to identify surviving Neanderthal lineages in contemporary individuals, and recovered over 600 Mb of the Neanderthal genome present in modern non-African populations [1]. The map of surviving Neanderthal sequences shows marked heterogeneity across the genome, and we identified many "deserts of Neanderthal sequence" that are almost entirely devoid of Neanderthal sequence. These genomic regions are of particular interest because they delimit sequences that may confer uniquely human characteristics. For example, the largest Neanderthal desert is a 15Mb region on Chromosome 7, centered around the FOXP2 gene, which has previously been implicated in speech and language. Here, we present a detailed characterization of Neanderthal deserts by analyzing surviving archaic sequences in an expanded sample of geographically diverse individuals. We have developed a formal statistical test to identify genomic regions significantly depleted of Neanderthal lineages, and performed extensive simulations to infer the strength of purifying selection acting on these Neanderthal deserts. Additionally, we have utilized extensive bioinformatics analyses superimposing heterogenous functional genomics data to identify candidate causal variants. These analyses provide significant new insights into regions of the human genome that harbor sequences that have played a critical role in the evolution of anatomically modern humans, and suggest that regulatory sequences responsible for muscle, bone, and brain development were key differences between humans and Neanderthals. [1] Vernot and Akey, Science, 2014.

2016M

Coalescent times of 63 males estimated using the complete Y-chromosomes. *E. WONG[1], S. Limborska[2], A. Valouev[1].* 1) Department of Preventive Medicine, Keck School of Medicine of University of Southern California, Los Angeles, CA; 2) Department of Human Molecular Genetics, Russian Academy of Sciences, Moscow, Russia.

The Y-chromosome contains the longest non-recombining region in the genome and is useful for inferring human population history. Previous studies on the Y-chromosome mostly rely on rapidly mutating microsatellites or selected single nucleotide polymorphisms ascertained from small panels of individuals. These approaches likely underestimate the genetic diversity of the Y-chromosomes. Complete sequencing of the Y-chromosome to high coverage would enable us to more accurately estimate the coalescent times of populations across the world and reconstruct past events in human evolution. We sequenced the complete Y-chromosomes of 12 males sampled from Asia and Europe. We analyzed these chromosomes together with publicly available high-coverage complete Y-chromosomes of 51 males from across the world with known ethnicities. In our analysis, we excluded regions of the Y-chromosome that are subject to recombination and regions that are ill-suited for short-read sequencing due to low complexity and homology to the X-chromosome. By focusing our analysis on 10.45 million sites, we identified over 15,000 single nucleotide variants (SNVs) in 63 males from 38 populations across the world and constructed a maximum likelihood tree. We performed haplogroup analysis for these Y-chromosomes, based on the SNVs present in the International Society of Genetic Genealogy (ISOGG) database. We estimated the Y-chromosome coalescent time to be around 129 thousand years, which is in line with previous estimates. By estimating the coalescent time of these Y-chromosomes, we are able to understand more about the history of these populations and the migrations of their ancestors.

2017S

Detecting and dissecting the fine-scale genetic population structure of Spain. *C. Bycroft[1], A. Carracedo[2], C. Fernandez-Rozadilla[1,2], C. Ruiz-Ponte[2], I. Quintela-García[2], S. Myers[1], P. Donnelly[1].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 2) Galician Public Fundation of Genomic Medicine (FPGMX)-Grupo de Medicina Xenómica-Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERer)-IDIS, Santiago de Compostela, Spain.

The Iberian Peninsula has been the subject of diverse population movements, both from Europe and North Africa, over millennia. The patterns of genetic variation within Spain reflect these movements, together with local isolation of populations due to cultural, linguistic, and geographical factors. Here we describe the largest genome-wide study of population structure across Spain, involving over 500,000 markers in ~600 individuals, analysed with powerful, recently developed statistical tools which exploit patterns of linkage disequilibrium. These analyses reveal striking patterns of genetic population structure at very fine geographic scales and shed light on the demographic history of populations within Spain. The genetic differences we identify distinguish individuals from many parts of Spain, including Galicia, the Basque Country, Catalonia, and the Balearic Islands. These differences correlate with both physical barriers to migration, and different languages spoken within Spain. Interestingly, we detect some genetic clusters that span from the north to the south of Spain, while groups differ more strongly east-west, suggesting predominantly north-south movements of people within Spain. Differing amounts of ancestry shared with individuals from Basque-speaking regions in populations from different parts of Spain also suggests a historic population movement out of the Basque region, largely in a southerly direction. Finally, we use a mixture model-based approach and a recently developed method, GLOBETROTTER, to quantify and date the genetic impact of migrations into Spain from groups including the Moors from North Africa and other European populations.

2018M

Using constraints on F_{ST} to interpret genetic differentiation in genomic data. *M.D. Edge[1], M. Jakobsson[2], N.A. Rosenberg[1].* 1) Department of Biology, Stanford University, Stanford, CA; 2) Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

The availability of genomic polymorphism data from large, worldwide samples has made it possible to study population structure at loci with rare alleles and low genetic diversity. Computation of F_{ST} is often a first step in population-genomic analyses of data from multiple populations. F_{ST} is a theoretically rich measurement of genetic differentiation, interpretable as an index of excess homozygosity due to population structure, as an excess coalescence time for alleles drawn from different subpopulations, and as the proportion of variance in allelic indicator variables attributable to population structure. At the same time, F_{ST} has sometimes behaved in surprising ways in genomic studies of human populations. For example, subpopulations within Africa often have low estimated F_{ST}, despite having considerable population structure as detected via individual-based clustering. Rare variants can generate low F_{ST} despite being more geographically localized than common variants. We suggest that one way to understand these surprising phenomena is to consider mathematical bounds on F_{ST} as a function of other population-genetic statistics. We have previously determined strict upper bounds on F_{ST} as a function of homozygosity and as a function of the frequency of the most frequent allele for loci with arbitrarily many alleles. Here, we extend these bounds to accommodate loci with specified numbers of alleles, showing that F_{ST} is highly constrained both for low and high levels of diversity and can only reach values near the theoretical maximum of 1 for loci with intermediate diversity levels. These bounds provide an explanation for the surprising behaviors of F_{ST}: for example, the low values for African subpopulations can be explained in terms of the particularly high genetic diversity in Africa. As genomic studies continue to examine rarer and more localized alleles - which generate loci with low genetic diversity -- our bounds can facilitate the sensible interpretation of F_{ST} results.

2019S

Population structure of Amerindian 19 populations in South America. *S. Nishikawa[1], R. Saito[1], N. Konno[1], G. Tamiya[1], N. Fuse[1], Y. Nakamura[2], J. Yasuda[1], I. Danjoh[1].* 1) Department of Integrative Genomics , Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 2) Cell Engineering Division, BioResource Center, RIKEN Tsukuba Institute, Tsukuba, Ibaraki, Japan.

<Introduction> Ancestors of Amerindians migrated from Eurasia to North American continent via Bering Strait, and then reached South American continent over 10 thousand years ago. The study of HTLV-1 in South American populations indicates that they can be divided into two major ethnic groups. However, detailed population structure and genetic diversity in Amerindians are still unclear. To address these questions, we obtained genome-wide SNP data from the Sonoda-Tajima Cell Collection, that is a vast collection of B lymphoblastoid cell lines established from 550 native Americans belong to 29 tribes in South America (Danjoh et al., 2011), and analyzed population structure within and among tribes, and then analyzed the genetic structure of Amerindian populations.

<Methods> The standard SNP and samples quality control (QC) procedures including the estimation of "cryptic relatedness" were applied to the genome-wide SNP data on Genome Studio software and PLINK software, and removed closely related samples. After this step, 10 tribes retained insufficient number of samples, therefore we did not process them for further analysis. Remaining 19 tribes were used for the following analysis. The between-population Fst estimation, neighbor-joining (NJ) tree construction, and PCA were carried out with Arlequin, MEGA, and EIGENSTRAT software, respectively. The number of clusters (K) in Amerindian populations was estimated with ADMIXTURE.

<Results and Conclusion> NJ-tree showed that Samuna and Chipaya are distant from other populations. Moreover, PCA and cluster analysis showed that Sanuma formed a distatnt cluster along PC1 and PC2, on the other hand, Chipaya was clustered with the other populations. The result of cluster analysis showed that our 19 populations of Amerindian was constituted of 7 clusters. This is the first report for the detailed population structure of these Amerindians including Samuna.

2020M

Distribution of *CYP2C9*2* and *CYP2C9*3* variants in two Native populations of Mexico. *L.C. León-Moreno, J. Sánchez-Corona, A.M. Saldaña-Cruz, A.G. García-Zapién, M.C. Morán-Moguel, S.E. Flores-Martínez.* Molecular Medicine, Centro de Investigación Biomédica de Occidente, IMSS., Guadalajara, Jalisco, Mexico.

Presence of polymorphisms in the CYP450 enzymes genes is one of the principal determinants for individual variability in drug response. Particularly, CYP2C9 enzyme metabolizes almost 20% of the existing drugs. *CYP2C9* gene exhibits more than 57 variants, of which *CYP2C9*2* (rs1799853) and *CYP2C9*3* (rs1057910) have been associated with a reduced catalytic activity compared with wild type (*CYP2C9*1*). Variation in frequencies has been described for these two variants in Mexico, mainly between mestizos and native groups. This variation might be the cause, at least in part, of the disparities in drug disposition between ethnic groups. The present study is aimed to analyzing the *CYP2C9*2* and *CYP2C9*3* variants in two Native populations of Mexico to increase the information about the polymorphisms of pharmacogenomic relevance in the Mexican population. **Methods.** DNA samples from two Native Mexican populations (Lacandonians n=223 and Yaquis n=103) were genotyped for two variants in the *CYP2C9* gene (*CYP2C9*2* and *CYP2C9*3*). The variants were identified using PCR-RFLPs. Genotypic and allele frequencies and the Hardy-Weinberg equilibrium were performed using a package of statistical analysis for genetic marker data, ARLEQUIN v3.5.1.2. **Results.** In Lacandonian population, both heterozygous genotypes to *CYP2C9*2* and *CYP2C9*3* variants were found in a frequency of 0.4%. In Yaquis population, the heterozygous genotype to *CYP2C9*2* variant was found with a frequency of 3.9%, while heterozygous genotype to *CYP2C9*3* variant with a frequency of 14.6%. The compound heterozygous genotype *CYP2C9*2/*3* was observed only in Yaquis population. The *CYP2C9*2* or *CYP2C9*3* variants in homozygote state were no detected in none population. No significant deviations from HWE expectations were observed in any of the two studied populations. **Conclusions.** Variation in allele and genotype frequencies is evident between northern and southern Mexican native population, reflecting the differences among Mexican populations. The presence of these variants could represent a risk factor of drug resistance or adverse reactions to those drugs metabolized by CYP2C9 enzyme in individuals of both populations.

2021S

Full ancient mitochondrial genomes using the Ion Proton platform: Analyzing the genetic diversity of the Neo-Eskimo. *J. Tackney[1], A.M. Jensen[2], S. Watkins[3], J. Brenner-Coltrain[1], D.H. O'Rourke[1].* 1) Department of Anthropology, University of Utah, Salt Lake City, UT; 2) UIC Science LLC, Barrow, AK; 3) Department of Human Genetics, University of Utah, Salt Lake City, UT.

A cultural shift occurred in the North American Arctic archaeological record beginning around 1000 AD. This culture, known as the Neo-Eskimo Thule, quickly occupied the region from Alaska to Greenland, and their descendants are modern Iñupiat/Inuit. Nuvuk is a long-term Thule village at Pt. Barrow, AK. Its cemetery has yielded archaeological material with radiocarbon dates between 980-1300 AD, suggesting that human remains from the site should span from the classic Thule period to modern Iñupiat Eskimo.

Mitochondrial haplogroup D is the second most common haplogroup in northern Asia and it is present at high frequency in central and eastern Asia, as well as in the Americas. One subclade of D4 seems to have originated in southern Siberia and spread northward into northern Asia and the Americas post-Last Glacial Maximum. This subclade is present in the Inuit/Iñupiat populations of North America at low frequency (~5%), is shared with some Siberian Chukotkans, and is absent in more southern Amerindian populations. To help further characterize the sequence variation present in this haplotype, we assessed the mitochondrial genomes of two samples previously typed for HVS1 SNPs, and assigned to D3/D4b1a2a1a. The bone samples were directly dated with median intercept calibrated dates of 1269 and 1313 AD.

We optimized the Ion Torrent sequencing library creation and the Torrent Suite bioinformatics pipeline for ancient DNA analysis by limiting reaction clean-up steps, determining PCR cycle numbers using qPCR, and expanding the mapping algorithms performed on the unaligned reads. Using mtDNA hybridization probe capture, we generated complete mtDNA for each sample on the Proton sequencing platform. The full genomes allowed us to place these samples with high phylogenetic confidence in the mtDNA tree. Sequences suggest population continuity with modern Iñupiat/Inuit and the D3/D4b1a2a1a haplotypes confirm that the Thule belong in the Beringian, and not Asian or Pan-American, gene pools.

This research was supported by NSF grants OPP-0732846 AND OPP-0637246 to DHO'R, OPP-0820790 to JBC, and ARC-0726253 to AMJ.

2022M

Large scale whole-genome sequencing of the Icelandic population isolate. *D.F. Gudbjartsson[1,2], H. Helgason[1,2], S.A. Gudjonsson[1], F. Zink[1], A. Oddsson[1], A. Gylfason[1], G. Magnusson[1], B. Halldorsson[1,3], S. Besenbacher[4], A. Kong[1], G. Masson[1], U. Thortsteinsdottir[1,5], A. Helgason[1,6], P. Sulem[1], K. Stefansson[1,5].* 1) Dept Statistics, deCODE Genetics/Amgen, Inc, 101 Reykjavik, Iceland; 2) School of Engineering and Natural Sciences, University of Iceland, 101 Reykjavik, Iceland,; 3) Institute of Biomedical and Neural Engineering, Reykjavík University, 101 Reykjavík, Iceland,; 4) Bioinformatics Research Centre, Aarhus University, C.F. Mollers Alle, 8000 Aarhus C, Denmark; 5) Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland,; 6) Department of Anthropology, University of Iceland, 101 Reykjavik, Iceland.

Here we describe the insights gained from sequencing the whole genomes of 2,636 Icelanders to a median depth of 20X. Twenty million single nucleotide polymorphisms (SNPs) and 1.5 million insertions/deletions (indels) passed stringent quality control. We report the number of sequence variants by minor allele frequency (MAF) and functional annotation. Almost all SNPs with derived allele frequency (DAF) over 2% in Iceland have been recorded in international databases of sequence variants, while for SNPs with a DAF of 0.1%, half of those in protein coding regions and 20% of those in non-protein coding regions have been recorded outside Iceland. The corresponding fractions of Icelandic indels that have been recorded elsewhere are substantially lower. Sequence variants in genes involved in the function of the nucleus have the lowest density and the highest fraction of rare variants (DAF < 0.5%, FRV), while variants in genes involved in sensory perception have the highest density and the lowest FRV. In particular, variants in olfactory receptors have a higher density and are more common than the genome average. Loss of function (LoF) and missense variants have a higher density and are more common in the first and last exons of genes than they are in the middle exons. Missense and LoF variants in genes that have been associated with disease in the Online Mendelian Inheritance in Man, catalog through a dominant mode of inheritance, are less dense and rarer than variants in genes that have not been associated with disease. Variants with a chromatin state annotation have a density and FRV that fall between intergenic and untranslated regions. Sequence conservation between mammals, measured by the GERP score, predicts variant density and frequency. However, after accounting for the GERP score, sequence annotation has a substantial effect on variant frequency but not on variant density. Consistent with a founder effect and the small size of the Icelandic population, there are 56% more stop gained SNPs in the frequency range between 0.1% and 1% in Iceland than in the European American part of the Exome Sequence Project data. Icelanders are more likely to be homozygotes than Hardy-Weinberg equilibrium would predict by a factor of 3 for variants with a MAF of 0.1%. We have described the sequence diversity and structure of the Icelandic population isolate and evaluated several types of sequence annotations based on the strength of selection signature.

2023S

Identifying Clusters of Rare Skeletal Genetic Disoders in Brazil - preliminary data. *D. Cavalcanti, C. Moreno.* Skeletal Dysplasia Group. Medical Genetics Department, FCM, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil.

The identification of clusters of rare genetic diseases has significant medical and scientific implications. The establishment of an investigation group of skeletal genetic disorders [osteochondrodysplasias (OCD) and skeletal malformations or dysostosis] at our institution has contributed to the creation of a register of these conditions in which familiar and demographic data in addition to the clinical ones have been sistematically collected. This presentation aims to show the preliminary results related to the identification of clusters of OCD and dysostoses in Brazil. The data base was revised with special attention to the clinical and demographic data. The more common registered cases as FGFR3 group, osteogenesis imperfecta as well as campomelic dysplasia, type 2 collagenopathies, and inconclusive diagnosis were all excluded of the present analysis. Thus, among 359 cases 52 different diagnoses were established. The autosomal recessive (AR) inheritance pattern was observed in 62% (32/52) of the cases and parental consanguinity was present in 28% (30/107). The molecular investigation found a pathogenic mutation confirming the diagnosis in 25 cases of 21 families. Among 11 cases for whom parental consanguinity was referred as negative a homozygous mutation was found in seven patients and for five ones both paternal and maternal ancestors came from small and identical/close cities, suggesting a common origin with an unknown and distant consanguinity. The following were considered evidence of a cluster: the presence of a rare mutation in different families associated or not with families from the same region, and cluster evidence already known from the literature. So far, five new clusters were found, and they are distributed in different states: Blomstrand dysplasia (Alagoas), Pycnodysostosis (Ceará), Short-rib polydactyly dysplasia type 3 (Maranhão and Pernambuco), Opsismodysplasia (Pernambuco), Spondyloenchondrodysplasia (Minas Gerais). All these conditions are associated with parental consanguinity. We also found a case of Grebe dysplasia (Bahia), and a case of Richeiri-Costa-Pereira syndrome (São Paulo), reinforcing clusters already known from the literature. In conclusion, we found five new clusters of AR skeletal conditions in Brazil, for whom the associated inbreeding suggests founder effect. [Supported by grants from CNPq # 402008/2010-3, 590148/2011-7].

2024M

The CARTaGENE Genomics Project: How successive bottlenecks have shaped the present French Canadian founder population. *H. Gauvin[1,2], JP. Goulet[1], JC. Grenier[1], M. Capredon[1], T. de Malliard[1], V. Bruat[1], E. Gbeha[1], E. Hip-Ki[1], A. Hodgkinson[1], P. Awadalla[1,3].* 1) Sainte-Justine Hospital Research Center, Université de Montréal, Montreal, QC, Canada; 2) Department of Social and Preventive Medicine, Université de Montréal, Montreal, QC, Canada; 3) Department of Pediatrics, Université de Montréal, Montreal, QC, Canada.

French Canadians (FC) are a population of *Nouvelle-France*'s settlers who established themselves before the British conquest. The population has undergone recent rapid growth, with limited genetic exchange with other populations, and has colonized the whole province in successive waves leading to regional founder effects contributing to population structure. It is known that levels of kinship and disease prevalence vary from region to region within Quebec, making the population a particularly interesting case-study for understanding the origins of rare genetic diseases; in this study we attempt to disentangle and describe the processes driving these patterns. Using participants from the CARTaGENE project, the provincial biobank, we generated whole genome genotyping data for ~1000 individuals sampled in three regions (Montreal, Quebec City and Saguenay) with the OMNI2.5 chip, and then examined the fine structure of the Quebec population and searched for signals of genetic drift and selection using a haplotype-based approach. Focussing on individuals with FC descent, which was inferred from principal component analysis and information provided by participants, we performed a high-resolution analysis of the genetic structure of the population using the Chromopainter and FineSTRUCTURE clustering algorithm. FineSTRUCTURE analysis showed differences between two geographically close regional populations, Montreal and Quebec City, and provided greater resolution of the migration patterns within the province. Furthermore, using HaploPS, we identified genomic signatures of positive selection and genomic regions that have potentially drifted to particularly high frequency among FC. Among these regions we located one, at a 50% frequency among FC from Saguenay, encompassing the RYR1 gene found to be implicated in malignant hyperthermia. This is a pharmacogenetic disorder for which cases were previously identified in families coming from Abitibi-Temiscamingue, another region of Quebec. It was reported that common ancestors of those cases were found before the colonization of this recently inhabited region meaning that mutations causing the disease could have spread in others regions of Quebec. Collectively, these results provide a refinement of previous analyses of the structure of the Quebec population and suggest that the unique demographic history of FC may have affected specific regions of the genome that are relevant to the population's health.

2025S

Dominant mutation in GJB2 causes hearing loss in northeast mexican family. *F. Loeza-Becerra[1,2,3], R. Rivera-Vega[2,3], E. Gutierrez-Contreras[4], P. Berruecos-Villalobos[2,3], SA. Cuevas-Covarrubias[2,3].* 1) UMSNH Morelia, Michoacán, Mexico; 2) Secretaria de Salud; 3) UNAM, Mexico DF; 4) Centro de Rehabilitacion Gpe. H. de De las Fuentes DIF-Torreon Coah.

To know mutations in GJB2 and GJB6 that causes hearing loss in mexican population, we make genotyping and audiometric study in n=95 families (trios and more) from the Mexican Republic to the north, north-east, north-west and west with at least one case non-syndromic sensorineural hearing loss (NSSNHL). It was 70% wild type (mutations in no-GJB2 nor GB6 or mADN), eighteen different mutations and this: c.36GA (p.G12R). Most of the cases are sporadic or reccesives, and only one north-eastern family has the mutation c.36GA (p.G12R) in nine members, three generations, both sexes affected and transmission male-to-male. All affected has neurosensorial hearing loss. Now corroborated no anatomical anomalies. We remarks that its importance is clinical and epidemiological because the hearing loss by GJB2 is classically reccesive and it is northern that mutation.

2026M

Clinical Implications and Frequency of Variants in Nonsyndromic Hearing Impairment Genes in a Population-Based Sample of African-Americans and European-Americans. *H. Dai, G.T. Wang, R.L.P. Santos-Cortez, S.M. Leal.* Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX, 77030.

A large number of studies on nonsyndromic hearing impairment (NSHI) have been performed using samples collected from the Indian subcontinent, Middle East and Europe. There has been limited study of NSHI in African-Americans (AA) and sub-Saharan Africans. We evaluated the frequency of previously reported "pathogenic" variants in NSHI genes using data from the NHLBI-Exome Sequencing Project (ESP) which is a population-based study of AA (N=2203) and European-Americans [EA (N=4300)]. Of 201 observed variant sites which are pathogenic according to ClinVar and/or the Deafness Variation Database, we reclassified 121 (60.2%) variant sites as non-pathogenic based on literature, high allele frequencies in ESP and bioinformatics tools. In ESP, 80 pathogenic variant sites were observed in 5 autosomal dominant (AD) and 14 autosomal recessive (AR) NSHI genes, of which 24 variant sites in 7 genes cause syndromic hearing impairment (HI), e.g. *MYO7A* (MIM 276903) variants cause both NSHI and Usher syndrome. Of these variant sites, 49 (91 alleles) were found only in EA, 18 (23 alleles) only in AA, and 13 (194 alleles) in both EA and AA. *GJB2* (MIM 121011) c.35delG (p.G12Vfs) was the ARNSHI variant site with the highest allele frequency in EA [N=89; 1.09% (95%CI:0.88%,1.34%)] but with a much lower allele frequency in AA [N=4; 0.094% (95%CI:0.03%,0.2%)]. Additionally *OTOF* (MIM 603681) c.2348delG (p.G783Afs) was identified in 4 AA alleles [0.095% (95%CI:0.03%,0.2%)] but is very rare in EA [N=1; 0.01% (95%CI:0.0003%,0.07%)]. For EA, pathogenic variant sites were identified in 11 AR and 5 AD NSHI genes including variant sites in 7 genes which cause syndromic HI, while for AA pathogenic variant sites were identified in 13 AR and 2 AD NSHI genes, of which 6 genes have variant sites which cause syndromic HI. *MYO7A* c.3764delA (p.K1255Rfs) which was reported to cause Usher syndrome, is homozygous in 1 AA and 3 EA individuals. The use of EVS in order to exclude nonpathogenic variants must be done cautiously due to NSHI-causal variants within EVS. Knowledge of the frequency of NSHI variants in large population samples from different ethnic backgrounds is not only important to evaluate clinical significance but also aids in evaluating pathogenicity. Although population-based samples of AA aids in evaluating population-specific frequency of pathogenic variants, AA and sub-Saharan Africans with HI need to be studied to better understand the genes and variants underlying disease etiology.

2027S

Human population growth and purifying selection have increased the burden of autosomal and X-linked private mutations. *F. Gao, A. Keinan.* Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY.

Several recent projects have sequenced the exome or whole-genome of a large number of individuals. A common feature discovered in these data sets is the significant elevation of rare variants, which are the segregating sites that have their minor allele in only one or a few chromosomes across the entire sequenced sample. This phenomenon can be explained by the effect of recent human population growth and purifying selection. Here, we considered the amount of rare variants in the prism of the burden of private mutations. We defined the burden of private mutations as the proportion out of all heterozygous sites in an individual that are unique compared to the rest of the sample. This quantity also answers the following question: Considering a large number of sequenced individuals, how many new variants will be ascertained with each additional individual sequenced? We studied the burden of private mutations by calculating the prediction from different demographic models and comparing with empirical estimates based on the Neutral Regions (NR) dataset that we recently published and that from the NHLBI Exome Sequencing Project (ESP). By analyzing autosomal SNVs, we observed a significant excess in the proportion of private mutations in the empirical data compared with models of demographic history without a recent epoch of population growth. Incorporating recent growth into the model provides a much improved fit to empirical observations. This phenomenon becomes more marked for larger sample sizes. For example, after sequencing 10,000 individuals from the same population, still about 1 in 400 heterozygous sites (~6,000 variants) at the 10,001st individual are predicted to be novel, 18-times more than predicted in the absence of recent population growth. The burden of private mutations is additionally increased by purifying selection, generally with intergenic SNVs exhibiting the lowest burden and splice sites the highest. Finally, we contrasted the burden of private mutations between the autosomes and the X chromosome, across different functional annotations and across matched allele sample sizes ranging from $n = 1$ to $n = 3027$. The results of this comparison point to a combination of sex-biased demographic history and purifying selection differentially affecting X and the autosomes. The very large numbers of private mutations in human genomes suggest that careful consideration is needed in the design and analysis of sequencing-based association studies.

2028M

Whole Genome Sequence of Japanese from Tohoku Medical Megabank Prospective Cohort Study. *M. Nagasaki[1,3], F. Katsuoka[1], N. Nariai[1], K. Kojima[1], I. Danjo[1], Y. Kawai[1], S. Saito[1], X. Pan[1], J. Yokozawa[1], R. Saito[1], Y. Sato[1], T. Mimori[1], Y. Yamaguchi-Kabata[1], K. Tsuda[1], Y. Kuroki[1], K. Kinoshita[1,3], J. Yasuda[1], M. Yamamoto[1,2], Tohoku Medical Megabank Organization Tohoku University Project.* 1) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Sendai, 980-8575, Japan; 2) Department of Medical Biochemistry, Tohoku University Graduate School of Medicine, Sendai, 980-8575, Japan; 3) Department of Applied Information Sciences, Graduate School of Information Sciences, Tohoku University, 6-6-05 Aramaki Aza Aoba, Aoba-ku, Sendai, Miyagi 980-8579, Japan.

Tohoku University Tohoku Medical Megabank Organization (ToMMo) located at the north-eastern Japan is now developing a biobank (the final goal is 150,000 volunteers with millions samples) that combines medical and genome information during the process of rebuilding the community medical system and supporting health to foster the reconstruction from the Great East Japan Earthquake on Mar/11/2011. One of the missions of ToMMo is to reveal a fine genetic architecture of Japanese population to tackle the further GWAS analysis by combining the knowledge, which is daily accumulated in the ToMMo prospective genome cohort project, e.g. questionnaire data, physiological data, medical treatment records and other omics data from serum, plasma and immortalized lymphocytes. The first goal is to sequence 1K samples to cover MAF > 0.5% variants including short indels and large structural variants in Japanese for constructing a Japanese 1K reference panel. In our experimental design, to minimize the biases caused by the different equipment, protocol and bioinformatics analysis, we performed whole genome sequencing of 1K samples with 30X high coverage using the HiSeq 2500 rapid run mode and analyzed by the same bioinformatics pipeline. We have catalogued reliable SNPs with MAF > 0.5. In addition, variants categorized into rare frequency (MAF ≤ 0.5%) were detected with the same precision thanks to the variant call from high coverage sequencing without imputation. We also present other preliminary findings while constructing Japanese 1K reference panel.

2029S

Analyzing the impact of consanguinity and admixture on the Middle East Variome. *E.M. Scott[1], E. Spencer[1], B. Copeland[1], M. Abdellateff[1], S. Gabrielle[2], J. Gleeson[1].* 1) Neurogenetics Laboratory, Howard Hughes Medical Institute, Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093, USA; 2) The Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA.

The Middle East has been an important hub of human migration, population admixture, and contribution of genetic diversity. Although details are disputed, several large-scale human migrations into an out of Africa, along with many more recent invading populations could have contribute to a complex heterogeneous Middle Eastern genome. Additionally, cultural practices predisposing consanguineous marriage have resulted in some of the highest known inbreeding coefficients, which theoretically should depress burdens of deleterious variants. Despite the size and uniqueness of these populations, so far they have been severely underrepresented and understudied. Here we investigate the genetic landscape of the Middle Eastern genome by repurposing a cohort of 2000 exomes samples resulting from three separate recruitment efforts focusing collection in the Gulf region, North Africa and Central Asia, collectively termed Middle East. Analysis of this cohort provides evidence for unique admixture components, and confirmatory evidence for inter-mixing of Sub-Saharan African and European populations. High levels of consanguinity resulted in a markedly higher rate of inbreeding coefficients compared with 1000G controls along with longer runs of homozygosity. Finally, we investigated potential impact these trends on the load of deleterious variant, demonstrating surprisingly a higher than expected burden of deleterious variants compared with European ancestry controls.

2030M

Rare variant stratification in a German/Austrian sample set. *E. Tilch[1,2], B. Schormair[1,2], B. Pütz[3], D. Czamara[3], M. Müller-Nurasyid[4,5,6], P. Lichtner[2], C. Trenkwalder[7,8], W. Paulus[9], B. Högl[10], K. Berger[11], I. Fietze[12], W. Oertel[13], T. Meitinger[1,2,14], C. Gieger[5], B. Müller-Myhsok[3,14], J. Winkelmann[1,2,14,15,16].* 1) Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 2) Institut für Humangenetik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; 3) Max Planck Institute of Psychiatry, Munich, Germany; 4) Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-University, Munich, Germany; 5) Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 6) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-University, Munich, Germany; 7) Paracelsus Elena Klinik, Kassel, Germany; 8) Department of Neurosurgery, University Medical Center, Georg August Universität Göttingen, Göttingen, Germany; 9) Department of Clinical Neurophysiology, University Medical Center, Georg August Universität Göttingen, Göttingen, Germany; 10) Neurologische Klinik, Medizinische Universität Innsbruck, Innsbruck, Austria; 11) Institut für Epidemiologie und Sozialmedizin, Westfälische Wilhelms Universität Münster, Münster, Germany; 12) Zentrum für Schlafmedizin, Charité Universitätsmedizin, Berlin, Germany; 13) Neurologische Klinik, Philips Universität Marburg, Marburg, Germany; 14) Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; 15) Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; 16) Department of Neurology and Neurosciences, Stanford Center for Sleep Medicine and Sciences, Stanford University, Palo Alto, California, USA.

Genome wide association studies of complex traits with common genetic variants could only explain a part of a trait's heritability. The missing heritability is expected to be partially explained by rare genetic variation. But this variation was shown to already stratify populations across relatively short geographical distances within Europe. A German/Austrian sample set of 3,654 patients with the same common complex disease from geographically distinct recruitment areas as well as 2,911 convenience controls (KORA) were genotyped for mainly rare variants using Illumina®'s Infinium HumanExome BeadChip. After quality control, all polymorphic markers were grouped into bins by MAF and conservation, which was defined by exceeding all median conservation scores (PhyloP, GERP++ and SiPhy annotations provided by CHARGE), and were pruned by LD ($r^2 > 0.2$). The bins were used for identity-by-state (IBS) calculations between individuals and the subsequent stepwise assignment to partially dependent sample clusters (as implemented in PLINK). Differentiation of samples within these clusters is negatively correlated with MAF and positively with degree of conservation. Furthermore, IBS calculations were used for multidimensional scaling (MDS) to visualize genetic distances between individuals. The first two MDS components were not correlated between common and rare variant marker bins. The results suggest that (especially conserved) rare variants might offer better discrimination between sub-populations in Germany/Austria than more common variants. Thus inflation of p-values may be a major concern in association analysis of rare variants.

2031S

Population structure in the UK: Rare variant analysis using whole genome sequencing in 3,621 samples in the UK10K cohorts project. *K. Walter[1], S. Metrustry[2], E. Zeggini[1], Y. Memari[1], J. Min[3], J. Huang[1], M. Cocca[4], S. Schiffels[1], I. Mathieson[5], D. Lawson[6], N. Soranzo[1], UK10K Consortium Cohorts Group.* 1) Human Genetics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) Twin Research & Genetic Epidemiology, Kings College London, United Kingdom; 3) MRC CAiTE Centre, University of Bristol, United Kingdom; 4) Institute for Maternal and Child Health-IRCCS "Burlo Garofolo"-Trieste, University of Trieste, Italy; 5) Harvard Medical School, Boston MA 02115, United States; 6) Heilbronn Institute, School of Mathematics, University of Bristol, United Kingdom.

Population structure is a well-characterized potential confounder of association studies based on common variants, but the structural pattern for rare variants and their influence on association studies is less understood. The cohorts arm of the UK10K project undertook whole-genome sequencing at low-read depth (median ~7x) in nearly 4,000 individuals from two large population samples in the UK (TwinsUK, N=1,754 and the Avon Longitudinal Study of Parents and Children (ALSPAC), N=1,867) in a comprehensive exploration of associations between rare and common genetic variants and a set of 61 bio-medically important quantitative phenotypes. The two study cohorts have marked differences in demographic profile with ALSPAC participants originating from a geographically restricted area (Bristol) in the South West of the UK, while the TwinsUK participants were born in different parts of the UK. After stringent QC steps, the data set comprises 42 million SNPs, 3.5 million INDELs and about 18,000 large deletions across 3,621 study participants. Here we describe the extent to which geographic stratification exists at rare variants by focusing on 31 shared 'core' phenotypes in 1,139 twins with available place of birth data throughout the United Kingdom. We modeled genetic structuring using a Euclidian distance metric, a regional grid and generalized additive models (GAM) applied to latitudinal and longitudinal data, and for single nucleotide variants of different minor allele frequencies separately. We further modeled correlation of genotypic and phenotypic data at these geographical locations, and compared them to simulated datasets. Finally, we applied Mantel tests to analyze the significance of genotypic and phenotypic relationships given the distance metrics. Overall, these analyses suggested that there is a moderate genetic structuring of very rare alleles (MAF=0.1-0.3%), however this structure is not associated with phenotypic variation and is unlikely to pose a serious concern for association studies of complex quantitative phenotypes and rare variation in the UK.

2032M

Rare variation and the genomic context of allele-specific expression. *J.R. Davis[1], D.A. Knowles[2], S.B. Montgomery[1,3,5], A. Battle[4,5].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Computer Science Dept., Stanford University, Stanford, CA; 3) Department of Pathology, Stanford University, Stanford, CA; 4) Computer Science Dept., Johns Hopkins University, Baltimore, MD; 5) Co-senior authors.

The functional impact of rare regulatory variation remains largely unexplored, in part due to the limited power of existing genetic association studies and the difficulty of interpreting regulatory variation in general. Allele-specific expression (ASE) offers a powerful, within-individual view of regulatory effects that is robust to the frequency of potential causal variants. However, until now the regulatory context of ASE has primarily been explored in conjunction with common genetic variants underlying expression quantitative trait loci (eQTLs). In this study, we assess the use of ASE to evaluate the impact of non-coding genetic variation, with a focus on rare variants. We use publicly available RNA sequencing data in 373 lymphoblastoid cell lines from the GEUVADIS consortium to perform over 3.3M tests of ASE at roughly 10K genes. This data provides sufficient power to evaluate the association of diverse genomic sequence annotations at candidate regulatory loci with allelic imbalance of nearby genes. Specifically, individuals heterozygous at non-coding loci are pooled to estimate the odds of allelic imbalance in the context of each annotation of interest. Our results demonstrate a positive association between ASE and the level of heterozygosity near a gene in each individual, which is strongly modulated by the position of heterozygous variants with respect to the TSS. Additionally, we find that the likelihood of ASE is associated with the severity of nearby heterozygous mutations; in particular heterozygous SNPs have a weaker effect than indels, which are weaker than structural variants. We determine the value of roughly 65 additional genomic annotations in predicting ASE, including DNase hypersensitivity, transcription factor binding, allele frequency and evolutionary conservation. Finally, we extend this work to develop a statistical model that combines these genomic annotations to predict ASE from individual sequence information. Using this framework, we characterize the effects of rare variation and contrast the genomic context of rare versus common regulatory variants.

2033S

Targeted transcriptomics to compare the susceptibility of human naive and pre-immune volunteers to an infection challenge with viable *Plasmodium vivax* sporozoites. *ML. Rojas-Pena[1], A. Vallejo[4], S. Herrera[3,4], M. Arévalo-Herrera[2,3], G. Gibson[1].* 1) Biology, Georgia Institute of Technology, Atlanta, GA; 2) Faculty of Health, Universidad del Valle, Cali, Colombia; 3) Malaria Vaccine and Drug Development Center, Cali, Colombia; 4) Caucaseco Scientific Research Center, Cali, Colombia.

Malaria is one of the most serious infectious diseases around the world, with an estimated 300 to 500 million cases annually, most of which occur in tropical areas. Infection rates vary across regions of Colombia due to different geographical conditions. Individuals living in high-prevalence malaria regions acquire a natural immunity that gradually eases the symptoms of the disease, and is likely to impact the efficacy of vaccines for malaria. Additionally, little is known about the immunological response to early infection and how the immune system may be boosted during vaccination. This study hypothesizes that the delayed onset and decrease of clinical and parasitological manifestations of malaria observed in pre-immune volunteers from Buenaventura (high prevalence of malaria) compared to naive individuals from Cali (low prevalence of malaria) is associated with altered peripheral blood gene expression. Here we describe the use of nanofluidic Fluidigm quantitative RT-PCR arrays targeting a set of 96 transcripts that are broadly informative of the major axes of variation, to explore the difference in gene expression between Cali and Buenaventura individuals experimentally exposed to viable Plasmodium vivax sporozoites and monitored at six timepoints (Pre-challenge, day 5, day 7, day 9, week 3 and month 4). The largest sources of variation are individuals and time. Site has a relatively minor effect, although it does separate the two clusters in the overall profiles of expression among the samples. Our analyses show that week 3, at the peak of parasitemia, is the most perturbed of the different time points, while among the axes, axis 3 related to B-Cell activation is most perturbed. RNA of the samples will be used to assess transcript abundance genome-wide and to perform gene expression profiling by RNA-seq of the samples to identify differentially expressed genes. Gene expression profiling of lymphocytes can thus be used to identify the type and duration of the immune signals that are biomarkers for vaccine immunogenicity, and establish how pre-immune exposure modifies their activation.

2034M

Conflations of short IBD blocks can bias inferred length of IBD. *C.W.K. Chiang[1], P. Ralph[2], J. Novembre[3].* 1) Ecology and Evolutionary Biology, UCLA, Los Angeles, CA; 2) Computational Biology and Bioinformatics, USC, Los Angeles, CA; 3) Human Genetics, University of Chicago, Chicago, IL.

Identity-by-descent (IBD) is a fundamental concept in genetics with many genetic applications. A common working definition of IBD blocks is that they are contiguous segments of the genome inherited from a recent shared common ancestor without intervening recombination. Long IBD blocks (> 1cM) can be efficiently detected by a number of programs using high-density SNP array data of a population sample. However, all programs detect IBD based on contiguous segments of identity-by-state (IBS). As such, detected IBD blocks could often be due to the conflation of smaller IBD blocks inherited from different common ancestors. Here, we show through theory and simulation that the conflation of small IBD blocks can occur with appreciable frequencies and can lead to errors in estimating the length distribution of IBD blocks, thereby affecting downstream inferences. Specifically, we used coalescent simulations where we know the precise genealogy of the sample and found that, under a realistic demographic model of human history, >35% of the detected IBD segments of 1cM or longer are composed of at least two subsegments, oftentimes each of appreciable length. This effect is universally observed across different IBD detection programs and demographic histories, and was more pronounced for shorter segments (1 to 2cM long) compared to longer segments (>2cM long). Furthermore, we characterized the rate of subsegment conflations as a function of minimum subsegment lengths and gap sizes, and obtained a conflation rate of 0.67 per 1000 pairs of individuals due to segments as small as 0.2cM. Based on this rate, ~28% of all IBD segments between 1 to 2 cM would be due to conflation. Finally, we propose a novel estimator of the de novo mutation rate using IBD blocks detected in population samples. We have found via simulations that such an approach will only work if it explicitly models the conflation of short IBD segments (e.g. 15-fold higher estimates are obtained if conflation is ignored). Thus, the conflation effect should be carefully considered as methods to detect shorter IBD blocks using sequencing data are being developed.

2035S

A systematic analysis of genetic forms of dilated cardiomyopathy reveals numerous ubiquitously expressed and muscle-specific genes. *F.W. Asselbergs[1], G. Kummeling[1], A. Sammani[1], M.P.M. Linschoten[1], A.F. Baas[2], J.J. van der Smagt[2], P.A. Doevedans[1], D. Dooijes[2], M. Mokry[3], M. Harakalova[1].* 1) Department of Cardiology, University Medical Center Utrecht, Utrecht, Netherlands; 2) Department of Medical Genetics, University Medical Center Utrecht, Netherlands; 3) Division of Pediatrics, Wilhelmina Children's Hospital, University Medical Center Utrecht, Netherlands.

Background: Dilated CardioMyopathy (DCM) is a genetic disorder with several tens of genes known to underline its etiology. Despite considerable progress being made in developing next generation sequencing-based diagnostic panels using a subset of the most prevalent 40 DCM genes, the genetic cause remains unsolved in the majority of patients. This suggests that to increase the yield of diagnostic testing all relevant DCM-implicated genes need to be catalogued and carefully considered. Methods and Results: We have conducted a systematic literature search on PubMed, Embase, and OMIM, to find genes that have been implicated in syndromic and non-syndromic DCM and PeriPartum CardioMyopathy (PPCM). Besides strong evidence for mitochondrial inheritance, our search yielded a total of 113 nuclear protein-coding genes. In addition to 42 genes sufficiently reviewed elsewhere we provide a comprehensive annotation of the level of genetic evidence for the set of remaining 71 genes. Next, we investigated the tissue specificity of collected genes using public RNA sequencing data. We show that genes primarily expressed in heart more likely result in DCM with possible myopathies while genes expressed ubiquitously cause DCM with extramuscular manifestations. Conclusions: This comprehensive analysis of genetic evidence of DCM revealed a much higher number of genes than routinely screened at the moment. However, these finding have to be carefully considered and validated in larger cohorts. Our results suggest that targeted sequencing of all known protein-coding genes and the whole Mt-DNA together with consideration of the tissue specificity of mutated genes, will likely increase the yield of genetic testing and facilitate further genotype-phenotype studies in DCM.

2036M

The -844 GA PAI-1 polymorphism is associated with Acute Coronary Syndrome in Mexican population. *I.J. García-González[1,2], Y. Valle[1], E. Sandoval-Pinto[1,3], E. Valdes-Alvarado[1,3], FJ. Muñoz-Valle[1], HE. Flores-Salinas[4], LF. Figuera-Villanueva[5], NO. Dávalos-Rodríguez[6], JR. Padilla-Gutiérrez[1].* 1) Instituto de Investigación en Ciencias Biomédicas, Centro Universitario de Cienc, Guada, Mexico; 2) Doctorado en Genética Humana, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, 44350, Guadalajara, Jalisco, México; 3) Doctorado en Ciencias Biomédicas, Centro Universitario de Ciencias de la Salud, UdeG. Sierra Mojada 950, 44350, Guadalajara, JAL, Mexico; 4) CMNO, IMSS; Centro Médico Nacional de Occidente, 44350, Guadalajara, Jalisco, México; 5) Centro de investigación Biomédica de Occidente, 44350, Guadalajara, Jalisco, México; 6) fInstituto de Investigación en Genética Humana, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, 44350, Guadalajara, Jalisco, México.

Introduction: Acute coronary syndrome (ACS) has an important impact in public health with high morbidity and mortality. Pro-thrombotic and pro-inflammatory states are involved in the pathogenesis of the disease. Plasminogen activator inhibitor-1 (PAI-1) is the major inhibitor of the fibrinolysis and also is part of immune response. The -844 GA gene polymorphism is related to increased PAI-1 protein levels. Aim: We evaluate the association of -844 GA PAI-1 polymorphism with ACS. Methods: A total of 646 individuals were recruited from Western Mexico: 350 unrelated healthy subjects and 296 patients with diagnosis of ACS. PCR-RFLP and PAGE with silver stain were used to identified the polymorphism. Results: The most important risk factor in our population was hypertension, followed by smoking. The genetic distribution showed an association of the A allele (OR=1.27, p=0.04) and AA genotype (OR=1.86, p=0.02) with ACS. The recessive model displayed similar results (OR= 1.76, p=0.02). As additional finding, we observed significant differences in the genetic distribution of ACS dyslipidemic patients (p= 0.04). Conclusion: The A allele and AA genotype of -844 polymorphism of PAI-1 gene are risk factors for ACS. The AA genotype might be associated with the development of dyslipidemia in ACS patients.

2037S

MAT2A Mutations Cause Familial Thoracic Aortic Aneurysms and Aortic Dissections. *D. Guo[1], L. Gong[1], E.S. Regalado[1], R.L. Santos-Cortez[2], B. Cai[1], R. Zhao[1], S. Veeraraghavan[3], M. Willing[4], G. Jondeau[5], C. Boileau[6], M.J. Bamshad[7], J. Shendure[7], D.A. Nickerson[7], S.M. Leal[2], H. Pannu[1], R. Moran[8], C.S. Raman[3], E.C. Swindell[9], D.M. Milewicz[1].* 1) Dept Internal Med, Univ Texas/Houston Med Sch, Houston, TX; 2) Dept Mol & Hum Genet, Baylor College of Medicine, Houston, TX; 3) Sch of Pharmacy, Univ Maryland, Baltimore, MD; 4) Washington Univ/Sch Medicine, St. Louis, MO; 5) 5Centre National de Référence pour le syndrome de Marfan et apparentés, Hôpital Bichat, Paris, France; 6) 6INSERM U383, Hôpital Necker-Enfants Malades, Université Paris, Paris, France; 7) Dept Genome Science, Univ Washington, Seattle, WA; 8) Genomic Med Inst, Cleveland Clinic, Cleveland, OH; 9) Dept Pediatrics, Univ Texas/Houston Med Sch, Houston, TX.

We performed genome-wide linkage and exome sequencing using DNA from members of a large family (TAA059) with autosomal dominant inheritance of thoracic aortic aneurysms and dissections (FTAAD) associated with bicuspid aortic valve (BAV). Analysis of DNA from an affected family member identified no mutation in known FTAAD genes and no novel copy number variants (CNV) on an Illumina SNP array analysis. Exome sequencing of DNA from multiple affected family members identified only a single rare variant in MAT2A, c.1031A>C (p.Glu344Ala), that fell under the major linkage peak and segregated with TAAD in the family (two-point LOD score of 2.10). This variant was predicted to lead to loss of function of the protein, was located near the active site, and altered a conserved amino acid. Sequencing of an additional 409 unrelated probands with FTAAD identified another MAT2A variant, p.Arg356His, which was also in the active site, conserved, and predicted to be damaging; no additional samples were available to test the segregation of this variant with disease in the family. MAT2A is highly expressed in aortic smooth muscle cells and encodes S-adenosylmethionine synthase isoform type-2, which catalyzes the synthesis of S-adenosylmethionine (SAM). Reduction of cellular SAM levels would be predicted to decrease methylation capacity and increase oxidative stress in cells. To confirm the mutation disrupts protein function, an antisense mat2aa-specific morpholino was used to knockdown expression in zebrafish. The mat2aa-specific morpholino produced a range of cardiovascular phenotypes at 72 h postfertilization, ranging from mild (pericardial edema) to severe cardiovascular phenotypes (prominent cardiac and pericardial edema). Co-injection of the morpholino with wild-type MAT2A mRNA partially rescued these phenotypes, whereas mRNA with the MAT2A p.Glu344Ala alteration did not, indicating the mutation disrupts protein function. In conclusion, both genetic and functional data suggest that loss-of-function MAT2A mutations may be a rare cause of familial thoracic aortic aneurysms and dissections associated with bicuspid aortic valve.

2038M

Associations of Endothelial System Genes with Blood Pressure Changes and Hypertension Incidence: the GenSalt Study. *F. Liu[1,2], J. He[2], D. Gu[1], D. Rao[3], J. Huang[1], J. Hixson[4], C. Jaquish[5], J. Chen[1], C. Li[2], X. Yang[1], J. Li[1], T. Rice[3], L. Shimmin[4], T. Kelly[2].* 1) Department of Evidence Based Medicine, Fuwai Hospital, National Center of Cardiovascular Diseases, PUMC & CAMS, Beijing, China; 2) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine; 1440 Canal Street, Suite 2000, New Orleans, LA 70112, USA; 3) Division of Biostatistics, Washington University School of Medicine, St. Louis, MO; Campus Box 8067, 660 South Euclid Ave. St. Louis, MO 63110-1093, USA; 4) Department of Epidemiology, Human Genetics and Environmental Sciences, University of Texas School of Public Health, Houston, TX, USA; 1200 Pressler Street, Houston, TX 77030, USA; 5) Division of Prevention and Population Sciences, National Heart, Lung, Blood Institute, Bethesda, Maryland, USA.

High blood pressure (BP) is a complex trait, influenced by multiple environmental and genetic determinants. Although established as a heritable trait, the genomic mechanisms underlying BP regulation remain largely unknown. The objective of the current study was to examine the associations of 206 common variants in 15 endothelial system genes with BP changes and hypertension incidence among 1,775 Han Chinese participants of the family-based Genetic Epidemiology Network of Salt Sensitivity (GenSalt) follow-up study. Nine BP measurements were obtained at baseline and during each of two follow-up observations using a random-zero sphygmomanometer. The associations of 206 SNPs in 15 endothelial system genes with BP changes and hypertension incidence were assessed using mixed models to account for the correlations of repeated measures among individuals and within families. A genotype by time interaction term was used to model differences in longitudinal BP change according to genotype over time. Gene-based analyses were conducted using the truncated product method. The Bonferroni method was used to adjust for multiple testing in all analyses. Among those free from hypertension at baseline, 513 (32.1%) GenSalt participants developed hypertension during the average 7.2 years of follow-up. In single-marker analyses, each copy of the minor alleles of SELE markers rs4656704, rs6427212 and rs5368 was associated with increased risk of hypertension with relative risks (95% confidence intervals) of 1.42 (1.18, 1.71), 1.46 (1.21, 1.75) and 1.46 (1.21, 1.77), respectively (P= 1.66×10-4, 7.44×10-5 and 8.51×10-5, respectively). SELE marker rs3917436 predicted longitudinal DBP change (P=8.27×10-5). Results of gene-based analyses showed the SELE gene was significantly associated with SBP change, DBP change and hypertension incidence (all P<1.00×10-6). Furthermore, the DDAH1, SELP, and COL18A1 genes were associated with SBP change (P=2.00×10-6, 9.00×10-5, and 1.00×10-6, respectively), while EDNRA was associated with hypertension incidence (P=2.00×10-4). In conclusion, the current study provides strong evidence of a role of endothelial system genes in BP progression and hypertension incidence. Future studies will be required to identify the causal variants underlying the observed associations.

2039S

Novel NKX2.5 mutation associated with congenital heart disease in South Indian patients. *S. Mattapally[1], S. Nizamuddin[3], KS. Murthy[2], K. Thangaraj[3], SK. Banerjee[1].* 1) Division of Medicinal Chemistry and Pharmacology, CSIR-Indian Institute of Chemical Technology, Uppal Road, Hyderabad 500 007, India; 2) Innova Children's Heart Hospital, Tarnaka, Hyderabad 500017, India; 3) CSIR-Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500 007, India.

Background: NKX 2.5 is one of the transcription factors, it is a cardiac specific homeobox gene and act as early marker genes for heart field development. Heterozygous NKX 2.5 germline mutations were reported to cause congenital heart disease (CHD), hence the aim of this study is to find the association of NKX 2.5 mutations with CHD among the south Indian CHD patients. Method: NKX 2.5 gene was sequenced in 100 CHD patients (ASD, VSD, TOF and SV) and 200 controls. Functional significance of the observed NKX2.5 mutations were analyzed using in silico software such as; Polyphen, SIFT, PMut, plink, etc. Results: Our analysis with NKX2.5 gene revealed a total of 7 mutations, out of which, 3 were in intronic region, 3 mutations were in coding region and 1 mutation was in 3' UTR. Of the above mutations, one was found to be associated with Tetralogy of Fallot (TOF) and two (rs2277923 and a novel mutation) were strongly associated with VSD. Interestingly, one novel missence mutation (p-value = 0.009744; Asp16Asn) was most significant findings of this study. Our 'in silico' analysis also provides evidence that some of the mutations reported above are pathogenic. Conclusion: The present study found that NKX2.5 genetic variations are associated with TOF and VSD in South Indian patients.

2040M

A replication study for fifteen coronary artery disease susceptible loci in a Japanese population. *K. Ozaki[1], Y. Sakata[2], S. Suna[2], Y. Onouchi[1,3], T. Tsunoda[4], M. Kubo[5], I. Komuro[2,6], T. Tanaka[1,7].* 1) Lab for Cardiovascular Diseases, RIKEN, Ctr Integrative Medical Sci, Yokohama, Japan; 2) Dept of Cardiovascular Medicine, Osaka University Graduate School of Medicine, Suita, Japan; 3) Dept of Public Health, Chiba University Graduate School of Medicine, Chiba, Japan; 4) Lab for Medical Sci Mathematics, RIKEN, Ctr for Integrative Medical Sci, Yokohama, Japan; 5) Lab for Genotyping Development, RIKEN Ctr for Integrative Medical Sci, Yokohama, Japan; 6) Dept of Cardiovascular Med, Graduate School of Med, The University of Tokyo, Tokyo, Japan; 7) Dept of Human Genetics and Disease Diversity, Tokyo Medical and Dental University, Tokyo, Japan.

Recent large-scale association analysis for coronary artery disease (CAD) identified fifteen new loci with genome wide statistical significance. Replication for the genetic association in other ethnic population is one of important issue to investigate a clinical benefit in the future. We conducted here the reproducibility for the association of the 15 single nucleotide polymorphism (SNP) loci (*IL6R*; rs4845625, *APOB*; rs515135, *ZEB2-AC074093.1*; rs2252641, *VAMP5-VAMP8-GGCX*; rs1561198, *ABCG5-ABCG8*; rs6544713, *GUCY1A3*; rs7692387, *EDNRA*; rs6842241, *SLC22A4*; rs17689550, *KCNK5*; rs10947789, *PLG*; rs4252120, HDAC9; rs2023938, *LPL*; rs264, *TRIB1*; rs2954029, *FLT1*; rs9319428, *FURIN* rs17514846) and CAD with 7,990 cases and 6,582 controls in a Japanese population. We found a convincing association with statistical significance for rs9319428 in intron 6 of *FLT1*, encoding fms-related tyrosine kinase 1, for CAD susceptibility ($P = 5.98 \times 10^{-8}$). We also found the replication for the association of two SNPs, rs6842241 in 5'-fanking region of *EDNRA*, encoding endothelin receptor type A, and rs17514846 in intron1 of *FURIN*, encoding a calcium-dependent serine endoprotease, with CAD (P = 0.000207 and P = 0.00208, respectively). Our validation results revealed that the three loci were genetic risk factor for CAD in the Japanese population.

2041S

Relations between PAI-1 plasma levels and adipose tissue expression of PAI-1 and hsa-miR-421 microRNA. *DA. Tregouet[1], M. Civelek[2], L. Markku[3], MC. Alessi[4], AJ. Lusis[2], PE. Morange[4].* 1) UMR_S 1166, INSERM, Paris, France; 2) Departments of Microbiology, Medicine, and Human Genetics, UCLA School of Medicine, Los Angeles, USA; 3) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 4) Inserm U1062, Aix-Marseille University, Marseille, France.

Several works have demonstrated that increased plasminogen activator inhibitor (PAI)-1 levels, the main inhibitor of the fibrinolytic system, could predict the risk of metabolic syndrome (MetS) and associate with its separate components, such as visceral obesity, increased blood glucose levels and dyslipidaemia. Biological mechanisms responsible for these associations are still debated and several arguments are in favor of a key role of adipose tissue in these phenomena. As part of the METSIM study, we measured plasma PAI-1 activity in a sample of 850 healthy males for which the subcutaneous adipose tissue expression of the SERPINE1 gene coding for PAI-1 was also available. We demonstrated that adipose tissue SERPINE1 expression was correlated to plasma PAI-1 activity ($\rho = 0.23$, $p = 2.68\ 10^{-11}$) and all MetS components (ρ as low as 0.31, $p < 10^{-16}$). The influence of SERPINE1 expression on MetS components was mainly independent on plasma PAI-1 activity levels as their correlations were hardly modified by the adjustment on plasma PAI-1 activity. Conversely, all correlations were strongly reduced (ρ no higher than 0.14, $p = 2.30\ 10^{-5}$) after adjusting for BMI. In a subsample of 194 METSIM individuals, we were also able to show that both plasma PAI-1 activity and adipose tissue SERPINE1 mRNA expression positively correlated with adipose tissue expression of the hsa-miR-421 microRNA ($\rho = 0.22$, $p = 0.0018$, and $\rho = 0.23$, $p = 0.0014$, respectively), a microRNA whose plasma levels have been recently shown to correlate with plasma PAI-1 levels. Interestingly, the association between adipose tissue levels of hsa-miR-421 and plasma PAI-1 activity was still significant after adjusting for BMI ($p = 0.015$) and SERPINE1 mRNA levels ($p = 0.013$). This study confirms in a large epidemiological cohort the strong association between PAI-1 plasma levels and adipose tissue serpin 1 expression, and provides novel elements suggesting, for the first time, that this expression may be influenced by adipose tissue expression of the hsa-miR-421 microRNA.

2042M

Functional fine mapping of the genes involved in plasma lipid metabolism in the LD-block of *NCAN/CILP2/PBX4* region. *S. Boonvisut, S. Makishima, K. Watanabe, K. Nakayama, S. Iwamoto.* Division of Human Genetics, Center for Molecular Medicine, Jichi Medical University, Shimotsuke, Tochigi, Japan.

Dyslipidemia is a major risk factor of cardiovascular diseases. The estimated heritability ranges 0.20-0.60. *NCAN/CILP2/PBX4* region is the initially identified locus by GWAS that substantially influences plasma lipid concentrations. In addition, the genetic analysis of non-alcoholic fatty liver disease (NAFLD) reported that risk allele increasing LDL and TG levels in this region reduced the risk of NAFLD. These studies suggested that liver is one of the main organs where the responsible gene acts to regulate plasma lipid metabolism. However, *NCAN/CILP2/PBX4* region is in a tight linkage-disequilibrium block spanning over 300kb, in which 11 genes are encoded and the responsible gene has not yet been identified. This study aims to identify the molecule involved in lipid metabolism from this region. Four genes expressed in liver, *Tm6sf2*, *Sugp1*, *Gatad2a*, and *Yjefn3*, were chosen from the 11 genes referring to ENCODE database. The short-hairpin RNA (shRNA) templates against them were synthesized and inserted in adenovirus vectors. The recombinant adenoviruses were administrated into mice and the plasma lipid levels were followed up. Knocking-down of *Sugp1* showed increased serum cholesterol level in both feeding and fasting states ($p = 0.044$ and 0.043 respectively), and *shGatad2a* decreased it in fasting state ($p = 0.0411$). While, transcriptome Analysis of these mice revealed that knocking down of *Gatad2a* and *Sugp1* affected on several lipid metabolism pathways, such as fatty acid metabolism, steroid synthesis and bile acid biosynthesis, it was noteworthy that *shSugp1* enhanced the expression of *ApoB* expression (3.99 fold, p-value < 0.01). In order to study the effect of SUGP1 on the expression of *APOB* in another main organ, intestinal epithelium, Caco-2 cells with enforced expression and knockdown of *SUGP1* were established using lentiviral gene transduction. The established cells were induced to be differentiated by culturing on filter insert. The mRNA of *APOB*, secreted APOB-48 and -100 protein levels and the ratios of them were increased by the knock-down of *SUGP1*. The *APOB* mRNA-editing for the conversion from APOB-100 to -48 in differentiated Caco-2 cells was increased by the knocking-down and decreased by the enforced expression of *SUGP1*. SUGP1 (SURP and G patch domain containing 1) is estimated to be involved in mRNA processing. These data indicated that *SUGP1* is the most powerful candidate gene regulating serum lipid levels in *NCAN/CILP2/PBX4* region.

2043S

Mutations in genes NKX2.5, GATA4 and TBX5, associated to congenital heart disease with septal defect in pediatric patients from the Guadalajara Civil Hospital Fray Antonio Alcalde. *R. Diaz Martinez[1,2], G. Perez-Garcia[1], M.L. Ornelas-Arana[1], M.S. Hernandez-Flores[3], S.J. Jasso-Bernal[3], X.M. Boldo-Leon[2], J.M. Magaña-Cerino[2], C. Rodriguez-Perez[2], M.C. Martinez-Lopez[2].* 1) Clinical Genetic, Hospital Civil de Guadalajara, Guadalajara, Mexico; 2) Laboratorio de Diagnóstico Molecular, DACS-UJAT, Tabasco, Mexico; 3) Pediatric Cardiology, Hospital Civil de Guadalajara, Guadalajara, Mexico.

Introduction. Congenital heart disease is the leading cause of birth defects, affecting approximately 1%; of all live births. It presence has been associated with single mutations in transcription factors that regulate heart development. The genes frequently reported are GATA4, NKX2.5 and TBX5. These interact during cardiogenesis. Objectives. Investigate the mutations frequently reported in literature of the genes, GATA4 , NKX2.5 and TBX5 in children with congenital heart disease involving septal defects in HCFAA. Material and Methods. We selected 30 children with congenital heart disease (CHD) with septal defect, a genetic history was elaborated and by echocardiography CHD type in each patient was determined. DNA was extracted from peripheral blood for genotyping by PCR -HRM mutations in GATA4 (c.886G > A), TBX5 (c.238G > A) and NKX2.5 (c.73C>T and c.533 C>T) genes, to patients and their parents. The results were analyzed by descriptive statistics. Results. In total 84 samples of all participants were taken, 30 of these samples were from patients diagnosed with CHD with septal defect and 54 of their parents. We found seven patients with mutations in genes previously selected for this study, the mutations seen in patients were also seen in all occasions in one or both parents. We found 3 female parents with mutations, and 5 male. In 23.33%; of patients with CHD with septal defect found association between two or more of the studied mutations. In parents of patients with mutations we found that one or both parents had the same mutations. The frequency of mutations in index cases (N = 30) was 13.3%; (4) for mutation in TBX5 , 16.6%; (5) for mutation in GATA4 and 10%; (3) for mutation in NKX2.5 (c.73C > T) . Finding that approximately 60%; of patients with this mutations appeared in double or triple, mainly related to the GATA4 gene. Conclusions. We found a frequency of these mutations in the index cases of 13.3%; for the mutation in TBX5, 16.6%; for the mutation in GATA4 and 10%; mutations in NKX2.5 (c.73C> T). No patient had mutation c.533C> T in gene NKX2.5. From 30 families we found that 23%; of them had single or combined mutations of the genes studied. The type of inheritance resembles autosomal dominant. In carrier parents no abnormalities were reported, however they inherited the mutation.

2044M

Standard Schnauzer dogs with dilated cardiomyopathy have a 22 bp deletion and frame shift in RBM20. *D. Gilliam[1], M.W. Harmon[2], G.S. Johnson[1], T. Mhlanga-Mutangadura[1], L. Hansen[1], J.F. Taylor[3], R.D. Schnabel[3], S.B. Leach[2].* 1) Veterinary Pathobiology, University of Missouri, Columbia, MO; 2) Department of Veterinary Medicine, University of Missouri, Columbia, MO; 3) Animal Sciences, University of Missouri, Columbia, MO.

To identify the molecular genetic cause for dilated cardiomyopathy (DCM) in Standard Schnauzers, we generated a 30-fold average coverage whole genome sequence (WGS) with DNA from a male Standard Schnauzer that died at 13.5 months of age with DCM and congestive heart failure. This WGS contained 69 homozygous sequence variants that were predicted to alter the primary structure of the gene product and absent from the WGSs of 101 canids not exhibiting DCM. A homozygous 22 bp deletion and frame shift in exon 11 of *RBM20* was the sequence variant considered most likely to be causal because mutations in the human and rodent orthologs cause DCM. All 9 of the confirmed DCM cases for which DNA was available tested homozygous for the deletion allele (del/del). A prospective screen for the mutation identified 653 Standard Schnauzers that were homozygous for the wild type allele, 225 Standard Schnauzers that were heterozygous, and 13 additional Standard Schnauzers that were homozygous for the deletion allele. Eight of these 13 del/del Standard Schnauzers have subsequently been diagnosed with occult DCM via echocardiography. The other 5 were not available for evaluation and 3 of them were <1 year old, before clinical signs typically appear. Echocardiographic evaluation of 17 Standard Schnauzers that were heterozygous for the deletion allele, including a 15 year old Standard Schnauzer, has revealed no echocardiographic evidence of myocardial dysfunction. Thus, *RBM20*-associated DCM appears to be recessive in Standard Schnauzers; whereas, human *RBM20*-associated DCM is reported to be a dominant trait. In addition, there appears to be a gender difference in survival time among Standard Schnauzers. Male Standard Schnauzers with the homozygous *RBM20* deletion have survived to a mean age of 16.8 months (range 10.5 months to 30 months). Female Standard Schnauzers with the homozygous RBM20 deletion have survived to a mean age of 43.4 months (range 19 months to 55 months). To our knowledge, a similar gender difference has not been reported for human *RBM20*-associated DCM patients.

2045S

Disruption of the SEMA3D gene in a patient with congenital heart defects. *C. Le Caignec[1,2,3], M. Sanchez-Castro[2], O. Pichon[1], D. Poulain[1], A. Briand[1], V. Gournay[4], A. David[1].* 1) Medical genetics department, CHU Nantes, Nantes, France; 2) INSERM, UMR1087, l'institut du thorax, Nantes, France; 3) Université de Nantes, Nantes, France; 4) Pediatric cardiology department, CHU Nantes, Nantes, France.

Congenital heart disease (CHD) is the leading malformation among all newborns. Although genetic variation contributes to CHD, the genetic basis for the disease remains unknown in the majority of the patients. We report a child with transposition of the great arteries, ventricular septal defect and coarctation of the aortic isthmus. By array comparative genomic hybridization (aCGH), we identified a duplication of the 5' half of the semaphorin 3D (SEMA3D) gene. Breakpoint sequencing and fiber fluorescent in situ hybridization showed a tandem duplication. SEMA3D expression studies, performed on the patient's immortalized lymphoblasts, showed a higher level of expression of SEMA3D mRNA compared to the controls. Moreover, we demonstrated the presence of a truncated SEMA3D poly-A tailed mRNA, resulting from an abnormal transcription of the partial duplication of SEMA3D. Sema3D is an axon guidance protein essential for the correct migration of cardiac neural crest cells into the outflow tract during cardiogenesis. Sema3D null mice present with anomalous pulmonary venous connections and atrial septal defects but the role of SEMA3D in humans remains unclear. The results suggest that a truncated SEMA3D may have hampered the migration of cardiac neural crest cells during heart development, and consequently contributed to CHD in our patient.

2046M

Long QT Syndrome - Family Studies in Ion Channel Encoding Gene. *P. Nallari.* Dept. of Genetics, Osmania University, Hyderabad, Andhra Pradesh, India.

Congenital Long-QT syndrome (cLQTS) is an inherited arrhythmogenic disease characterized on ECG by a prolonged QTc interval. The ECG manifestation reflects an abnormally prolonged ventricular action potential, which can be the substrate for life-threatening arrhythmias that lead to syncope and sudden cardiac death. Heterogeneity of clinical manifestations corroborated with genetic etiology is well established. Screening of ion channel encoding genes was carried out by PCR-SSCP analysis followed by commercial sequencing in 46 LQTS probands and 69 available family members. We report novel variations in four LQTS probands and their family members in KCNQ1 gene. In-silico analysis predicts that these variations may lead to abnormal IKs channel and altered ion fluxes. The variation is also observed in the family members of the probands which helps in predictive testing and prognosis. The study is aimed at highlighting the importance of family screening in clinical management of LQTS with emphasis on personalized medicine.

2047S

Assessment of the *TNF*-α rs1799964 (-1031T>C) polymorphism and soluble protein concentration in Acute Coronary Syndrome: association with circulating levels. *E. Sandoval-Pinto[1,2], JR. Padilla-Gutiérrez[2], E. Valdes-Alvarado[1,2], IL. García-González[2,3], A. Valdez-Haro[2,3], JF. Muñoz-Valle[2], HE. Flores-Salinas[4], F. Rivas[5], Y. Valle[2].* 1) Doctorado en Ciencias Biomédicas, CUCS, UdeG, Guadalajara, Jal., Méx; 2) Instituto de Investigación en Ciencias Biomédicas, CUCS UdeG, Guadalajara, Jal., Méx; 3) Doctorado en Genética Humana, CUCS, UdeG, Guadalajara, Jal., Méx; 4) CMNO, IMSS; Centro Médico Nacional de Occidente, Guadalajara, Jal., Méx; 5) Hospital General de Occidente, Secretaria de Salud Jalisco, Guadalajara, Jal., Méx.

Introduction: The acute coronary syndrome (ACS) is a complex disease where genetic and environmental factors are involved. *TNF*-α is a candidate gene for ACS progression due to its contribution in the inflammatory process and endothelial function. The rs1799964 polymorphism in the *TNF*-α gene consists of a T>C change at the -1031 promoter region. The C allele has been associated with changes in gene expression and plasmatic levels thus associated with a decrease in the risk of cardiovascular disease. Objective: To associate the *TNF*-α rs1799964 polymorphism with ACS and to measure the serum levels of TNF-α. Methods: Were recruited 251 patients with ACS classified according to American College of Cardiology and 164 healthy subjects (HS) age-matched from Western Mexico. The study was made in accordance with de Declaration of Helsinki. All individuals accepted to participate and an informed written consent was obtained. The rs1799964 polymorphism was identified by PCR-RFLP. Fragments were separated in polyacrylamide gel electrophoresis. The TNF-α was measured using an enzyme-linked immunosorbent assay. The genotype and allele differences were estimated by Fisher's exact test. The association measure was evaluated by OR and 95% of confidence intervals. The Mann-Whitney U test was applied in order to compare the TNF-α serum lelevels. The significance level was $p<0.05$. Results: The genotype and allele frequencies of the rs1799964 polymorphism showed statistically differences between groups (OR= 0.317, p= 0.01; OR= 0.688, p= 0.03, respectively). These results suggest that the C allele carriers have 1.44 less susceptibility to ACS. Also, the goodness of fit was performed to identify the most likely heritage model. In this context, significant differences were found in the recessive genetic model (OR: 0.333, p = 0.02). The TNF-α levels were significantly higher in ACS patients compared to HS (38.05 vs 23.92 ng/mL, p=0.0001). The C/C genotype carriers showed lower levels of TNF-α when compared to heterozygous ACS patients; this finding could not be replicated in HS (36.95 vs 33.52 ng/ml, p=0.03). Conclusion: The *TNF*-α gene polymorphism rs1799964 (-1031T>C) is a susceptibility genetic marker for Acute Coronary Syndrome in Western Mexico population. In addition, serum levels of TNF-α may be a biological marker of ACS.

2048M

Whole exome sequencing identifies a *NEXN* mutation in a family with left ventricular noncompaction. *J.M. Taylor[1], K. Thomson[1], J. Williams[1], L. Witty[2], J. Craft[2], J. Taylor[2,3], A. Seller[1], E. Blair[4], H. Watkins[2,5].* 1) Oxford Medical Genetics Laboratories, Oxford University Hospitals NHS Trust, Oxford., UK. OX3 7LE; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK OX3 7BN; 3) NIHR Biomedical Research Centre, Oxford, UK; 4) Department of Clinical Genetics, Oxford University Hospitals NHS Trust, Oxford, UK OX3 7LE; 5) Department of Cardiovascular Medicine, University of Oxford, Oxford, UK. OX3 9DU.

Background: Hypertrophic cardiomyopathy (HCM) and dilated cardiomyopathy (DCM) are genetically and phenotypically heterogeneous. *NEXN*, which encodes a component of the cardiac Z disc, has been previously reported to contain mutations in patients with HCM and DCM. The age of onset and disease severity associated with *NEXN* mutation has shown some inter-individual variability.

Methods: Whole exome DNA sequence analysis was performed on a small nuclear family with maternal left ventricular noncompaction (LVNC) cardiomyopathy and in-utero "dilated" cardiomyopathy with severe hydrops fetalis in two affected children. All 3 affected individuals also had a large ASD. Post mortem analysis on one of the children revealed LVNC cardiomyopathy. Interestingly the other affected child's cardiac studies normalised ex-utero. Standard clinical molecular analyses including array CGH, candidate nuclear gene analysis and selected mitochondrial genome mutation analysis had failed to identify a pathogenic variant. Exome capture was undertaken using the NimbleGen SeqCap EZ v2.0 library and sequenced using an Illumina HiSeq2000. The data were masked and analysed for 54 genes previously implicated in causing HCM or DCM or LVNC.

Results: An inframe single amino acid deletion (p.Gly650del) in the last exon of the *NEXN* gene was identified in all three affected individuals. The p.Gly650del variant has been previously reported in adult patients with dilated cardiomyopathy; animal studies supported pathogenicity of this variant, however segregation studies were not reported. The extreme phenotypic diversity seen in our family and the normalisation of the cardiac studies in the proband remain unexplained. Further clinical evaluation of the maternal family is being planned.

Conclusion: Using an exome sequencing strategy we have successfully identified a likely pathogenic variant within the *NEXN* gene and provided evidence that LVNC is associated with variants in this gene. Further analysis is underway to examine the extreme phenotypic variability seen with the p.Gly650del variant and to determine whether this result can be used for prenatal diagnosis in this family.

2049S

Association of rs4340 ACE polymorphism with acute coronary syndrome in Mexican population. *A. Valdez Haro[1,2], Y. Valle[1], I.J. García-González[1,2], E. Sandoval-Pinto[1,3], E. Valdes-Alvarado[1,3], JF. Muñoz-Valle[1], JR. Padilla-Gutiérrez[1].* 1) Instituto de Investigación en Ciencias Biomédicas, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, 44350, Guadalajara, Jalisco, México; 2) Doctorado en Genética Humana, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, 44350, Guadalajara, Jalisco, México; 3) Doctorado en Ciencias Biomédicas, Centro Universitario de Ciencias de la Salud, UdeG. Sierra Mojada 950, 44350, Guadalajara, JAL, Mexico.

Introduction: Acute Coronary Syndrome (ACS) is an important public health problem and a main cause of death, in which the classic risk factors can barely explain half of the cases. A proposed explanation is that the sum of unfavorable gene polymorphisms and a proper environment may lead to the disease. Angiotensin converting enzyme (ACE) increase angiotensin II (AII) plasmatic levels, which is a powerful vasoconstrictor that leads to endothelial and myocardial oxidative stress and aterosclerosis. The polymorphism rs4340 consists in an Insertion/deletion (I/D), in which the D allele has been related to increase ACE levels, therefore to risk for ACS. Aim: To determine the association of rs4340 polymorphism of ACE gene in ACS. Methods: 266 ACS patients and 147 healthy subjects (HS) matched by age, were recruited from western Mexico. The ACS selection based on the American College of Cardiology criteria. An informed written consent was obtained from all patients. The rs4340 polymorphism was identified by PCR and PAGE with silver stain. Results: The genetic frequencies in healthy subjects were on Hardy-Weinberg equilibrium expectations (p=0.13). Allelic and genotype distributions were compared. Significant differences between groups were found. The allele D was 1.47 more frequent in ACS patients than HS subjects (OR = 1.47, p= 0.008). The genotype D/D was also more commonly found in the SCA group (OR = 1.96, p = 0.014). The recessive model supported this results (p=0.03). Conclusion: The D allele and DD genotype is a genetic marker for susceptibility to ACS in Western Mexican population.

2050M

HLA-DRB1*01 allele associates with Acute Coronary Syndrome (ACS) in Finnish population. *E. Vlachopoulou[1], M. Marchesani[1], J. Nokelainen[1], M.S. Nieminen[2], J. Sinisalo[2], M-L. Lokki[1].* 1) Transplantation Laboratory, Haartman Institute, Haartmaninkatu 3, University of Helsinki, Helsinki, Finland; 2) HUCH Heart and Lung Center, Division of Cardiology, Helsinki University Central Hospital, Finland.

Background: Acute Coronary Syndrome (ACS) is one of the leading causes of death in the world, still being a diagnostic and management challenge. Inflammation has a central role in the pathophysiology of ACS. However, the ultimate reasons for the inflammation remain unrevealed. The system that regulates inflammation and immunity lies in big part in the Human Leukocyte Antigen (HLA) system on chromosome 6p21.31. Our previous smaller candidate gene study suggested that HLA-DRB1*01 allele of MHC class II is associated with ACS. The aim of this study is to confirm this association in a larger material. Methods: We studied 2090 ACS patients and 1580 geographically-matched controls (Corogene study) having information of various risk factors and survival data. All subjects were genotyped for the HLA-DRB1*01 allele to determine allele copy numbers with the genomic real-time quantitative polymerase chain reaction. The HLA-DRB1*01 was tested for association with ACS. Results were replicated in two different replication materials from Finland. Results: HLA-DRB1*01 was associated with ACS (freq. cases 38.9% vs. freq. controls 32.4%; odds ratio 1.33; P = 7.7 ×10-5). Interaction analysis showed that total cholesterol, low-density lipoprotein cholesterol, and triglyceride levels interacted with HLA-DRB1*01 (odds ratio 1.23 [95% CI 1.01-1.49]; P = 0.04, odds ratio 1.20 [95% CI 1.06-1.36]; P = 0.004, odds ratio 1.18 [95% CI 1.02-1.37]; P = 0.03, respectively). The survival curves stratified for HLA-DRB1*01 copies did not show any statistical difference. Conclusion: We showed that HLA-DRB1*01 is associated with ACS in Finnish materials. This gene may play a role in promoting ACS by affecting antigen presentation. Peptide fragments of oxidized low-density lipoprotein cholesterol may connect the cholesterol pathway to inflammation and immunology in atherosclerosis progression.

2051S

Exploring the role of common and rare platelet traits-associated variants in cardiovascular disease risk. *N. Chami[1,2], S. de Denus[1,2], M.P. Dubé[1,2], J. Rioux[1,2], J.C. Tardif[1,2], G. Lettre[1,2].* 1) Montreal Heart Institute, Montreal, PQ, Canada; 2) Faculté de Médecine, Université de Montréal, Montréal, Québec, Canada.

Introduction: Platelets are pivotal components of the hemostatic and thrombotic responses. Their count and volume are also useful biomarkers for several human diseases, and platelet features and activation have been associated with cardiovascular diseases (CVD). Platelet phenotypes are highly heritable: h2~0.5-0.6. Genome-wide association studies (GWAS) have identified >65 loci that carry common DNA sequence variants associated with platelet count or Mean Platelet Volume (MPV). In this study, we sought to evaluate the relationship between the known platelet traits-associated variants and prevalent CVD. Methods: We genotyped 9,660 individuals from the biobank of the Montreal Heart Institute (MHI) on the IlluminaExomeChip array. We generated a genetic risk score (GRS) using genotypes from 35 SNPs and four rare variants associated with platelet traits, and tested the association between this GRS and CVD. We also considered the effect of deleterious rare variants in 12 thrombocytopenia and 7 myeloproliferative neoplasm (MPN) genes on platelet traits and CVD risk in the MHI biobank. Results: The GRS was significantly associated with platelet count and MPV (both P<1×10-16) and with increased risk of CVD (OR=1.46, P=0.025). Interestingly, inclusion of recently published rare variants associated with platelet traits in the GRS improved the association with CVD (OR=1.65, P=0.004). With rare variants in the model, the GRS was also associated with stroke (OR=2.11, P=0.035) and myocardial infarction (OR=1.49, P=0.023). Using the sequence kernel association test (SKAT), we tested the association between rare (minor allele frequency <0.1%) nonsense, missense and splice site variants in 19 thrombocytopenia or MPN genes with platelet count and MPV. JAK2 and MYH9 were significantly associated with platelet count (P=7.15×10-8 and P=5.50×10-3, respectively) and TUBB1 with MPV (P=0.0001). The gene-based results for JAK2 and TUBB1 had previously been described. Further, rare coding variants in ITGB3, a gene implicated in thrombocytopenia, were associated in aggregate with stroke (P=0.0019), although the association with platelet count or volume was not significant. Conclusion: Our results suggest that common and rare variants within genes associated with platelet traits and disorders influence CVD risk.

2052M

GWAS-identified loci for coronary heart disease are associated with intima-media thickness and plaque presence at the carotid artery bulb. *M. den Hoed[1], R. Strawbridge[2], P. Almgren[3], S. Gustafsson[1], T. Axelsson[4], G. Engström[3], U. de Faire[5], B. Hedblad[3], S.E. Humphries[6], C.M. Lindgren[7], A.P. Morris[7,8], G. Östling[3], A-C. Syvänen[4,9], E. Tremoli[10], A. Hamsten[2], E. Ingelsson[1], O. Melander[3], L. Lind[11].* 1) Medical Sciences, Mol. Epidemiol, SciLifeLab, Uppsala University, Uppsala, Uppsala, Sweden; 2) Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital Solna, Stockholm, Sweden; 3) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University and Lund University Diabetes Centre, Malmö, Sweden; 4) Department of Medical Sciences, SNP & SEQ Technology Platform, Uppsala University, Uppsala, Sweden; 5) Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 6) Centre for Cardiovascular Genetics, University College London, London, UK; 7) Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 8) Department of Biostatistics, University of Liverpool, Liverpool, UK; 9) Department of Medical Sciences, Molecular Medicine, Uppsala University, Uppsala, Sweden; 10) Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano & Centro Cardiologico Monzino, IRCCS, Milan, Italy; 11) Department of Medical Sciences, Cardiovascular Epidemiology and EpiHealth, Uppsala University, Akademiska sjukhuset, Uppsala, Sweden.

Large-scale genome-wide association studies (GWAS) have so far identified 45 loci that are robustly associated with coronary heart disease (CHD) in data from adult men and women of European descent. We examined whether these loci are also associated with measures of atherosclerosis in data from up to 9,582 individuals of European ancestry. Forty-five SNPs representing the CHD-associated loci were genotyped in middle-aged to elderly individuals of European descent from four independent population-based studies (IMPROVE, MDC-CC, ULSAM and PIVUS). Intima-media thickness (IMT) was measured by external B-mode ultrasonography at the far wall of the bulb and common carotid artery. Plaque presence was defined as a maximal IMT of the bulb >1.5 mm. We meta-analysed single-SNP associations across the four studies, and combined them in a genetic predisposition score. We subsequently examined the association of the genetic predisposition score with prevalent CHD and the three indices of atherosclerosis, adjusting for sex, age and Framingham risk factors. As anticipated, the genetic predisposition score was associated with prevalent CHD, with each additional risk allele increasing the odds of disease by 5.5% ($p=4.1\times10^{-6}$). Moreover, each additional CHD-risk allele across the 45 loci was associated with a 0.24% increase in IMT at the far wall of the bulb ($p=4.0\times10^{-3}$), and with a 2.8% increased odds of plaque presence ($p=7.4\times10^{-6}$), independently of traditional risk factors. The genetic predisposition score was not associated with IMT of the common carotid artery (p=0.47). Our results suggest that the association between the 45 previously identified loci and CHD at least partly acts through atherosclerosis.

2053S

Transethnic replication of the gene-gene interaction of the prostaglandin E_2 system in determining blood pressure reactivity. *X. Kong[1,2], Q. Zhao[1], TN. Kelly[1], C. Li[1], D. Gu[3], J. He[1].* 1) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, US; 2) Department of Endocrinology, China-Japan Friendship Hospital, Beijing, China; 3) State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

Increased blood pressure (BP) reactivity to cold stress has been suggested as a risk factor for hypertension and cardiovascular disease. Prostaglandin E_2 (PGE_2), the most abundant prostanoid in humans, plays a critical role in BP regulation. This study aimed to systematically investigate individual and interaction effects of genetic variants of the PGE_2 system on BP reactivity to cold stress in both Chinese and Whites. The cold pressor test (CPT) was performed among 1,881 Han Chinese individuals from the Genetic Epidemiology Network of Salt Sensitivity (GenSalt) study. BP reactivity variables were calculated as the changes between the highest BP levels during the CPT and those prior to the CPT. A total of 397 single nucleotide polymorphisms (SNPs) covering 10 genes from the PGE_2 system, including synthases (*PLA2G4A*, *PTGS1*, *PTGS2*, *PTGES*, *PTGES2*, *PTGES3*) and receptors (*PTGER1*, *PTGER2*, *PTGER3*, *PTGER4*) of the PGE_2, were genotyped for the analysis. In addition to single marker analyses, gene-based and gene-gene interaction analyses were conducted using the truncated *P* value method. Results were further replicated in 1,448 participants of European ancestry from the Coronary Artery Risk Development in Young Adults (CARDIA) study, in which a similar CPT was conducted. SNPs from the *PTGER3* and *PTGER4* genes were significantly associated with BP reactivity variables of diastolic BP (DBP) or mean arterial pressure (MAP) in both Han Chinese and Whites. In gene-based analyses, *PTGER4* was associated with DBP and MAP responses in both Han Chinese (both $P < 1.0\times10^{-6}$) and Whites ($P = 0.029$ and 0.038). In addition, gene-gene interactions were identified between several pairs of genes from the PGE_2 system. For example, *PLA2G4A* and *PTGES2* genes showed the strongest interaction for systolic BP and MAP responses in both Han Chinese (both $P < 1.0\times10^{-5}$) and Whites (both $P < 1.0\times10^{-5}$). Transethnic replication of genetic associations identified in this study may suggest an important role of genetic variation of the PGE_2 system in affecting BP reactivity to stress.

2054M

Role of common sarcomeric gene polymorphisms in genetic susceptibility to left ventricular dysfunction. *B. Mittal[1], S. Kumar[1], A. Mishra[1], A. Srivastava[1], N. Garg[2], S. Agarwal[3], S. Pande[3].* 1) Genetics, SGPGIMS, Lucknow, U.P., India; 2) Cardiology, SGPGIMS, Lucknow, UP, India; 3) CVTS, SGPGIMS, Lucknow, U.P, India.

Background: Mutations in sarcomeric genes are common genetic cause of cardiomyopathies. An intronic 25-bp deletion in MYBPC3 at 3' region is associated with dilated (DCM) and hypertrophic (HCM) cardiomyopathies in Southeast Asia. However, the frequency of sarcomeric gene polymorphisms and associated clinical presentation has not been established with left ventricular dysfunction (LVD). Therefore, the aim of the present study was to explore the association of MYBPC3 25 bp deletion, TTN 18 bp I/D, TNNT2 5bp I/D, and Myospryn K2906N polymorphisms with LVD. Methods: The study included 988 consecutive patients with angiographically confirmed CAD and 300 healthy controls. Among 988 CAD patients, 253 with reduced left ventricle ejection fraction (LVEF≤45%) were categorized as LVD. MYBPC3 25 bp deletion, TTN 18 bp I/D and TNNT2 5bp I/D polymorphisms were determined by direct polymerase chain reaction (PCR) method while Myospryn K2906N polymorphism by TaqMan assay. Results: Our results showed that MYBPC3 25 bp deletion polymorphism was significantly associated with elevated risk of LVD (healthy controls v/s LVD: OR=3.85, p-value<0.001; and Non-LVD v/s LVD: OR=1.65, p-value=0.035) while TTN 18 bp I/D, TNNT2 5bp I/D, and Myospryn K2906N polymorphisms did not show any significant association with LVD. The results also showed that MYBPC3 25 bp deletion polymorphism was significantly associated with other parameters of LV remodeling i.e. LV dimensions (LV end diastole dimension: LVEDD; p-value=0.037, and LV end systolic dimension: LVESD; p-value=0.032). Conclusion: Our data suggests that MYBPC3 25 bp deletion may play significant role in conferring LVD risk in Southeast Asian populations. Financial support from DBT, Government of India.

2055S

Identification of TMEM241 as the underlying gene in the chromosome 18q11.2 triglyceride region in Mexicans. *A. Rodriguez[1], L. Gonzalez[1], Y. Bhagat[1], E. Nikkola[1], T. Tusie-Luna[3,4], C.A Aguilar-Salinas[3], P. Pajukanta[1,2].* 1) Human Genetics, UCLA, Los Angeles, CA; 2) Molecular Biology Institute at UCLA, Los Angeles, USA; 3) Instituto Nacional de Ciencias Médicas y Nutrición, Salvador Zubiran, Mexico City, Mexico; 4) Instituto de Investigaciones Biomédicas de la UNAM, Mexico City, Mexico.

High serum triglyceride (TG) levels are a critical risk factor for cardiovascular disease (CVD). We recently identified a locus on chromosome 18q11.2 to be associated with high serum TGs in Mexicans in a genome-wide association study (GWAS). However, the biological mechanism(s) underlying this signal has not been elucidated yet. Characterization of GWAS hits is an important area of research in the field of functional genomics. The 18q11.2 locus in particular could have implications in the treatment and diagnosis of CVD in the rapidly growing Hispanic populations. To characterize the biological mechanism, we performed a cis-eQTL analysis utilizing 856 publicly available human adipose, skin, and lymphocyte RNA microarrays of the MuTHER resource, and discovered that the lead SNP rs9949617 is a genome-wide significant cis-eQTL, regulating the expression of one of the 5 regional genes on 18q11.2, the transmembrane protein 241 (TMEM241) gene. However, lead GWAS SNPs are often not the underlying causal variant. Therefore, to identify the causal variant, we performed a regional linkage disequilibrium (LD) analysis using PLINK and found a total of 9 single nucleotide polymorphisms (SNPs) in high LD (R2>0.8) with rs9949617, suggesting that anyone of these SNPs can be the underlying susceptibility variant. To prioritize the variants for functional analysis, we filtered the variants through systematic data mining including ENCODE ChIP-seq data and found that rs17259126 resides in a predicted nuclear receptor subfamily 2 (NR2F2), transcription factor binding site and an enhancer region identified by a co-occurrence of H3K27ac and H3K4me1. NR2F2 has previously been associated with lipid metabolism in humans. Therefore, we hypothesize that rs17259126 resides in an enhancer and that G/A alleles result in a differential TMEM241 expression, ultimately leading to the TG phenotype. To prove our hypothesis, we cloned a 500-bp sequence surrounding the SNP upstream of a minimal promoter in the pGL4.23 vector (Promega) and transferred the firefly/Renilla reporters via Lipofection into human hepatocellular carcinoma cells (HepG2). Luciferase assays at 48 hrs post transfection confirmed that the A allele in rs17259126 has an increased reporter expression ($p<0.05$) compared to the reference G allele in 3 biological replicates. These results suggest that TMEM241 is the potential candidate gene underlying the TG GWAS signal on chromosome 18q11.2 in Amerindian origin populations.

2056M

Contribution of genetic variation of ATP-binding cassette transporter A1 (*ABCA1*) to the regulation of plasma lipid/lipoprotein levels in US Non-Hispanic Whites. *F.Y. Demirci[1], V. Niemsiri[1], X. Wang[1], M.M. Barmada[1], J.E. Hokanson[2], R.F. Hamman[2], M.I. Kamboh[1].* 1) Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO.

Abnormalities in plasma lipid/lipoprotein levels have been linked to the risk of coronary heart disease (CHD). *ABCA1* (ATP-binding cassette, subfamily A (ABC1), member 1) encodes a member of the ATP-binding cassette transporters superfamily that functions as a cholesterol efflux pump and *ABCA1* mutations are known to cause hereditary high-density lipoprotein (HDL) deficiency. In this study, we sequenced *ABCA1* (exons, most introns, and ~1 kb of each of 5' and 3' flanking regions) in 95 US Non-Hispanic Whites (NHWs) with extreme HDL cholesterol (HDL-C) levels using the Sanger method and identified 404 variants (402 bi-allelic and 2 tri-allelic). A total of 237 bi-allelic variants, including those identified by sequencing (common tagSNPs with MAF≥5% and selected uncommon/rare variants) plus additional tagSNPs from the HapMap project, were genotyped in our entire NHW sample comprising 623 individuals. The resulting quality control (QC)-passed genotyping data (including 183 variants; 117 common and 66 uncommon/rare) were then evaluated for associations with 5 major lipid/lipoprotein traits (HDL-C, low-density lipoprotein cholesterol (LDL-C), triglyceride (TG), apolipoprotein (apo) A1, and apoB levels). In single-site analysis, 35 common variants showed significant association ($P<0.05$) with at least one lipid/lipoprotein trait. Although significant associations were observed with all interrogated lipid/lipoprotein traits (for at least 2 variants per trait), the strongest and highest number of associations (21 of 35 significant SNPs) were detected with TG levels. The rare variant and haplotype analyses are currently underway. Our preliminary findings suggest a role for common *ABCA1* variants in modulating the lipid/lipoprotein levels (and especially the TG levels), which is in line with the recent studies implicating *ABCA1* in the metabolism of various major plasma lipoproteins as a contributor to the metabolic link between them.

2057S

A high yield of variants with a putative role as modifiers in patients with hypertrophic cardiomyopathy. *S. Bardi[1], F. Girolami[1], M. Benelli[1], B. Tomberli[2,3], E. Contini[1], G. Marseglia[1], C. Pescucci[1], G. Castelli[2], A. Fornaro[2], F. Cecchi[3], I. Olivotto[2], F. Torricelli[1].* 1) Diagnostic Genetic Unit, Careggi University Hospital, Florence, Italy; 2) Referral Centre for Myocardial Diseases, Careggi University Hospital, Florence, Italy; 3) University of Florence, Department of Clinical and Experimental Medicine, Florence, Italy.

Next Generation Sequencing enables simultaneous screening of multiple genes for multiple patients in a single run. We designed a panel of 111 genes known to be associated to CMs to study 94 unrelated patients (80 with Hypertrophic Cardiomyopathy, HCM; 18 with Dilatative Cardiomyopathy, DCM and 6 with Arrhythmogenic Cardiomyopathy, AC). Targeted resequencing was performed on Illumina platform (98,13% of the regions with a depth of coverage of 20X or more, mean coverage on target of 530X). A mean of 1016 variants were found for each patient. Rare (frequency <0.05), non-synonymous, loss- of- function and splice-site variants were defined as candidates. Pathogenic or likely-pathogenic variants were all confirmed by Sanger and cosegregation was tested when possible. Excluding titin missense variants, we identified 48 variants (27 novel) in sarcomeric or associated genes in 48/70 HCM patients (68%), with 14% of complex genotype. MYH7, MYBPC3 and TNNI3 resulted the high-yield genes; 19 additional candidate variants (13 novel) in desmosomal and ion-channel genes in 14 patients (20%) were identified in this group. We identified 10 candidate variants (7 novel) in 7/18 DCM patients (39%) and 5 candidate variants in 3/6 AC patients (50%). A targeted protocol allowed the identification of likely pathogenic variants in a large proportion of patients with CMs, irrespective of phenotype. The unexpected finding of rare non synonymous variants in desmosomal and ion-channel genes among HCM patients raises important issues regarding their role as previously unappreciated modifiers of the disease, potentially relevant to risk prediction and counseling.

2058M

Genetic dissection of a novel X-linked congenital heart syndrome. *C. Preuss[1], S. Yang[2], M. Samuels[3], P. Awadalla[3], P. Chetaille[4], H. Björck[5], S. Mohamed[6], P. Eriksson[5], G. Andelfinger[1].* 1) Department of Pediatrics, Centre Hospitalier Universitaire Sainte Justine, Montreal , Quebec, Canada; 2) Department of Cardiology, Nanjing Children's Hospital, Nanjing Medical University, Nanjing, China; 3) Department of Pediatrics, Faculty of Medicine, Sainte-Justine Research Center, University of Montreal, Montreal, QC, Canada; 4) Cardiology Service, Centre Mère-Enfants, Centre Hospitalier Universitaire de Québec, Université de Laval, Québec City, Québec, Canada; 5) Atherosclerosis Research Unit, Center for Molecular Medicine, Department of Medicine, Karolinska Institutet, Stockholm, Sweden; 6) Department of Cardio and Thoracic Vascular Surgery, University Clinic of Schleswig-Holstein, Luebeck, Germany.

The major burden of congenital heart disease (CHD) is caused by left ventricular outflow tract obstructions (LVOTO). This heterogeneous group of cardiac malformations constitutes to an increased risk of valve replacements and aortic aneurysms. Although a strong male predominance has been described for the highly heritable trait, the genetic basis for these cardiac malformations remains poorly understood. Here, we describe a comprehensive genetic study in two French Canadian pedigrees with 15 affected male individuals with septal defects, aortic valve lesions and bicuspid aortic valve. Linkage analysis and genetic fine mapping using the Illumina Omni 5.0 platform revealed a significant interval on chromosome Xq28 (LOD score = 3.29) harboring a 260kb haplotype co-segregating with disease in both families. The disease associated haplotype is rare (MAF < 5%) among 960 genotyped French Canadian controls. Whole-exome sequencing of six patients and re-sequencing of the disease haplotype did not reveal rare (MAF < 2%) coding or splice-site mutations. Validation of all identified rare and novel mutations in the disease haplotype using a Sequenom panel confirmed only one mutation co-segregating with disease in an X-linked manner. This highly conserved intronic mutation (GERP score > 4) in *FLNA* is absent among public (dbSNP138, 1000 Genomes) and in-house sequencing data sets. In silico analysis predicts the hemizygous mutation to create a high affinity MyoD binding site. Expression data from patients with bicuspid aortic valve revealed that *FLNA* shows the highest expression patterns in the aortic media of dilated aortas compared to all expressed transcripts in the disease associated haplotype. Reconstruction of ascending genealogies supports the notion of a founder effect for this novel X-linked syndrome, dating back to a common founder couple in 1788.

2059S

Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *A. Ganna[1], S. Salihovic[2], J. Sundström[2], C.D. Broeckling[3], A.K. Hedman[4], P.K.E. Magnusson[1], N.L. Pedersen[1], A. Larsson[5], A. Siegbahn[6], M. Zilmer[7], J. Prenni[3,8], J. Ärnlöv[4,9], L. Lind[2], T. Fall[4], E. Ingelsson[4,10].* 1) Medical epidemiology and biostatistics, Karolinska institutet, Stockholm, Sweden; 2) Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden; 3) Proteomics and Metabolomics Facility, Colorado State University, Fort Collins, Colorado, USA; 4) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 5) Department of Medical Sciences, Biochemial structure and function, Uppsala University, Uppsala, Sweden; 6) Department of Medical Sciences, Coagulation and inflammation science, Uppsala University, Uppsala, Sweden; 7) Department of Biochemistry, The Centre of Excellence for Translational Medicine, University of Tartu, Tartu, Estonia; 8) Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO, USA; 9) School of Health and Social Studies, Dalarna University, Falun, Sweden; 10) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, United Kingdom.

Background: Analyses of circulating metabolites in large prospective epidemiological studies could lead to improved prediction and better biological understanding of coronary heart disease (CHD).

Methods and Results: We performed a mass spectrometry-based non-targeted metabolomics study for association with incident CHD events in 1,028 individuals (131 events; 10 y. median follow-up) with validation in 1,670 individuals (231 events; 4 y. median follow-up). Four metabolites were robustly replicated and independent of established risk factors [lysophosphatidylcholine 18:1 (hazard ratio [HR] per standard deviation [SD] increment= 0.77, P-value<0.001), lysophosphatidylcholine 18:2 (HR=0.82, P-value= 0.002), monoglyceride 18:2 (MG 18:2; HR=1.16, P-value=0.032) and sphingomyelin 28:1 (HR=0.86, P-value=0.010)]. Together they contributed to moderate improvements in discrimination and re-classification in addition to traditional risk factors (C-statistic: 0.76 vs. 0.75; NRI: 9.2%). MG 18:2 was associated with CHD independently of triglycerides. Lysophosphatidylcholines were negatively associated with BMI, glucose, C-reactive protein and with less evidence of subclinical cardiovascular disease in 970 PIVUS participants; a reverse pattern was observed for MG 18:2. MG 18:2 showed an enrichment (P-value=0.002) of significant associations with CHD-associated SNPs (P-value=1.2×10^{-7} for association with rs964184 in the ZNF259/APOA5 region) and a weak, but positive causal effect (odds ratio=1.05 per SD increment in MG 18:2, P-value=0.05) on CHD, as suggested by Mendelian randomization analysis.

Conclusions: We identified four lipid-related metabolites with evidence for clinical utility, as well as a causal role in CHD development.

2060M

Genetic and Metabolic Causes of Neonatal Cardiomyopathy. *C. Prada[1,2], I. Villamizar-Schiller[1], J. Castro[3], L. Pabon[1], A. Duran[3].* 1) Centro de Medicina Genómica y Metabolismo, Cardiovascular Foundation of Colombia, Floridablanca, Colombia; 2) Dept Pediatrics Genetics, Cincinnati Children's Hosp Med Ctr, Cincinnati, OH; 3) Division de Cardiologia Pediatrica, Cardiovascular Foundation of Colombia, Floridablanca, Colombia.

Background: Neonatal cardiomyopathy is an important cause of pediatric cardiac deaths. This is a heterogeneous disease with a strong genetic component. The objective of this study was to determine the prevalence of genetic and metabolic etiologies in patients with neonatal cardiomyopathy. Study Design: A retrospective analysis of data from clinical records of 40 neonates with cardiomyopathy seen at the Cardiovascular Foundation of Colombia between 2011 and 2014 was performed. Results: In a total of 29 neonates (72.5%) a genetic or metabolic etiology was identified. The most common group of causes was metabolic (13/29, 45%) followed by syndromic (10/29, 34%), and maternal diabetes (6/29, 21%). The single most common diagnosis were RASopathies (7/25, 28%) followed by maternal diabetes, mitochondrial and fatty acid oxidation defects. Mortality was higher in the group of metabolic disorders (7/13, 54%). Hypertrophic cardiomyopathy was identified in 25 of the 40 patients (62.5%). Sixty percent of patients with cardiomyopathy (24/40) survived beyond their first year of life. Extra-cardiac manifestations were more prevalent in the syndromic group. Family history of a sibling with unexplained death was seen in the metabolic group. Conclusion: Neonatal cardiomyopathy has a strong genetic component in the majority of patients. An underlying metabolic or syndromic cause was identified in 72.5% of neonates. Identification of etiology is important for management and family counseling and recurrence.

2061S

Targeted Oligonucleotide-Selective Sequencing for Genetic Diagnostics of Pulmonary Arterial Hypertension. *E.H. Seppälä[1], S. Vattulainen[2], J. Aho[3], J. Tallila[1], M. Gentile[1], M. Sankelo[4], T. Laitinen[5], J.W. Koskenvuo[1,3], T.-P. Alastalo[1,2], S. Myllykangas[1,6].* 1) Blueprint Genetics, Helsinki, Finland; 2) Childrens Hospital Helsinki, University of Helsinki, Helsinki, Finland; 3) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Finland; 4) Department of Internal Medicine, Tampere University Hospital, Tampere, Finland; 5) Division of Medicine, Department of Pulmonary Diseases and Allergology, Turku University Hospital and University of Turku, Turku, Finland; 6) Institute of Biomedicine, University of Helsinki, Helsinki, Finland.

Pulmonary arterial hypertension (PAH) is a severe and progressive disease. The genetic basis of idiopathic PAH is well recognized but rarely utilized in a diagnostic setting. Hundreds of mutations in seven genes have been reported to associate with idiopathic and familial forms of PAH. Next-generation sequencing for comprehensive genetic testing improves diagnostics, prognostics and treatment optimization of the index patient and allows an effective screening of asymptomatic family members. We utilized novel next-generation sequencing and bioinformatics approaches for genetic diagnostics of PAH. We applied Oligonucleotide-Selective Sequencing (OS-Seq) and custom data analysis and interpretation pipelines to identify pathogenic base substitutions, insertions and deletions in seven genes associated with PAH (BMPR2, BMPR1B, ACVRL1, ENG, SMAD9, CAV1 and exon 2 of KCNK3). Targeted sequencing covered all coding exons, exon-intron boundaries and known intronic mutations (12,638 bases). In OS-Seq, oligonucleotide-functionalized Illumina flow cells are used for both capture and sequencing of DNA. We automated target DNA capture and sequencing using the MiSeq Sequencing system. We analyzed 21 Finnish PAH patients. Sequencing depth was evaluated in each target base to demonstrate the efficiency of targeted sequencing. Average sequencing depth was 791 and 99.88% of target regions were covered >15x. We identified six idiopathic PAH patients with likely pathogenic variants in the BMPR2 gene. The identified mutations caused stop, frameshift and missense changes and had minor allele frequency of <0.001 reported in the 1000 genomes database. Mutations or potential variants were not identified from 6 other analyzed genes. Our results show that a comprehensive next-generation sequencing panel is an effective tool for genetic diagnostics of PAH and 29% of the Finnish PAH patients carried mutations in BMPR2.

2062M

Use of a Gene Expression Score in a Primary Care Setting to Evaluate African American Patients Presenting with Symptoms Suggestive of Obstructive Coronary Artery Disease. *L. Wilson[1], M. Brown[2], D. Smith[3], B. Rhees[1].* 1) CardioDx, Inc., Redwood City, CA; 2) Providence - Dayton Primary Care, Dayton, OH; 3) Novant Heart and Vascular Institute, Huntersville, NC.

Purpose: Approximately 3 million pts without diabetes present annually to primary care clinicians with symptoms suggestive of obstructive coronary artery disease (CAD). After a detailed examination, physicians still rely heavily on advanced imaging diagnostic tests, to determine the etiology of symptoms. Given the heterogeneity in clinical manifestations of CAD amongst different ethnic populations in the US, the use of a gene expression score (GES) may allow clinicians to more consistently evaluate across the spectrum of primary care pts, thereby appropriately avoiding unnecessary referrals and advanced diagnostic tests. We hypothesized that use of the GES would show clinical utility for a clinician evaluating African Americans pts presenting with symptoms suggestive of CAD. Methods: A previously validated gene expression diagnostic test (Corus® CAD, CardioDx, Inc.) has a 96% NPV in ruling out obstructive CAD among symptomatic pts with no previous history of diabetes or myocardial infarction. GES results are predefined as low (GES ≤15) or elevated (GES >15), with low score pts having a low likelihood of obstructive CAD. Previous evaluation showed no significant difference in test performance between non-white and white populations. This was a single primary care practice study with a large African American pt population. De-identified pt data from 2011-2013 was collected. Results: This cohort included 582 African American pts who received a GES, with 325 (56%) female pts and 201 (35%) being >65yrs. Approximately 90% (518/582) of pts presented with typical or atypical symptoms suggestive of obstructive CAD. Mean GES was 19 (range, 1-40), and 245 pts (42%) had low scores. In this analysis, 11/245 (5%) of low GES pts were referred to cardiology and/or further diagnostic testing, whereas 248/337 (74%) of elevated GES pts were referred to cardiology and/or further diagnostic testing (p-value<0.0001). Conclusions: The personalized GES showed clinical utility in the evaluation of symptomatic African American pts. The findings show fewer cardiology referrals and/or further diagnostic testing for those African American pts with low GES, thereby improving pt care by avoiding unnecessary, and potentially harmful, advanced diagnostic testing. This low rate of referral among low GES pts is similar to that observed in largely non-minority pt cohorts from previous GES studies: COMPASS, IMPACT-PCP, IMPACT-Cardiology, Registry I, and the PRESET Registry.

2063S

Extended genetic diagnosis of Familial Hypercholesterolemia using next-generation sequencing. *M.M. Motazacker[1], B. Sjouke[2], O.R.F. Mook[1], M.A. Haagmans[1], G.K. Hovingh[2], J.C. Defesche[3], A.R. Mensenkamp[4], M.M.A.M. Mannens[1].* 1) Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands; 2) Department of Vascular Medicine, Academic Medical Center, Amsterdam, The Netherlands; 3) Department of Experimental Vascular Medicine, Academic Medical Center, Amsterdam, The Netherlands; 4) Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands.

Familial Hypercholesterolemia (FH) is a major risk factor for coronary artery disease. FH is caused by mutations in the genes coding for the low-density lipoprotein receptor (*LDLR*), apolipoprotein B (*APOB*) and proprotein convertase subtilisin/kexin 9 (*PCSK9*). Routine genetic diagnosis of FH is often limited to sequencing *LDLR* followed by partial sequencing of *APOB* and *PCSK9* in cases with no *LDLR* mutations. This is mainly due to the large size of *APOB* and rarity of *PCSK9* mutations which makes Sanger sequencing inefficient. Using DNA from 20 patients with 31 known mutations (including single-nucleotide coding and promoter variants, insertions, deletions, indels and large copy-number variants), we employed an Ion AmpliSeq™ enrichment of coding sequence (and 25 bp flanking intronic regions) of all three FH genes followed by sequencing using Ion-PGM™ sequencing platform (Life technologies™). The sequence data were then analyzed using SeqNext software (JSI medical systems). We could successfully identify all previously detected mutations. Interestingly, we also identified additional 8 rare variants including *PCSK9* p.Cys679X and 7 *APOB* variants (e.g. p.Arg532Trp, p.Asp1113His, p.Lys3076Met, etc.), some of which already reported in the literature to be functional while others predicted to be functional using *in silico* prediction models. Identification of these extra variants may help in explaining unexpected phenotypes (which are usually justified by incomplete penetrance of the phenotype) seen frequently in the families with dyslipidemia. Our study suggests a fast, cost-effective and accurate approach for extended genetic diagnosis of FH which can increase the yield of FH diagnosis and improve phenotype-genotype correlation studies in pedigrees with phenocopy or non-penetrant FH mutations.

2064M

Evidence for the novel variant, c.1937 C>T (p.Ser646Phe) in the membrane binding domain of the *ANK2* gene contributing to Long QT syndrome in a First Nations community of Northern British Columbia. *L.T Arbour[1,2,3], A.K.J Boyce[2], J. Christensen[2], S. MacIntosh[1], S. Lauson[3], S. Tung[4], C. Kerr[4], L.A Swayne[2].* 1) Department of Medical Genetics, University of British Columbia, Victoria, BC; 2) Division of Medical Sciences, University of Victoria, Victoria, BC; 3) Island Health and the BC Inherited Arrhythmia Program, Victoria BC; 4) Department of Medicine, University of British Columbia, Vancouver, BC Canada.

Long QT syndrome (LQTS) is a rare inherited cardiac condition named for a prolonged QT interval on ECG (corrected for rate) conferring susceptibility for life-threatening ventricular arrhythmias. An estimated 75 percent of hereditary LQTS is caused by mutations in 12 genes that encode ion channels or functionally related proteins. In Northern British Columbia, through participatory methods, a First Nations community was previously identified to have a disproportionately high rate of LQTS, accounted for largely by a novel missense founder mutation (p.V205M) in the *KCNQ1* gene causing LQTS1. Mechanistic and clinical evaluation supported pathogenesis and studies on the variable phenotype and natural history are on-going. Although the majority of cases in the community have been confirmed with the founder mutation, there are some families with LQTS who are mutation negative. We report on two distinct families within this First Nations community where clinical 12 gene sequencing on 2 apparently unrelated probands revealed the same likely pathogenic variant (c.1937 C>T) in exon 18 of the *ANK2* gene (NM_001148.4) resulting in a serine to phenalanine substitution at position 646 in the membrane binding domain. This variant was undetected in ~6500 controls in the NHLBI exome sequencing project and noted to be conserved across species. Nine variant carriers from the two families have been identified, with a range in peak QTc (QT interval corrected for rate on ECG) from 426-509 ms. Five of the 9 with the variant have a peak QTc > 470 ms. A sudden death at age 50 in an obligate carrier and a recent sudden death in a 36 year old male relative at high risk for the variant exemplify the family implications. The protein encoded by the *ANK2* gene, Ankyrin2, acts as a scaffold for ion channels and transporters critical in cardiomyocyte excitability. Although mutations in the Ankyrin2 C-terminal domain have been reported to cause LQTS (type 4), this is the first mutation identified in its membrane binding domain. To determine the effects of the Ser646Phe mutation on Ankyrin2 localization and cellular metrics (such as cell size, shape proliferation, and viability), cardiomyoblasts were transfected with GFP-tagged wildtype and Ser646Phe variant Ankyrin2 for confocal microscopy and live cell kinetic imaging experiments. Our findings support pathogenesis for this uniquely situated mutation in the *ANK2* gene.

2065S

Novel variants in *VINCULIN* and *TROPOMYOSIN1* combinatorially predispose patients to dilated cardiomyopathy. *D.C. Deacon[1,2,3,5], A.M. Manso[1,4], B.C. Nelson[1,5], R.S. Ross[1,2,4], E.D. Adler[1], N.C. Chi[1,2,3].* 1) School of Medicine, University of California San Diego, La Jolla, CA; 2) Biomedical Sciences Graduate Program, UCSD; 3) Medical Scientist Training Program, UCSD; 4) Veterans Affairs San Diego Healthcare System; 5) California Institute of Regenerative Medicine Fellow.

Dilated cardiomyopathy (DCM) may affect as many as 1 in 250 individuals and is the leading cause of heart failure necessitating heart transplant. We have identified a large family presenting with multiple cases of DCM across four generations. The proband for this study, a 14-year old male patient, presented with severe heart failure and required heart transplantation. His father was also diagnosed with DCM while his sister, who had a normal heart by echocardiography, died suddenly and, on autopsy, showed mild cardiac hypertrophy. We have identified sequence variants in the genes encoding the costameric protein VINCULIN (VCL) and the sarcomeric regulatory protein TROPOMYOSIN 1 (TPM1) in this family. Targeted sequencing at these loci in 31 family members showed that the combination of *VCL* and *TPM1* variants cosegregated with all family members diagnosed with cardiomyopathy. Given the genetic cosegregation of these novel heterozygous variants in *VCL* and *TPM1*, we hypothesize that the functional interaction between costameric and sarcomeric structures is critical to the proper regulation of cardiomyocyte force generation and transmission. In order to study the effects of these variants on cardiac structure and function, we have utilized CRISPR-mediated genomic editing to introduce these patient-specific variants into a mouse model. Both of these genes are over 99% conserved between human and mouse at the protein level and mutations in both have been previously associated with cardiomyopathies, though the molecular mechanisms governing these pathophysiologies have not been fully described. We will determine the consequences of these variants on cardiac contractility and sarcomeric and cytoskeletal organization in single and double variant heterozygous mice to elucidate the mechanisms underlying the observed combinatorial disease inheritance pattern. A further understanding of the genetic and molecular interactions between costameric and sarcomeric proteins in proper cardiac function could pave the way for new therapies to treat heart failure.

2066M

A novel mutation in the *RYR2* gene (c.527 G>T, p.R176L) identified in a 4 generation family presents with a catecholaminergic polymorphic ventricular tachycardia (CPVT) phenotype with variable penetrance. *S. Lauson[1,2], B. Sinclair[2,3], A.A. Collier[4,5], F. Curtis[4,5], F. Van Petegem[6], B.A. Fernandez[4,5], A.E. Williams[7], S. Connors[7], C.G. Templeton[7], K. Hodgkinson[5,9], R. Leather[2,8], S. Sanatani[2,10], L. Arbour[1, 2, 11].* 1) Medical Genetics, Island Health, Victoria, BC, Canada; 2) BC Inherited Arrhythmia Program, Victoria/Vancouver, BC; 3) Pediatric Cardiology, Island Health, Victoria, BC; 4) Provincial Medical Genetics Program, Eastern Health, St. John's, NL; 5) Discipline of Genetics and Medicine, Memorial University of Newfoundland, St. John's, NL; 6) Department of Biochemistry, University of British Columbia, Vancouver, BC; 7) Division of Cardiology, Discipline of Medicine, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL; 8) Division of Cardiology, Island Health, Western Cardiology, Victoria, BC; 9) Clinical Epidemiology, Discipline of Medicine, Memorial University, St. John's, NL; 10) Department of Pediatrics, Children & Women's Hospital, University of British Columbia, Vancouver, BC; 11) Department of Medical Genetics, University of British Columbia, Vancouver BC.

Catecholaminergic Polymorphic Ventricular Tachycardia (CPVT) is a rare condition with population incidence estimated to be 1:10,000 and confers substantive risk for sudden cardiac death in the young. Four known genes cause the condition (*RYR2, CASQ2, TRDN*, and *CALM1*) but the identification of large extended families are rare. Stress and exercise-induced releases of catecholamines can lead to aberrant calcium ion flux, and the hallmark feature of bidirectional ventricular tachycardia (VT). Other inherited arrhythmia conditions can mimic the presentation, including Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC) and Long QT syndrome (LQTS) with overlapping implicated genes. We present a large Newfoundland family presenting with a significant history of cardiac arrests. The proband had two previous cardiac arrests (age 28 and 32) one while undergoing surgery. She was initially investigated and tested for ARVC due to the common regional founder gene mutation, *TMEM43* (p.S358L). Subsequent comprehensive testing for ARVC included the *RYR2* gene, revealing a novel *RYR2* gene variant (c.527 G>T, p.R176L). Confirmation of bidirectional VT in her son suggested the CPVT phenotype. He had no evidence of ARVC. Of interest, this family also harbours a known pathogenic LQTS gene mutation (R518Q in *KCNQ1*), and two siblings are heterozygous for both. The R176L variant was not seen in 600 alleles from controls, and results in a non-conservative amino acid substitution of a polar Arginine with a non-polar Leucine at a highly-conserved region. Crystallographic studies show that R176 is involved in interactions with a neighboring RyR2 subunit. These interactions stabilize the RyR2 closed state, and any weakening through R176L would result in facilitated channel opening, leading to premature or prolonged release of calcium ions. Currently 16 individuals in 4 generations have been identified with the mutation. Cardiac arrest during surgery (age 29), was documented in a variant carrier (cousin to the proband), and bidirectional VT has been confirmed in another during exercise testing supporting clinically that the phenotype segregates with the mutation alone. However, the carrier father of the proband (age 76) remains apparently unaffected. ECGs, echocardiograms, Holter monitors and stress tests are being carried out on family members. This family provides a unique opportunity to explore the variable penetrance and expression of an *RYR2* mutation causing CPVT.

2067S

Harnessing genomic data to identify drug targets for reduction of LDL cholesterol and CAD risk that do not impact upon glycaemic status. *V. Tragante do O[1], F.W. Asselbergs[1], M.V. Holmes[2].* 1) Divisie Hart & Longen, University Medical Center Utrecht, Heidelberglaan 100, 3584CX Utrecht, Netherlands; 2) Division of Transplant Surgery, Perelman School of Medicine, University of Pennsylvania, 3400 Spruce St, Philadelphia, PA 19104, USA.

Introduction LDL cholesterol (LDL-C) reduction is recognized as an efficacious therapeutic means to reduce risk of coronary artery disease (CAD). However, the most widely prescribed LDL-reducing drugs (statins) are associated with an increased risk of type 2 diabetes (T2D). We used data from genome-wide association studies for CAD and glycaemic traits to identify potential drug targets for LDL-C reduction that have no impact on glycaemic status. Methods and Results We obtained summary-level results from a GWAS on LDL cholesterol from Global Lipids Genetics Consortium (GLGC), CAD/MI from CARDIoGRAMplusC4D, T2D from DIAGRAM consortium and glucose, insulin and related traits from MAGIC consortium, and standardized all effects to 1 unit of increase in LDL-C. A p-value filter of P<5E-08 for GLGC results was employed, and the SNPs surpassing this threshold were further filtered with P<0.05 for CARDIoGRAMplusC4D and DIAGRAM. Directions of effect were the same for GLGC and CARDIoGRAM for 81 out of 84 SNPs passing the significance threshold (binomial P=4.93E-21), whereas all 17 SNPs passing significance thresholds for DIAGRAM and MAGIC had the same direction of effect (binomial P=7.63E-06). We then performed a multiple-trait meta-analysis of all MAGIC subcomponents (referred to as "metabolic burden") and searched for SNPs that are significant for LDL and CAD/MI, but did not associate with metabolic burden (P>0.05). Eighteen loci met these criteria, and were further investigated for their potential druggability. Six of those present interaction with one or more commercially available drugs, including LDLR, HNF1A, OASL, PLG, PTPN11 and SLC22A3. Conclusions These findings provide novel information for prioritizing therapeutic targets for reduction of LDL-C and CAD risk that should be free from adverse consequences on glycaemic status. Keywords Coronary artery disease, type 2 diabetes, LDL cholesterol, drug target.

2068M

From Identification of Differing TIE2 Mutations with Distinct Cellular Characteristics in Four Types of Venous Anomalies towards a Murine Model and a Therapeutic Pilot Study. *M. VIKKULA[1], N. LIMAYE[1], J. SOBLET[1], M. UEBELHOER[1], M. NETYNKI[2], E. BOSCOLO[3], L. EKLUND[2], J. BISCHOFF[3], L.M. BOON[1,4].* 1) Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Brussels, Belgium; 2) Oulu Center for Cell-Matrix Research, Biocenter Oulu and Department of Medical Biochemistry and Molecular Biology, University of Oulu, Finland; 3) Vascular Biology Program and Department of Surgery, Boston Children's Hospital, Harvard Medical School, MA, USA; 4) Center for Vascular Anomalies, Division of Plastic Surgery, Cliniques universitaires Saint Luc, Brussels, Belgium.

Venous anomalies are composed of ectatic veins with irregular smooth muscle coverage. They are commonly cutaneous. They usually occur as a single lesion without family history (sporadic Venous Malformation, VM). Some sporadic patients have multifocal lesions (Multifocal Sporadic Venous Malformation, MSVM). In the sporadic Blue Rubber Bleb Nevus syndrome (BRBN), patients also have multifocal lesions; pathognomonic are rubbery palmoplantar lesions and those located in the GI-track. In rare cases, venous malformations are multifocal because of autosomal dominant inheritance (Mucocutaneous Venous Malformation, VMCM). VMs progressively expand causing deformity, pain and local intravascular coagulopathy. Despite sclerotherapy or excision, lesions often progress or recur. We have identified activating mutations in the endothelial tyrosine kinase receptor TIE2 in all four forms. VMs are mostly due to a single somatic amino acid change L914F. MSVMs and BRBNs are due to double mutations in cis. The BRBN mutations are somatic, whereas MSVM mutations seem mosaic. Moreover, a distinct cis-mutation is seen in MSVM. The inherited VMCM is due to a germline mutation combined with a somatic second-hit. These clinicogenetic entities are reflected by phenotypic differences in cells overexpressing mutant receptors. Activation of AKT is yet a common phenomenon. The capacity to form lesions clearly resides in mutant endothelial cells, which when injected to immunodeficient mice generate lesions mimicking human VM. Interestingly, an mTOR inhibitor is able to deter lesion development. Finally, in our therapeutic pilot study comprising five patients with VMs refractory to standard-of-care, an mTOR inhibitor diminished pain, intravascular coagulopathy and improved quality of life.

2069S

A novel pathway involved in the susceptibility of non-alcoholic fatty liver diseases. *S. Makishima, S. Boonvisut, K. Watanabe, K. Nakayama, S. Iwamoto.* Division of Human Genetics, Center for Molecular M, Jichi Medical University, Shimotsuke, Tochigi, Japan.

Background. *Mammalian tribbles homolog 1* (*TRIB1*) is a locus that has convincing impact on cardiovascular diseases and levels of plasma triglyceride (TG) and LDL-cholesterol across several ethnic groups. The genetic associations have been shown in both genome-wide SNP association studies (GWAS) and the replication studies. In addition, we recently showed a deep association of *TRIB1* SNP with non-alcoholic fatty liver diseases (NAFLD) in Japanese (P = 9.39E-7). The risk allele of NAFLD decreased transactivation activity in reporter gene assay. Furthermore, knockdown of *Trib1* expression in mouse liver increased plasma and hepatic lipid levels, whereas the over expression decreased them. Although the enhanced lipogenesis in mouse liver was estimated to result from the reduced decay of carbohydrate response element binding protein (ChREBP), molecular pathways for the hepatic lipid accumulation were still uncovered. **Methods.** The novel molecular targets of TRIB1 were explored using Yeast two-hybrid system. The positive clones were further screened functionally using shRNA template via adenoviral gene transduction system into mice liver. Plasma and hepatic lipid levels of the mice were measured to identify the genes involved in the hepatic lipid accumulation. **Results and discussion.** Nine cDNA clones were repeatedly identified through the Yeast two-hybrid screening, four of them simultaneously showed molecular interaction with TRIB1 protein in mammalian cells. Knocking-down one of the four genes, *SAP18*, reduced plasma TG levels and increased hepatic lipid accumulation, suggesting impaired lipid secretion. Transcriptome analysis of the mice liver using gene chip revealed significant reduction of the expression of *microsomal triglyceride transfer protein* (*MTTP*), one of the causal gene of abetalipoproteinemia. SAP18 is a subunit of Sin3A-HDAC complex. ChIP studies using halo-chip system and anti-Sin3A showed enrichment of *MTTP* regulatory sequences. These results showed a possible molecular mechanism of NAFLD associated with *TRIB1* expression levels.

2070M

Association of the eNOS -786T > C gene polymorphism and coronary artery disease in Iranian population. *S. Mehrtashfar[1], M. Safarpour[2], A. Esmaeli Khatir[3], A. Ebrahimi[2].* 1) Biology Dept, Guilan University, Rasht, Guilan, Iran; 2) Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 3) Sasan Hospital, Tehran, Iran.

Introduction: Coronary artery disease (CAD) is the most common type of cardiovascular disease and one of the major causes of mortality worldwide. CAD is a complex, multi system disease that is caused by plaque formation along the coronary arteries , which restricts the heart blood supply. This process is called atherosclerosis. CAD is a multifactorial disease which is influenced by both genetic and environmental factors. One of the most important genes involved in cardiac disease is nitric oxide synthase 3 (NOS3) gene which its product synthesize nitric oxide from L-Arginine. The endothelium plays an important role in maintaining vascular tone and blood pressure and that is largely mediated by nitric oxide (NO). Since reduced NO synthesize has been involved in the development of coronary atherosclerosis, polymorphisms of the NOS3 gene can be associated with increased susceptibility to CAD. So, The aims of the present study was to evaluate the possible association between the endothelial nitric oxide synthase (eNOS) gene polymorphism and occurrence of Coronary artery disease. Methods: In this case-control study, one variation with more clinical significance in cardiac disease (T786C) was studied in 50 patients with CAD and 100 healthy individuals as control group. This variation was studied using ARMS-PCR method on extracted DNA from peripheral blood cells. Results: According to the findings, the C allele in T786C variation in patients was significantly higher than control group in Iranian population. (OR=1.67, 95%, P-value= 0.041). Conclusion: This study indicates a possible association between the presence of C allele in T786C and the risk of coronary artery disease in Iranian population and suggests this variation as a potential marker for estimating the risk of CAD.

2071S

Novel association between genetic variation in the platelet endothelial aggregation receptor 1 (*PEAR1*) gene and platelet count is not modified by anti-platelet treatment with clopidogrel and aspirin. *A.S. Fisch, L.M. Bozzi, K.A. Ryan, R.B. Horenstein, J.R. O'Connell, B.D. Mitchell, A.R. Shuldiner, L.M. Yerges-Armstrong, J.P. Lewis.* University of Maryland School of Medicine, Department of Medicine, Division of Endocrinology, Diabetes, and Nutrition, Program in Personalized and Genomic Medicine, Baltimore, MD.

While an increasing body of evidence suggests that *PEAR1*, a recently identified transmembrane receptor on the surface of platelets, influences platelet activation and aggregation upon exposure to anti-platelet agents, this protein has also been implicated in megakaryopoiesis and thrombopoiesis via modulation of the PI3K/PTEN pathways. To validate and extend these findings, we evaluated whether genetic variation in *PEAR1* significantly affects several platelet-related traits including platelet count (PC), mean platelet volume (MPV), and platelet distribution width (PDW) in 687 healthy old order Amish (OOA) participants of the Pharmacogenomics of Anti-Platelet Intervention (PAPI) Study. Platelet phenotypes were measured at baseline and after exposure to 7 days of clopidogrel (75 mg/day) treatment as well as 1 day of dual anti-platelet therapy (DAPT) including both clopidogrel and aspirin (325 mg/day). Genotyping was performed using the Taqman system, and association analyses were conducted under a dominant model using a multivariable linear regression model that simultaneously adjusted for the effects of age, sex, and the relationship structure of the OOA. We observed a significant association between a well-characterized single nucleotide polymorphism in intron 1 (rs12041331) and PC at baseline (β = 12.6 x 10^3, P = 0.013) as well as after clopidogrel (β = 13.2 x 10^3, P = 0.009) and DAPT (β = 13.4 x 10^3, P = 0.008) exposure. In contrast, no evidence of association was observed between rs12041331 and MPV or PDW regardless of time point. Consistent with prior reports, these results indicate that *PEAR1* likely influences the production of platelets, yet does not influence platelet volume or distribution width. Future studies will be required to more precisely understand the mechanisms by which *PEAR1* and its genetic variants influence these platelet measurements, which could ultimately lead to the development of more individualized and effective anti-platelet therapy.

2072M

The Association of the Vanin-1 N131S Variant with Blood Pressure is Mediated by Endoplasmic-Reticulum-Associated Degradation and Loss of Function. *Y.J. Wang[1,2], B.O. Tayo[4], A. Bandyopadhyay[5], H. Wang[1], T. Feng[1], N. Franceschini[6], H. Tang[7], J. Gao[5], Y.J. Sung[9], R.C. Elston[1], S.M. Williams[8], R.S. Cooper[4], T.W. Mu[3], X.F. Zhu[1].* 1) Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Center for Proteomics and Bioinformatics, School of Medicine, Case Western Reserve University, Cleveland, OH; 3) Department of Physiology and Biophysics, School of Medicine, Case Western Reserve University, Cleveland, OH; 4) Department of Public Health Sciences, Loyola University Chicago, Stritch School of Medicine, Maywood, IL; 5) Department of Chemistry, Boston College, Chestnut Hill, MA; 6) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 7) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 8) Department of Genetics, Geisel School of Medicine, Dartmouth College, Hanover, NH; 9) Division of Biostatistics, Washington University School of Medicine, St Louis, MO.

High blood pressure (BP) is the most common cardiovascular risk factor worldwide and a major contributor to heart disease and stroke. We previously discovered a BP-associated missense SNP -rs2272996-in the gene encoding vanin-1, a glycosylphosphatidylinositol (GPI)-anchored membrane pantetheinase. In the present study, we first replicated the association of rs2272996 and BP traits with a total sample size of nearly 30,000 individuals from the Continental Origins and Genetic Epidemiology Network (COGENT) of African Americans (P=0.01). This association was further validated using patient plasma samples; we observed that the N131S mutation led to significantly lower plasma vanin-1 protein levels, and thus the N131S mutation retains apparent protective effects in individuals of lowering systolic BP. We hypothesized that variant vanin-1 is subjected to rapid endoplasmic reticulum-associated degradation (ERAD) as the underlying mechanism for reduction in the gene product. Using HEK293 cells stably expressing vanin-1 variants, we show that N131S vanin-1 was degraded significantly faster than wild type (WT) vanin-1. Consequently, there were only minimal quantities of variant vanin-1 present on the plasma membrane and greatly reduced pantetheinase activity. Application of MG-132, a proteasome inhibitor, resulted in accumulation of ubiquitinated variant protein. Application of hypertension drugs reduced the endogenous vanin-1 protein levels. Our study provides strong biological evidence for the association of the identified SNP with BP and suggests that vanin-1 misfolding and degradation are the underlying molecular mechanism.

2073S

Cardiac enhancers in sub-threshold genetic loci reveal candidate repolarization genes. *X. Wang[1,2,3], N.R. Tucker[4], R. Mills[4], X. Nguyen[4], J. Ye[4], J. Leyton-Mange[4], E.V. Dolmatova[4], P.T. Ellinor[4,5], C. Newton-Cheh[2,4,5], D.J. Milan[4], M. Kellis[2,3], L.A. Boyer[1].* 1) MIT, Cambridge, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA; 4) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA; 5) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA.

The majority of disease-associated loci identified by genome-wide association studies (GWAS) do not affect protein-coding genes and current loci often only explain a modest proportion of heritability for any trait, suggesting that genetic variants in non-coding regions may contribute to disease phenotypes. Here, we integrate genetic information from GWAS and epigenomic maps to show that enhancers significantly overlap known loci associated with QT interval and QRS duration traits. We find that enhancers that overlap QT/QRS loci share similar features including increased H3K27ac density, tissue-selective activity, and regulatory motif enrichment that persist below the genome-wide significance threshold. Using these criteria, we developed a machine-learning framework to identify potential novel loci associated with myocardial repolarization that do not meet genome-wide significance. Remarkably, knockdown of the predicted enhancer gene targets led to repolarization phenotypes in zebrafish, demonstrating that epigenomic signals can reveal new disease genes. Our work suggests that enhancer regions harbor sub-threshold variants, and provides a general strategy for the discovery of novel disease genes for complex human traits.

2074M

Mini Brain Related Kinase / DYRK1B: Novel Gene for Metabolic Syndrome. *M. Fathzadeh[1,3,4], A.R. Keramati[1], G.W. Go[1], R. Singh[1], K. Sarajzadeh[3,5], M. Kasaei[5], M. Babaee Bigi[5], M. Babaei[6], M. Choi[2], S. Faramarzi[1], Sh. Mane[2], J. Hwa[1], A. Hosseinian[6], A. Noorafshan[7], M. Amini[8], R. Malekzadeh[3], J. Tavakkoly[4], R.P. Lifton[2], A. Mani[1].* 1) Department of Internal Medicine, Yale Cardiovascular Research Center, YALE UNIVERSITY, New Haven, CT, USA; 2) Yale Center for Mendelian Genomics, YALE UNIVERSITY, New Haven, CT, USA; 3) Digestive Disease Research Institute, Shariati Hospital , Tehran University of Medical Sciences, Tehran, Iran; 4) Department of Medical Genetics, Tehran University of Medical Sciences, Tehran, Iran; 5) Cardiovascular Research Center, Shiraz University of Medical Sciences, Shiraz, Iran; 6) Ardabil University of Medical Sciences, Ardabil, Iran; 7) Histomorphometry & Stereology Research Centre Shiraz University of Medical Sciences, Shiraz, Iran; 8) Laparoscopy Research Center, Shiraz University of Medical Sciences, Shiraz, Iran.

Metabolic syndrome (MetS), a cluster of risk factors for coronary heart disease (CAD) and type 2 diabetes, is one of the fastest growing epidemics worldwide. Although, a heritable disease, identification of disease genes that underlie the association of its multiple risk factors has been elusive. As with many common disorders, genetic approaches have been complicated by unknown mode of inheritance and number of alleles involved. To overcome this complexity we characterized a subpopulation, which included multiple kindreds with cluster of early onset CAD, central obesity, diabetes and hypertension (HTN). Using combined Whole Genome Linkage analysis and Exome Sequencing we identified a single missense mutation (c.304C>T) (p.[R102C]) in DYRK1B (Gene ID: 9149), which segregated with most features of MetS in affected family members. Overexpression of the mutant allele in vitro demonstrated gain of function effects, characterized by increased expression of the key gluconeogenic enzyme glucose-6-phosphatase (G6PC) and adipogenic proteins PPARGC1A and PPARG. Administration of i.p. Lentivirus vectors carrying wildtype and mutant DYRK1B to C57Bl/6 mice resulted in dramatic rise of blood glucose level and increased body weight compared to empty vector. This was associated with increased hepatic FOXO1 and G6PC expression. Accordingly, oral Glucose Tolerance Test (OGTT) in mutation carriers revealed increased insulin resistance compared to unaffected family members. Mutation carriers had higher ratio of skeletal muscle fast-twitch fibers, indicative of preferential glycolytic catabolism compared to unaffected relatives. Correspondingly, the mutation carriers had lower levels of skeletal muscle peroxisome proliferator-activated receptor gamma protein, coactivator 1 alpha (PPARGC1A), indicative of downregulated oxidative metabolism and mitochondria biogenesis. In conclusion, altered DYRK1B function is associated with MetS, likely due to combination of increased hepatic gluconeogenesis and skeletal muscle insulin resistance. Thus, Dyrk1B is a potential target for development of novel therapeutics against diabetes and obesity.

2075S

Revealing the genetic and gene expression variation in a case-control study for Acute Myocardial Infarction in a Pakistani population. *N.I. Panousis[1,2,3], S. Tuna[4], L. Lataniotis[4], A. Rasheed[5], N. Shah[5], J. Danesh[6], E.T Dermitzakis[1,2,3], D. Saleheen[5,7], P. Deloukas[4,8].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) Institute of Genetics and Genomics in Geneva (iGE3), University of Geneva Medical School, Geneva, Switzerland; 3) Swiss Institute of Bioinformatics, Geneva, Switzerland; 4) William Harvey Research Institute, Queen Mary University of London, London, UK; 5) Center for Non-Communicable Diseases, Karachi Pakistan; 6) Department of Public Health and Primary Care, University of Cambridge, UK; 7) Department of Biostatistics and Epidemiology, University of Pennsylvania, USA; 8) Wellcome Trust Sanger Institute, Hinxton, Cambridge UK.

The risk of coronary heart disease (CHD), the leading cause of death worldwide, is higher in South Asians compared to individuals of European descent. Although lifestyle and congenital factors, which stimulate metabolic dysfunctions, have been proposed to increase the risk for myocardial infarction (MI) there is limited knowledge of the genetic basis of the elevated incidence of CHD in these populations. Here we report an RNA-seq analysis of monocytes from 71 cases of confirmed acute MI and 77 healthy individuals from the Pakistan Risk Of Myocardial infarction study (PROMIS). RNA was extracted and sequenced on the Illumina HiSEQ2000 (median 42,8 million reads, 75 bp paired-end). All individuals have genome-wide SNP data. We identified 972 differentially expressed genes (FDR 5%, fold chance ≥ 2) and performed pathway and functional annotation analyses. Multidimensional scaling analysis based on gene expression reveals distinct clusters within and between cases and controls suggesting different expression patterns. Allele-specific expression (ASE) was estimated for every individual for all the heterozygous sites. We identified on average 182 sites in ASE (FDR 5%) per individual. In order to find genetic variants that affect the gene expression levels of monocytes we performed expression quantitative trait locus (eQTL) analysis. We discovered 5267 eQTLs (FDR 5%, significance level based on permutations) using a 1 Mb window upstream or downstream of the TSS of the genes. Following these results we are examining if there are ASE sites induced or repressed in cases and controls and will undertake a functional annotation analysis of all ASE sites. We are also conducting an enrichment analysis in MI GWAS SNPs for MI-specific eQTLs in order to identify putative causative variants. Overall, these findings together with the ongoing analyses will allow us to investigate further the genetic architecture of CHD in the Pakistani population and better understand the relationship of genetic regulatory variation and gene expression.

2076M

Association of Cytochrome P450 2C9 (CYP2C9) and VKORC1 polymorphisms and warfarin dosage in Iranian patients refer to Shahid Rajaie Heart Center. *r. kameli[1], m. Babanejad[1], m. Imeni[1], m. Soveizi[1], m. Afshari[2], s.h. jamaldini[1].* 1) Cardiogenetics Research Center , Shahid Rajaie Cardiovascular Medical & Resear, Tehran, IranÆ; 2) Zabol University of Medical Sciences, Zabol, IranÁ.

Introduction: Warfarin is a commonly prescribed oral anticoagulant for the treatment and prevention of thrombotic diseases. Warfarin dose has a large interindividual variation. Patients with CYP2C9*2 and/or CYP2C9*3 and VKORC1 (-1639G>A) polymorphisms need lower dose than patients with wild-type variant. These patient metabolized warfarin more slowly than wild patients and traditional dose results in overdose and bleeding in this patients. Objective: This study was conducted to identify the associations between demographic characteristics (sex, height, weight, age, ethnicity), and genetic polymorphisms of CYP2C9 and VKORC1 (-1639G>A) with warfarin dose among Iranian patients. Method: Our study concluded 200 patients that reached to a stable dose of warfarin. By PCR-RFLP method CYP2C9*2 and CYP2C9*3 polymorphisms of CYP2C9 gene and VKORC1 (-1639G>A) was genotyped. Results and Conclusion: Our study showed that CYP2C9 polymorphisms had significant influence on Iranian daily warfarin dose(P= 0.007). Our results suggested that patients with AA genotype in VKORC1 (-1639G>A) require lower doses of warfarin than those with AG or GG genotype(P=0). Height and weight did not have a significant correlation with the warfarin maintenance dose. In addition there was no significant relationship between sex and ethnicity with the maintenance dose of warfarin ($p < 0.05$).

2077S

Regulatory Polymorphisms in DBH Affect Peripheral Gene Expression and Sympathetic Phenotypes. *E.S. Barrie[1], D. Weinshenker[2], A. Verma[3], S.A. Pendergrass[3], L.A. Lange[4], M.D. Ritchie[3], J.G. Wilson[5], H. Kuivaniemi[6], G. Tromp[6], D.J. Carey[6], G.S. Gerhard[7], M.H. Brilliant[8], S.J. Hebbring[8], J.F. Cubells[2], W. Sadee[1].* 1) Pharmacology, Center for Pharmacogenomics, The Ohio State University Wexner Medical Center, Columbus, OH; 2) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA; 3) Center for Systems Genomics, Pennsylvania State University, University Park, PA 16802, USA; 4) Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA; 5) University of Mississippi Medical Center, Jackson, MS 39216, USA; 6) The Sigfried and Janet Weis Center for Research, Geisinger Health System, Danville, PA 17822, USA; 7) Institute for Personalized Medicine, The Pennsylvania State University College of Medicine, Hershey, PA 17033, USA; 8) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI 54449, USA.

Rationale: Dopamine β-hydroxylase (DBH) catalyzes the conversion of dopamine to norepinephrine in the CNS and peripherally. DBH variants have been associated with large changes in circulating DBH and norepinephrine and are implicated in multiple disorders, yet causal relationships and tissue-specific effects remain unresolved. **Objective**: To characterize regulatory variants in *DBH* mRNA, effect on expression in human tissues, and role in modulating sympathetic tone and disease risk. **Methods and Results**: Analysis of *DBH* mRNA in human tissues confirmed high expression in the brain (locus coeruleus, LC) and adrenal gland, but also unexpectedly in sympathetically innervated organs (liver>lung>heart). Allele-specific expression assays of mRNA in LC and adrenals yielded small allelic differences (<twofold), whereas pronounced allelic differences in the liver (2-11 fold) revealed regulatory effects. The minor alleles of two variants, promoter region rs1611115 and exon 2 rs1108580, were associated with significantly reduced DBH mRNA expression in liver and lung, but not brain and adrenals. In mice, *Dbh* mRNA levels in the liver correlated with cardiovascular risk phenotypes. Using a PheWAS (phenome-wide association study) analysis, the minor alleles of rs1611115 and rs1108580 were associated with sympathetic phenotypes including angina pectoris. Testing combined effects of rs1611115 and rs1108580 indicated robust protection against myocardial infarction in two separate clinical cohorts, which was replicated in a third cohort demonstrating increasing protective effect with an increasing number of minor alleles. **Conclusions**: These results demonstrate profound effects of common *DBH* variants on expression in sympathetically innervated organs, modulating clinical phenotypes responsive to peripheral sympathetic tone. This work was supported in part by NIH grant U01092655.

2078M

Obstructive heart defects associated with candidate genes, maternal obesity, and folic acid supplementation. *X. Tang[1], M.A. Cleves[2], T.G. Nick[1], M. Li[1], S.L. MacLeod[2], S.W. Erickson[2], J. Li[1], G.M. Shaw[3], C.A. Hobbs[2].* 1) Biostatistics Program, Department of Pediatrics, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR; 2) Division of Birth Defects Research, Department of Pediatrics, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR; 3) 3Division of Neonatal and Developmental Medicine, Department of Pediatrics, Stanford University Medical School, Palo Alto, CA.

Right-sided and left-sided obstructive heart defects (OHDs) are subtypes of congenital heart defects, in which the heart valves, arteries, or veins are abnormally narrow or blocked. Previous studies have suggested that the development of OHDs involved a complex interplay between genetic variants and maternal factors. Using the data from 569 OHD case families and 1644 control families recruited from the National Birth Defects Prevention Study (NBDPS) between October, 1997 and December, 2008, we conducted an analysis to investigate the genetic effects of 877 single nucleotide polymorphisms (SNPs) in 60 candidate genes for association with the risk of OHDs, and their interactions with maternal use of folic acid supplements, and prepregnancy obesity. Applying log-linear models based on the hybrid design, we identified a SNP in methylenetetrahydrofolate reductase (MTHFR) gene (C677T polymorphism) with a main genetic effect on the occurrence of OHDs. In addition, multiple SNPs in betaine-homocysteine methyltransferase (BHMT and BHMT2) were also identified to be associated with the occurrence of OHDs through significant main infant genetic effects and interaction effects with maternal use of folic acid supplements. We also identified multiple SNPs in glutamate-cysteine ligase, catalytic subunit (GCLC) and DNA (cytosine-5-)-methyltransferase 3 beta (DNMT3B) that were associated with elevated risk of OHDs among obese women. Our findings suggested that the risk of OHDs was closely related to a combined effect of variations in genes in the folate, homocysteine, or glutathione/transsulfuration pathways, maternal use of folic acid supplements and prepregnancy obesity.

2079S

Notch1 haploinsufficiency increases risk of congenital heart defects in the setting of maternal diabetes by an epigenetic mechanism. *M. Basu[1], K. Bosse[1], V. Garg[1, 2].* 1) Center for Cardiovascular and Pulmonary Research, The Research Institute at Nationwide Children's Hospital , 700 Children's Drive, Columbus, OH 43205, USA; 2) Departments of Molecular Genetics & Pediatrics, The Ohio State University, Columbus, OH., USA.

Rationale: Congenital heart disease (CHD) is the leading noninfectious cause of infant morbidity and mortality. Epidemiologic studies have demonstrated the importance of genetic and environmental factors in the multifactorial etiology of CHD. Pre-gestational maternal diabetes is one of the non-genetic risk factors that predispose individuals to CHD. Diabetes is associated with endothelial cell dysfunction and we recently demonstrated a genetic interaction between endothelial nitric oxide synthase and Notch1, which encodes a transmembrane receptor, is important for heart development. We hypothesized that maternal diabetes in combination with Notch1 heterozygosity of the developing embryo will predispose to the development of CHD. **Methods:** We tested the hypothesis using streptozotocin-induced mouse model of diabetes in mice heterozygous for a null allele for Notch1. To elucidate the molecular mechanism underlying this gene-environment interaction, studies were performed using cardiomyoblast (H9C2) and endocardial-derived cell lines, chick embryos and mouse models. Gene expression was measured using qRT-PCR and Western blotting. **Results:** Histologic analysis of E13.5 *Notch1*$^{+/-}$ embryos exposed to maternal hyperglycemia demonstrated an 85% (12/14) incidence of ventricular septal defects compared to 22% (2/9) in wildtype littermates. Gene expression studies in non-diabetic wildtype, diabetic wildtype and diabetic *Notch1*$^{+/-}$ embryos showed that hyperglycemia was associated with a gradual decrease in Notch1 mRNA and a concomitant upregulation in Jarid2, a histone H3K9 demethylase known to regulate Notch1. In H9C2 and endocardial-derived cells along with chick embryos, we showed that hyperglycemia decreases expression of Notch1 and its downstream targets Hey1, Hey2, and Nrg1 in a dose-dependent manner. Similarly, Jarid2 expression increased with hyperglycemia. Further link between hyperglycemia and Notch signaling was demonstrated by reduced CBF-luciferase activity in H9C2 cells transfected with a constitutively active Notch1 intracellular domain that were exposed to high glucose. Studies demonstrating the relative enrichment of Jarid2 on Notch1 locus with hyperglycemia by ChIP-qPCR are in process. **Conclusions:** Our data demonstrate that maternal hyperglycemia disrupts cardiac development by deregulating Notch1 signaling pathway. Our preliminary findings suggest that this gene-environment interaction is mediated by an epigenetic mechanism involving Jarid2.

2080M

eNOS GENE POLYMORPHISM IN NEWBORN PATIENTS WITH PERSISTENT PULMONARY HYPERTENSION. *ML. Lemus-Varela[1], LM. García[1], R. Ramírez[2], MP. Gallegos[2].* 1) Departamento de Neonatología, Hospital de Pediatría, UMAE, CMNO, IMSS, Guadalajara, Jal,. Mex, Guadalajara, Jalisco, Mexico; 2) Laboratorio de Genética Molecular, División de Medicina Molecular, CIBO. IMSS.

Introduction: Persistent pulmonary hypertension of the newborn (PPHN) is a serious condition with high morbidity and mortality , is characterized by increased pulmonary vascular resistance and activity of endothelial nitric oxide synthase (eNOS), that is critical for pulmonary vasodilation. Objective: Identify the T786C and I/D (VNTR) polymorphisms of eNOS gene in patients with PPHN. Material and Methods: Genomic DNA extraction was performed 42 newborns (NB), 21 with PPHN, confirmed by echocardiography (study group) and 21 healthy RN without PPHN (control group). PCR method was performant with specific primers of T786C and I / D (VNTR 27 bp) polymorphisms of eNOS gene. Statistical analysis was analyzed by Student t, and X2 test of Hardy- Weinberg equilibrium. Statistical significance was considered when P value was < 0.05 with confidence intervals of 95%. Results: The study group with PPHN was lower gestational age (35.6 ± 2.81 vs 38.1 +1.9 weeks , p = 0.008) , therefore the average body weight was lower in this group: 2622 ± 626 g, in contrast to the control group : 2,992 ± 565 g . The T786C polymorphism of the eNOS gene, showed statistically significant differences in both genotype distribution and binding of genotypes T / CC / C , observed in 59% (13 /21) of patients with PPHN in contrast to 25 % (5 /21) in the control group and behaved as a risk factor [OR 4.33 (95 % CI 1.15-16.2) , p = 0.029] . In contrast, the T / T genotype, observed in 43% (9/21) of patients with PPHN and 76 % (16/ 21) of the control group (without PPHN), behaved as a protective factor [OR 0.23 (95 % CI 0.06-0.88), p < 0.029] . The I / D genotype , observed in 50% (11/21) of patients with PPHN compared with 10% (2/21) in the control group ; behaved as a risk factor [OR 10.4 (95 % CI 1.9-76.38) , p = 0.002] . The identified mutations alter the function of eNOS , a key pathway in nitric oxide / cGMP pathway , thus favoring pulmonary vasoconstriction and perhaps poor or no response to inhaled nitric oxide . Conclusions: Polymorphisms identified in the eNOS gene partly explain the vasoreactivity characteristic of this serious disease, and vascular remodeling. To our knowledge there are no previous studies exploring characteristic polymorphisms in patients with PPHN.

2081S

A transcriptomic study reveals KLF15 as a circadian metabolic switch in the heart. *L. Zhang, M. Jain.* Dept Med/Human Gen, Case Western Univ SOM, Cleveland, OH.

Physiological parameters of the cardiovascular (CV) system and pathological states such as heart attacks show a circadian rhythm. We have previously identified that in multiple CV diseases, including atherosclerosis and heart failure, KLF15, a zinc-finger transcription factor, expression is reduced. Interestingly, its expression also oscillates in a circadian fashion under the direct regulation of core clock. To study the role of KLF15 in the circadian regulation of CV health and disease, we performed a transcriptomic study of the heart in wild type and cardiomyocyte specific *Klf15* null mice (cKO). We have identified 1335 genes oscillating in the wild type mice heart. These genes fall into two distinctive groups, one peaks at rest to active phase, another peaks at active to rest phase. Interestingly, in addition to the core clock genes, genes peaking at the beginning phase of activity are strongly associated with metabolic processes including amino acid and fatty acid catabolism, ABC transporters, as well as "de-tox" pathways, such as drug and xenobiotic metabolism. This is consistent with the increased energy demand and food intake. The genes expressed highly at the resting phase are associated with cell migration, cell shape and cell cycle regulation. This suggests the resting phase is vulnerable to both hypertrophy and hyperplasia, and is a critical phase for cardiac repair and remodeling. This novel discovery has important implication for chronotherapy in management of patient with CV disease. There are 1003 oscillating genes showed autonomous KLF15 dependency, specifically, cKO mice showed a loss of all metabolic gene induction upon active phase, including lysine, tryptophan, branch chain amino acids and fatty acid metabolism. The oscillation of core clock genes, "de-tox" genes and ABC transporters are preserved. This observation suggests KLF15 is the switch for metabolic control in the heart for the active phase and loss of function of KLF15 in the diseased state likely results in an energy deficiency state in the heart. We have also identified 473 genes, which gained oscillation in the cKO mice without an increase of other cell types including inflammatory cells. De novo motif prediction has identified enrichment of EGR2, NR5A2, MYC and several others. This is the first report of gaining circadian oscillation in a genetic mutant mice. We are in the process of confirming and studying the molecular mechanism of this phenotype.

2082M

A Large Scale Genome Wide Association Study of Varicose Veins in the 23andMe Cohort. *R.K. Bell, E.Y. Durand, C.Y. McLean, N. Eriksson, J.Y. Tung, D.A. Hinds.* 23andMe Inc., Mountain View, CA.

Varicose veins are a common chronic condition in which veins become enlarged and twisted due to deficient functioning of one-way valves that normally return blood to the heart. They usually appear in the legs, and are more common in women than men. While they are most commonly considered a cosmetic problem, in some cases they can lead to, or signal, more serious circulatory problems. We carried out a GWAS of self-reported varicose veins in the legs in the 23andMe participant cohort, including approximately 20,000 cases and 65,000 controls of European ancestry, imputed against 1000 Genomes reference haplotypes. We identified several novel associations for varicose veins near or in genes involved in various circulatory and blood related functions, including blood group determination, regulation of blood pressure and/or circulation, vascular growth factors. Our strongest association was with SNP rs507666 (p=5.2e-20), found in ABO, which encodes a protein defining the ABO blood group. Variation in SNP rs507666 determines A1 blood subgroup status. This finding is consistent with a previous study that reported the association between the A blood group and varicose veins. It also identified an additional SNP, rs966562 (p= 1.8e-16), near XKR5 involved in Kell blood group determination. Our study also identified SNPs in genes which have previously been characterized as players in the regulation of blood pressure or circulation, including SNP rs11121615 (8.1e-12) in CASZ1, associated with blood pressure variation and SNP rs7111987 (2.2e-10) near ADM which encodes a peptide which may function as a hormone in circulation control. Finally, the GWAS of varicose veins points to a role of genes involved in circulatory development: SNP rs1433196 (5.2e-11) is found in ANGPT1 that encodes a type of angiopoietin, a group of proteins with important roles in blood development; SNP rs966562 (mentioned above) near ANGPT2, that encodes an antagonist for ANGPT1; and SNP rs6905288 (1.5e-8) near VEGFA which encodes a protein whose effects include angiogenesis, vasculogenesis and endothelial cell growth.

2083S

An extreme phenotype approach to identify genes in Caribbean Hispanics for carotid plaque, a preclinical marker of atherosclerosis. *S.H. Blanton, A.H. Beecham, H. Gardener, L. Wang, C. Dong, D. Cabral, R. Sacco, T. Rundek.* Human Genetics, Univ Miami, Miami, FL.

Carotid plaque is a measure of subclinical cardiovascular disease and is a risk factor for stroke and heart attack. The genetic underpinnings of plaque are not known. The study goal was to identify risk/protective loci for plaque using an extreme phenotype approach in the analysis of our genome-wide association study. In individuals from the population-based Northern Manhattan Study, plaque was detected by high-resolution B-mode ultrasound and plaque area was measured by automated edge detection software and expressed as total plaque area, a sum of all plaque areas within an individual. Genotyping was done with the Affymetrix 6.0 SNP array. Among the 908 Hispanics, we first calculated the residual score for each participant by regressing plaque area on significant known risk factors (age, sex, packyears of smoking, SBP, diabetes, LDL:HDL ratio, homocysteine levels, high school completion, LDL, lipid lowering medication, and WBC count). To enrich the genetic effects and increase our power to detect associations, we next identified the individuals in the extreme 10% and 20% of the residual distribution, for a total of 90 and 200 individuals on each side respectively, thus ensuring that the phenotypic extremes were largely unexplained by these known risk factors. For each of the extreme thresholds, we performed logistic regression analysis on 877K SNPs, while controlling for the top 3 PCs. A gene based analysis was then performed based on the SNP results using VEGAS. The top gene from the 10% threshold (p=3.80E-05) was member RAS oncogene family (RAB2B). Members of the RAS family are thought to play a critical role in endothelial function. The top gene from the 20% threshold (p=7.60E-05) was retinal pigment epithelium-specific protein 65kDa (RPE65). In addition to the gene based analysis, we also conducted a pathway based analysis using genes with a p<0.05 as input into WebGestalt. After adjusting for multiple testing, several pathways of interest were identified. In particular for the 10% threshold, the taste transduction pathway was significant (p=5.0E-04). An analysis of our diet data found differences (p<0.05) in the intake of vegetables (but not fruit) between the lower 10% and upper 10%, as well as the lower and upper 20%, with individuals with smaller than expected plaque area consuming more vegetable, suggesting that genetic variants affecting taste could contribute to plaque development via influencing vegetable consumption.

2084M

Genetic variants in *LEKR1* and *GALNT10* modulate sex-difference in carotid intima-media thickness: A genome-wide interaction study. *C. Dong[1], A. Beecham[2], L. Wang[2], D. Cabral[1], S.H. Blanton[2], R.L. Sacco[1,2,3], T. Rundek[1,3].* 1) Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL; 2) John T. McDonald Department of Human Genetics, John P Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL; 3) Department of Public Health Sciences, Miller School of Medicine, University of Miami, Miami, FL.

There is an established sex difference in carotid artery intima-media thickness (IMT), a recognized marker of atherosclerosis. However, the genetic underpinnings that modulate sex-differences in gene-IMT associations are largely unknown. With a multistage design using 722,379 single nucleotide polymorphisms (SNP) genotyped with the Affymetrix 6.0 chip, we performed a genome-wide sex-by-SNP interaction study in 931 Hispanics (mean age: 69±8 years, 61% women, mean IMT: 0.70±0.06 mm), followed by replication in 257 non-Hispanic blacks (mean age: 73±9 years, 62% women, mean IMT: 0.74±0.10 mm) and 153 non-Hispanic whites (mean age: 73±9 years, 50% women, mean IMT: 0.76±0.10 mm). Assuming an additive genetic effect for each SNP based on the minor allele number, we performed multiple linear regression analysis to test for sex-by-SNP interaction on the IMT while controlling for age and the top 3 principal components estimated to capture ancestry by EIGENSTRAT. Among 14 SNPs with an interaction of p<5.0E-6 in Hispanic discovery sample, replicated interactions were found for four SNPs in leucine, glutamate and lysine rich 1 (LEKR1) gene in non-Hispanic whites (p≤0.03) and for one SNP in polypeptide N-acetylgalactosaminyltransferase 10 (GALNT10) gene in non-Hispanic blacks (p=0.043). Specifically, for the top SNP rs7616559 in LEKR1 gene, the adjusted mean difference between men and women was 0.003 mm (p=0.67) for AA-carriers, 0.044 mm (p=2.1E-9) for AG carriers, and 0.064 mm (p=3.0E-5) for GG carriers in the combined sample. For SNP rs2081015 in GALTN10 gene, the adjusted mean difference between men and women was -0.017 mm (p=0.13) for TT-carriers, 0.022 mm (p=0.001) for TC carriers, and 0.051 mm (p=1.2E-10) for CC carriers in the combined sample. Genetic variants near LEKR1 gene have been associated with the measures of adiposity in newborns and variants near GALNT10 gene have been associated with body mass index. Given the consistent findings across different-ethnic groups, further studies are warranted to perform in-depth investigations of functional genetic variants in these regions.

2085S

The Kaiser Permanente/UCSF Genetic Epidemiology Research study on Adult Health and Aging: A blood pressure genome-scan in ~100,000 Subjects. *T. Hoffmann[1], A. Chakravarti[2], G. Ehret[2,3], C. Iribarren[4], Y. Banda[1], E. Jorgenson[4], M.N. Kvale[1], C. Schaefer[4], N. Risch[1,4].* 1) University of California San Francisco, San Francisco, CA; 2) John's Hopkins University, Baltimore, MD; 3) Geneva University Hospitals, Switzerland; 4) Division of Research, Kaiser Permanente, Northern California, Oakland, CA.

Blood pressure (BP) is quantitatively the major risk factor for cardiovascular disease. The Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort has detailed data on longitudinal blood pressure, prescription medication, and genotypes on ~675,000 markers for nearly all of its 100,000 participants (81% white, 7.5% Asian, 7% Latino, and 3.5% African American). Approximately 1.3 million systolic (SBP) and diastolic (DBP) blood pressure measures from electronic health records were used to derive long term average phenotypes and association tests with genotyped variants were run based on the directly genotyped markers.

We conducted a genome-wide association study among the subjects of European ancestry. All analyses were adjusted for age, age^2, sex, BMI, and the first ten principal components. Follow-up analysis was conducted in the other races/ethnicities. We compared the results of our association analysis of SBP to previously reported GWAS hits from the NHGRI GWAS catalog, and found that, at a p<0.0016 (0.05/32 SNPs), 23 SNPs replicated, 8 did not replicate, and 1 was not informative. Further we identified 3 novel loci at genome-wide significance (p<5×10^-8). When comparing to DBP, we found that, at a p<0.0016, 23 SNPs replicated, 6 did not replicate, and 1 was not informative. Further we identified 7 novel loci at genome-wide significance.

In summary, we replicate the great majority of BP SNPs listed in the NHGRI GWAS catalog, and identify 10 new common BP loci.

2086M

Single variant and burden analysis of low frequency variants for Fibrinogen, FVII, FVIII, and vWF. *J.E. Huffman[1,2], CHARGE Hemostasis Working Group.* 1) MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom; 2) National Heart, Lung, and Blood Institute's Framingham Heart Study and Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD, USA.

Fibrinogen, coagulation factors VII (FVII) and VIII (FVIII, and von Willibrand factor (vWF) occupy key roles in blood coagulation. Plasma levels are associated with risk of arterial and venous thrombosis and fibrinogen is also a marker of inflammation. The CHARGE Hemostasis Working Group previously identified common variants associated with plasma levels of these factors. We now explore associations with rare and functionally relevant variants.

Individuals of European, African, East Asian and Hispanic ancestries were genotyped using the Illumina “Exome Chip” genotyping array and variants called using a common algorithm across cohorts. Association of rare variants with fibrinogen (n=76,316), FVII (n=25,372), FVIII (n=28,291), and vWF (n=23,272) were tested using both single-variant and burden methods. Burden tests included both unidirectional (T5 & T1) and bidirectional (SKAT) tests. These were restricted to include only missense, nonsense, or splice site single nucleotide variants. Fixed effects meta-analyses were used to combine results across studies using the seqMeta package for R. Conditional analyses were also undertaken for loci where common genetic variation was known to be associated with plasma levels.

Single variant associations were found within previously reported genes for fibrinogen (*FGG*, *FGB* & *HNF4A*), FVII (*F7*) and FVIII/vWF (*ABO*, *VWF* & *STAB2*). Conditioning on known common variants, we identified newly associated rare coding variants with independent associations. These included rs201909029 ($p=5.21\times10^{-13}$) in *FGB* and *FGG* SNPs rs148685782 ($p=2.30\times10^{-150}$) and rs145051028 ($p=4.84\times10^{-9}$) in addition to the previously reported association of rs6054 ($p=3.70\times10^{-42}$) in *FGB*. Significant rare variants tended to have larger effect sizes than previously reported common variants. In some instances, association of different rare variants within the same gene were noted when comparing ethnicity-specific results. Results of the gene-level burden tests were driven primarily by single variant associations rather than a burden of rare variants.

The investigation of rare, functional variants associated with circulating clotting factor levels can add to the knowledge of molecular pathways and mechanisms related to hemostatic disorders and their potential prevention or treatment.

2087S

Carotid intima-media thickness: a genome-wide association analysis among African Americans. *S.M. Tajuddin[1], M.A. Nalls[2], M.F. Keller[2], A.B. Zonderman[1], M.K. Evans[1].* 1) Laboratory of Epidemiology and Population Sciences, National Institutes of Health, Baltimore, MD; 2) Laboratory of Neurogenetics, National Institutes of Health, Bethesda, MD.

Although the age-adjusted death rate from coronary heart disease (CHD) continues to decline in the population overall, there remains significant morbidity and mortality disparities for African Americans (AAs). The atherosclerotic process of CHD leads to thickening of the intimal and medial layers of the common carotid artery. Carotid intima-media thickness (CIMT), measured by B-mode ultrasound, is a non-invasive assessment of subclinical atherosclerosis and has been shown to predict cardiovascular events. Genetic factors influence CIMT. Identification of genetic factors linked with CIMT may facilitate early detection of atherosclerosis and improved treatment strategies. Recent genome-wide association studies have identified genetic loci associated with CIMT in European ancestry and Hispanic populations. However, the association between genetic variants and CIMT has never been investigated among AAs. To identify genetic variants associated with CIMT in AAs, we conducted a genome-wide association analysis of CIMT in AAs in the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study, an epidemiological, longitudinal study of health disparities. CIMT in the left common carotid artery was measured by high-resolution B-mode ultrasound. We analyzed 16.2 million genotyped and 1000 Genomes Project imputed SNPs that passed stringent quality control criteria in 584 AA (55% men). We estimated beta coefficients and p values using multiple linear regression models adjusted for age, sex and first 10 principal components under additive genetic effect model. We identified eight SNPs in three novel loci (*MMAA*, *LINC00299*, and *CAMTA1*) associated with CIMT at genome-wide significance level of $P < 5 \times 10^{-8}$. The lead SNPs were *MMAA*-rs142277468 $P = 2.6 \times 10^{-9}$, *LINC00299*-rs12477055 $P = 2.7 \times 10^{-8}$, and *CAMTA1*-rs72863082 $P = 2.8 \times 10^{-8}$. We identified three genetic regions that are associated with CIMT in this African ancestry population. SNPs in *CAMTA1* and *LINC00299* have been implicated in plasma fatty acid levels and obesity related traits, respectively. *MMAA* is involved in cobalamin processing and transportation into the mitochondria and may have an effect on atherosclerosis through homocysteine metabolism. These findings may shed light on atherosclerosis development and provide a biomarker of susceptibility in AAs. Replication studies in other AA cohorts will be conducted to validate these findings.

2088M

Genetic determinants underlying hypertension in multi-ethnic populations. *N. Vasudeva[1], L. Wang[1,2], Z. Liu[2], P. Goldschmidt[3], M. Pericak-Vance[1,2], D. Seo[3], G. Beecham[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami, FL., U.S.A; 2) Dr. John T. Macdonald Foundation, Department of Human Genetics, University of Miami, Miami, FL; 3) Division of Cardiology, Miller School of Medicine, University of Miami, Miami, FL.

Hypertension is a major cardiovascular health risk that is well-established as a heritable trait. The environmental factors influencing it, as well as their complex interactions make it challenging to discover the genetic determinants affecting the trait. A cumulative effect of multiple variants has been believed to contribute to this polygenic trait. To dissect the genetic basis of this trait, we analyzed European, Hispanic and African ancestry samples from Miami Cardiovascular Registry (MCR). Genome-wide genotyping was performed on about 2000 subjects using an Affymetrix SNP array. The subjects were separated into European, Hispanic and African ancestry subgroups using Eigenstrat. Sample and SNP quality control were followed by single variant association test. To account of the effect of blood pressure lowering medications, we added 7% and 6% to systolic blood pressure (SBP) and diastolic blood pressure (DBP) measures, respectively, for each antihypertensive medication taken. Linear regression was used to analyze single variant association adjusting for covariates such as sex, age, age^2, anti-cholesterol medicines and diabetes. SNPs rs937421, rs7025928, rs11965507 and rs10488371 showed significant association with BP phenotypes. We further performed a gene-based association test using the SKAT R package. Significant association was observed in Hispanics for genes AURKAIP1 (p= 5.39E-06), FAM53C (p= 6.59E-06) for DBP and TSNAX-DISC1 (p= 6.05E-06) for mean arterial pressure (MAP). Various genes approached the genome-wide significant level in the European ancestry subgroup as well as African Americans. In order to find pathways connected to these large numbers of genes, we used web based tool WebGestalt. The results showed that in the European subgroup, cytokine-cytokine receptor interaction was significantly associated with SBP (p=5.83E-06) and RIG-I-like receptor signaling pathways with pulse pressure(PP); for Hispanics, metabolic pathways were significant for multiple phenotypes, with DBP being most prominent phenotype (p=5.06E-11); in African Americans metabolic pathways were also highly associated with SBP (p=5.15E-07). We will further analyze these genotypes and pathways using alternate methods such as VEGAS to better understand the genetic etiology of hypertension in our population.

2089S

Contribution of Global Copy Number Variants to Down Syndrome-associated Atrioventricular Septal Defects. *D. Ramachandran[1], J.G. Mulle[2], A.E. Locke[3], L.J. Bean[1], T.C. Rosser[1], P. Bose[1], K.J. Dooley[4], C.L. Cua[5], G.T. Capone[6], R.H. Reeves[7], C.L. Maslen[8], D.J. Cutler[1], S.L. Sherman[1], M.E. Zwick[1].* 1) Department of Human Genetics, Emory University, Atlanta, GA; 2) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA; 3) Center for Statistical Genetics and Department of Biostatistics, University of Michigan School of Public Health, MI; 4) Sibley Heart Center Cardiology, Children's Hospital of Atlanta, GA; 5) Heart Center, Nationwide Children's Hospital, Columbus, OH; 6) Down Syndrome Clinic and Research Center, Kennedy Krieger Institute, Baltimore, MD; 7) Department of Physiology, McKusick Nathans Institute for Genetic Medicine, School Of Medicine, John Hopkins University, MD; 8) Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR.

Atrioventricular septal defects (AVSD), a severe congenital heart defect, occur in the general population in ~1 in 10,000 births. Nearly 20% of infants with Down Syndrome (DS) have an AVSD, representing an approximately 2000-fold increased risk compared to the euploid population. Our central hypothesis was that in the presence of an extra chromosome 21, otherwise benign copy number variants (CNVs) act to increase the risk of AVSD in individuals with DS. Here we sought to test two specific questions: (1) Do common CNVs of large effect contribute to the elevated risk of AVSD in the DS population? (2) Do rare CNVs increase the risk of AVSD in the DS population? We used the Affymetrix SNP 6.0 genotyping platform to comprehensively characterize CNVs in 452 ethnically matched individuals with DS, comprising of 210 cases (DS + complete AVSD) and 242 controls with a structurally normal heart (DS + NH). We implemented strict quality control filters to minimize false positive calls, including 3 algorithms to make the CNV calls (BEAST, GADA and GLAD). We also required each putative CNV call to contain > 20 SNPs within the interval. After excluding CNVs overlapping centromeres, we identified 541 deletions (253 in cases and 288 in controls) and 383 duplications (177 in cases and 206 in controls). Results from burden and region-wise analyses using PLINK revealed that despite the 2000 fold elevated risk, common CNVs of large effect (OR > 2.0) do not account for the increased risk observed in DS-associated AVSD. In contrast, cases do harbor a significantly elevated burden of large rare variants (> 100kb, < 1% frequency) ($p < 0.01$) and case deletions intersect genes more often than those observed in controls ($p < 0.007$). Gene enrichment analysis showed a trend for enrichment among deletions impacting the ciliome pathway in cases compared to controls. No significant differences were observed for large rare duplications between cases and controls. Our findings suggest that the etiology of AVSD is highly complex and does not arise from the action of a few common variants of large effect. Instead, our data support a multifactorial model, wherein large rare deletions play a significant role in elevating the risk of AVSD in a trisomic background. Our study further suggests an important role for ciliome genes in AVSD.

2090M

Identification of Loci associated with Bicuspid Aortic Valve (BAV). *S.C. Body[1], M. Heydarpour[1], J.G. Seidman[2], S. Prakash[3], D. Milewicz[3], Y. Bosse[4], G. Limongelli[5], the Bicuspid Aortic Valve Consortium (BAVCon).* 1) Anesthesia, Brigham & Womens Hosp, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) University of Texas, Houston, TX; 4) University of Laval, Quebec, Canada; 5) Monaldi Hospital, Naples, Italy.

OBJECTIVE: Bicuspid aortic valve (BAV) is observed in 0.6-1% of humans and is associated with premature aortic stenosis, aortic aneurysm and other congenital cardiac anomalies. BAV has been associated with high-penetrance rare mutations, mainly in NOTCH1, in a few families, but the majority of BAV is sporadic with unidentified inheritance. The embryologic molecular mechanisms causing BAV are unknown. The objective of this study was to identify pathway/genes associated with sporadic BAV. METHODS: We genotyped 456 Caucasian cases (339 males, 117 females) using Illumina Omni2.5. We used three dbGaP cohorts genotyped with either Omni2.5 or Omni5 arrays containing 1,869, 897, and 2,450 Caucasian controls respectively. Quality control and population stratification of the genotype data were performed in PLINK. An additive logistic regression model was performed for association analysis across all three GWAS studies accounting for gender and variants associated with aortic stenosis and aortic aneurysm in order to reduce confounding by presentation. We performed three analyses, one for each control group versus the cases, then pooled significant P-values (<1*10-5) for common SNPs (>5%) with the same direction for genetic marker effects (ORs). RESULTS: 22 SNPs in 12 regions with P<1*10-5 in all three GWA studies remained significant after correction for multiple testing. Five regions (1p21, 4q32, 5q14, 8p23, 12q22) contained SNPs of genome-wide significance (P<5*10-8) with odds ratios of 1.5-2.6. The most significant SNPs with high LD (r2>0.8) were identified on chromosome 8p23. CONCLUSION: Our findings suggest that several polymorphisms could explain the higher predisposition of BAV in Caucasians. Novel regions were identified on chromosomes 1,4,5,8,10,12,16, and 18, which have important role in heart development and can be good candidates for fine mapping and replication in other cohorts. In addition, prior work has demonstrated BAV-linked genes involved in cardiac development of other vertebrates (e.g. Notch1-deficient-mice, Gata5-null-mice). Such studies combined with our ability to delete genes in specific cells at specific developmental states in animal models will undoubtedly unravel numerous candidate Coronary Heart Disease causing genes that can be directly tested in human genetic studies. Conversely loci identify in human cohorts can be tested in animal models to confirm (or not) their causative link to disease. These findings are being replicated.

2091S

Relationship Between Plasma Betaine Levels and Cardiovascular Disease: Results of a Genome-Wide Association Study. *J. Hartiala[1], W.H. Tang[2], Z. Wang[2], S.L Hazen[2], H. Allayee[1].* 1) Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90033; 2) Department of Cardiovascular Medicine, Cleveland Clinic, Cleveland, OH 44195.

We recently identified several metabolites generated through gut microbiome and hepatic-mediated metabolism of dietary choline and L-carnitine as novel risk factors for cardiovascular disease (CVD). Elevated levels of trimethyl amine N-oxide (TMAO) in particular were found to be pro-atherogenic in mice and humans but the role of betaine, which can also be generated from choline, in CVD is less clear. Following our recently published comparative genomics analysis of TMAO, we sought to identify the genetic determinants of plasma betaine levels. In stage 1, a GWAS with ~1900 subjects from the GeneBank study identified 4 loci on chromosome (chr) 1q32, 2q34, 5q14, and 16q24 that were significantly or suggestively associated with betaine levels. In stage 2 with ~2000 additional GeneBank subjects, 2 independent SNPs on chr 5q14 (rs617219 and rs16876503) and rs715 on chr 2q34 demonstrated association with betaine, which became even more significant in all ~4000 subjects (p=6.0E-9 - 9.0E-13). A weighted genetic risk score with these 3 SNPS also yielded highly significant association with betaine (p=1.1E-23). The SNPs on chr 5q14 are located near the BHMT gene (rs617219), which catalyzes the simultaneous conversion of betaine and homocysteine to dimethylglycine (DMG) and methionine, respectively, and the DMGDH gene (rs16876503), which further metabolizes DMG to sarcosine and subsequently glycine. Of these additional metabolites, rs617219 significantly lowered DMG levels (p=4.0E-4) whereas rs16876503 modestly increased methionine (p=0.04). By comparison, rs715 on chr 2q34 is located within the CPS1 gene, which is the rate-limiting enzyme in the urea cycle. In addition to the association with betaine, rs715 increased ornithine levels (p=9.0E-3) and decreased citrulline (p=3.0E-4), arginine (p=4.0E-3), and urea (p=0.07). These results suggest that rs715 reduces flux down the urea cycle by decreasing CPS1 activity and/or expression since the strongest effects were on the most proximal urea cycle metabolite (citrulline) with increasingly weaker effects on more distal metabolites (i.e. arginine and urea). Lastly, of the betaine-associated SNPs on chr 2q34 and 5q14, rs715 was significantly associated with decreased risk of CVD (OR=0.71, 95% CI 0.58-0.87; p=0.004) in all ~10,000 GeneBank subjects. Taken together, these studies provide evidence for a genetically mechanistic link between betaine metabolism, the urea cycle, and CVD in humans.

2092M

Genome-wide association study identifies novel susceptibility loci for venous thromboembolism in African Americans. *W. Hernandez[1], E.R. Gamazon[1], A. Konkashbaev[1], R.A. Kittles[2], L.H. Cavallari[3], M.A. Perera[1].* 1) The University of Chicago, Department of Medicine, Section of Genetic Medicine, Chicago, IL; 2) University of Illinois, Department of Medicine, Institute of Human Genetics, Chicago, IL; 3) University of Florida, Department of Pharmacotherapy and Translational Research, College of Pharmacy, Gainesville, FL.

Venous thromboembolism (VTE) is a chronic multifactorial disease encompassing deep vein thrombosis (DVT) and pulmonary embolism (PE). It is a major public health burden and the third most common life-threatening cardiovascular condition resulting in high rates of hospitalization and mortality. In the US, African Americans (AAs) have the highest incidence and mortality rates. Although studies suggest that VTE is highly heritable, few associations between VTE risk and genetic variants have been established. Furthermore, the role of these polymorphisms in the risk of VTE among AAs is conflicting at best. The lack of genetic susceptibility to VTE among AAs, in great part, may be due a combination of identifying susceptibility genetic variants in Caucasian cohorts and subsequently testing these variants in AAs, as well as, the very low frequency of these variants among AAs. In this study, we present preliminary findings from the first genome-wide association study (GWAS) and risk of VTE conducted exclusively on AAs (137 cases and 437 controls). Our results revealed 15 independent signals that reached a genome-wide significance level of 5.0×10-8. Five variants in PTGER3 (rs570021), PPP1R12B (rs705744), SLC22A3 (rs9364552), SLC39A11 (rs951284), and MGST1 (rs2975139) and eight intergenic variants (rs7503127, rs7669402, rs7729075, rs9276835, rs11219138, rs2535607, rs2749490, rs711831) were associated with increased risk of VTE (ORs= 2.2 - 2.6). Two variants, rs10234060 and rs711831, were found to decrease risk of VTE (OR=0.40 and 0.44, respectively). These ORs are larger than those previously identified in GWASs on VTE risk among Caucasians. In addition, we conducted bioinformatics analysis of the top signals to identify potential expression quantitative trait loci (eQTLs) utilizing whole blood from AAs. The risk variant, rs9364552 located in SLC22A3 (OR=2.3; P=4.30×10-8), was found to be a cis-eQTL (β=-0.19; P=0.02) for plasminogen (PLG) which is crucial in breaking down fibrin in blood clots. Individuals carrying the rs9364552 minor allele would express lower levels of PLG, consequently increasing their risk of VTE. Furthermore, SLC22A3 and PLG have been shown to affect lipoprotein(a) levels - an independent risk factor for VTE. Currently, we are in the process of validating these finding in an independent AA cohort of VTE. Our study provides new molecular insight into the underlying mechanism that may regulate VTE susceptibility in AAs.

2093S

Nervous system-related loci determining sex-difference in blood pressure reactivity to cold stress in both Chinese and Whites. *Q. Zhao[1], X. Kong[1,2], TN. Kelly[1], C. Li[1], D. Gu[3], J. He[1].* 1) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 2) Department of Endocrinology, China-Japan Friendship Hospital, Beijing, China; 3) State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

Exaggerated blood pressure (BP) response during the cold pressor test (CPT) has been suggested as a risk factor for developing cardiovascular disease. Women were reported to exhibit greater BP increase than men during the CPT in populations of different ethnicities. The aim of the study was to identify sex-specific genetic determinants for BP reactivity to cold stress. The discovery sample was from the Genetic Epidemiology Network of Salt Sensitivity (GenSalt) study, which was conducted among Han Chinese. During the CPT, BP was measured prior to and after the participants immersed their right hand in ice water for 1 minute. A total of 1,881 GenSalt participants completed the CPT and were genotyped using the Affymetrix SNP array 6.0. A genome-wide association analysis was conducted to examine the interaction effect of SNPs and sex on BP reactivity during the CPT. A total of 13 loci showed potential sex difference in the association with BP reactivity variables (P for interaction < 1.0×10^{-5}). These loci were further tested among 1,448 participants of European ancestry from the Coronary Artery Risk Development in Young Adults (CARDIA) study, in which a similar cold stress test was conducted. Five of the 13 loci (1p32.2, 2q33.1, 3q26.31, 5q15, and 13q33.3) identified in the GenSalt study showed potential transethnic replication in the CARDIA study (P for interaction < 0.1). Four of the five loci (1p32.2, 3q26.31, 5q15, and 13q33.3) include nervous system-related genes which have been implicated in neurological disorders, such as Parkinson's disease (*CDCP2* in 1p32.2 and *CAST* in 5q15), bipolar disorder (*TMEM212* in 3q26.31), and cerebral calcification (*COL4A1* in 13q33.3). Stratified analyses by sex for significant interactions between sex and SNPs were further conducted. For example, the index SNP rs3766457 of the locus 1p32.2 was associated with lower pulse pressure (PP) response in men but greater PP response in women (-2.64 vs. 1.48 mm Hg for each minor G allele) in the GenSalt study. Meanwhile, this SNP was associated with lower diastolic BP response in men but greater response in women (-2.42 vs. 4.02 mm Hg for each G allele) in the CARDIA study. In conclusion, our study provides the first evidence for the transethnic replication of sex-specific genetic factors for BP response to cold stress and implicates multiple potential nervous system-related genes in determining the sex difference in BP reactivity to stress.

2094M

Finding low-frequency causal genetic variants for lipids by genotyping subjects with extreme HDL-c levels. *W. Zhou[1], OL. Holmen[2,3], H. Zhang[1], J. Chen[1], M. Boehnke[4], GR. Abecasis[4], K. Hveem[2,5], CJ. Willer[1,6,7].* 1) Department of Internal Medicine, Division of Cardiology, University of Michigan Medical School, Ann Arbor, Michigan, 48109, United States of America; 2) HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, 7600 Levanger, Norway; 3) St. Olav Hospital, Trondheim University Hospital, Trondheim, Norway; 4) Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, 48109, United States of America; 5) Department of Medicine, Levanger Hospital, Nord-Trøndelag Health Trust, 7600 Levanger, Norway; 6) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, 48109, United States of America; 7) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan, 48109, United States of America.

Genome-wide association studies have identified more than 157 loci associated with lipid levels, which are treatable and heritable risk factors for cardiovascular diseases. Most of these variants are common with modest effect sizes. Missing heritability of blood lipid levels suggests less frequent (1 - 5% minor allele frequency) or rare (< 1% minor allele frequency) coding variants may exist for lipids. To detect low-frequency causal variants, samples with extreme phenotypes likely have increased power compared to population-based. We successfully genotyped 1,751 non-MI samples with extremely high HDL-c levels, 1,744 non-MI samples with extremely low HDL-c levels from the Norwegian HUNT-study using the Illumina HumanExome Beadchip, allowing for direct genotyping of 72,399 polymorphic coding variants. We evaluated the power of the extreme HDL samples (N=3,495) in comparison to a genotyped set of non-extreme samples (N = 3,495 from the same HUNT cohort) to detect the known HDL variants and genes. Single SNP association results for the 50 HDL known variants are more significant using extreme HDL samples than population-based samples (Pwilcox = 5.22×10^{-8}), as were the gene-based burden test results for the 11 known HDL genes (Pwilcox = 9.3×10^{-3}). Single SNP and gene-based association tests showed that no novel lipid loci were found with the extreme HDL-c samples, suggesting that even with extreme phenotypes, larger sample size or more complete genotypes from sequencing may be needed to detect novel loci.

2095S

Identification of blood pressure related genes by population-based transcriptome analyses. *C. Müller[1,2], K. Schramm[3,4], C. Schurmann[5,6], S. Kwon[7], A. Schillert[2,8], D. Lau[1,2], C. Herder[9,10], G. Homuth[5], S. Wahl[11], H. Grallert[11], T. Illig[11,12], A. Peters[11,13], M. Dörr[14,15], X. Guo[17], W. Palmas[18], T. Meitinger[3,4,16], A. Teumer[22], M. Cartensen[9,10], P.S. Wild[19,20,21], H. Völzke[15,22], M. Roden[9,10], D.M. Herrington[23], U. Völker[5,15], A. Ziegler[2,8], Y. Liu[24], T. Zeller[1,2], S. Blankenberg[1,2], H. Prokisch[3,4,16], S.B. Felix[14,15].* 1) Clinic for general and interventional Cardiology, University Heart Center Hamburg, Hamburg, Germany; 2) DZHK (German Centre for Cardiovascular Research), partner site Hamburg/Kiel/Lübeck; 3) Institute of Human Genetics, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany; 4) Institute of Human Genetics, Technical University Munich, München, Germany; 5) Interfaculty Institute for Genetics and Functional Genomics, University Medicine and Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany; 6) The Genetics of Obesity and Related Metabolic Traits Program, The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 7) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, USA; 8) Institute of Medical Biometry and Statistics, University of Lübeck, UNIVERSITY HOSPITAL Schleswig-Holstein, Lübeck, Germany; 9) Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Germany; 10) Member of the German Center for Diabetes Research (DZD e.V.), partner site Düsseldorf; 11) Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 12) Medical School Hannover, Hannover Unified Biobank, Hannover, Germany; 13) Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 14) Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany; 15) DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany; 16) DZHK (German Centre for Cardiovascular Research), partner site Munich, Munich, Germany; 17) Institute for Translational Genomics and Populations Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, California, United States; 18) Department of Medicine, Columbia University, New York, NY 10032, USA; 19) Department of Medicine 2, University Medical Center Mainz, Germany; 20) Center for Thrombosis and Hemostasis, University Medical Center of the Johannes Gutenberg-University Mainz, Germany; 21) DZHK (German Centre for Cardiovascular Research), partner site Rhine-Main, Germany; 22) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; 23) Section on Cardiology, Department of Internal Medicine, Wake Forest School of Medicine, Winston Salem, North Carolina, USA; 24) Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest University, Winston-Salem, North Carolina, USA.

High blood pressure (BP) is a major risk factor for cardiovascular diseases. We analyzed associations between the blood transcriptome and BP traits within large cohorts of the MetaXpress consortium by applying standardized protocols.

The study population included 4,533 individuals with available transcriptomics data from three German cohorts: Gutenberg Health Study (GHS), Cooperative Health Research in the Region of Augsburg (KORA F4) and Study of Health in Pomerania (SHIP-TREND) and one US cohort: Multi-Ethnic Study of Atherosclerosis (MESA). Expression levels were measured in monocyte (n=2,549) and whole blood cell (n=1,984) samples using Illumina HumanHT-12 BeadChip arrays. Associations with systolic BP (SBP), diastolic BP (DBP) and pulse pressure (PP) were computed by linear regression models adjusted for sex, age, BMI and technical covariates within each study. A meta-analysis was conducted within GHS and MESA using the inverse variance method. Significant associations (FDR≤0.05) were selected for replication in KORA F4 and SHIP-TREND. Genes with consistent effect directions in all four studies and p-values ≤ 0.05 in the replication cohorts were reported as candidate genes.

In total, transcripts of eight distinct genes were consistently associated with at least one of the traits SBP, PP or DBP in discovery and replication steps: *CEBPA*, *CRIP1*, *F12*, *LMNA*, *MYADM*, *TIPARP*, *TPPP3* and *TSC22D3*. Effect sizes were comparable between associations in monocytes and whole blood. In total, the candidate genes explained between 4-13%, 4-6% and 2-8% of inter-individual variance of SBP, DBP and PP, respectively.

This is the first study investigating the associations between BP traits and whole transcriptomes across different blood cell populations on a large scale. The comprehensive analyses highlight eight genes, correlated with BP.

2096M

Genetics of Plasma Lactate. *P. Balakrishnan[1], A. Tin[1], J. Pankow[2], E. Boerwinkle[3], R. Hoogeveen[4], J.H. Young[5], WH.L. Kao[1].* 1) Epidemiology, Johns Hopkins University, Baltimore, MD; 2) University of Minnesota School of Public Health, Minneapolis, MN 55454; 3) Human Genetics Center, University of Texas School of Public Health, Houston, TX 77030; 4) Department of Medicine, Baylor College of Medicine, Baylor College of Medicine, Houston, TX 77030; 5) Department of Medicine, The Johns Hopkins University, School of Medicine, Baltimore, MD 21205.

Background: Type 2 diabetes affects 25.8 million of the U.S. population. Recent studies have shown that increased plasma lactate levels are associated with insulin resistance and type 2 diabetes. Plasma lactate is a measure of oxidative capacity, the ability to meet increased energy demand. The genetics of plasma lactate have not been well-characterized. **Materials and Methods:** We performed a genome-wide association study (GWAS) in the Atherosclerosis Risk in Communities (ARIC) study - an ongoing prospective epidemiological study conducted in four U.S. communities. Plasma lactate (mg/dl) was measured from blood collected at Visit 4, at which the participants' age ranged between 54-73. Single nucleotide polymorphisms (SNPs) were also genotyped from ARIC blood samples and imputation was done using the 1000 Genome (NCBI build 37 - hg 19) reference. Linear regression models of single nucleotide polymorphisms (SNPs) were assessed for association with log transformed plasma lactate, separately in European Americans and African Americans. Models were additionally adjusted for age at Visit 4, sex, ARIC center, body mass index, waist circumference and significant principal components of ancestry. The p-values from the European American GWAS SNPs with minor allele frequency >0.01 were used as input for pathway based analysis using MAGENTA. **Results:** Four SNPs reached genome-wide significance (P < 5*10-8) in the European Americans and were also associated in the African Americans based on locus specific threshold. The four SNPs are located within GCKR gene, involved in glucokinase inhibition, and PPP1R3B gene, involved in glycogen synthase inhibition. The vasoconstriction and vasodilation regulation pathway was the pathway most statistically significantly associated with plasma lactate. **Conclusions:** The genetics of plasma lactate may provide important clues regarding the pathogenesis of type 2 diabetes.

2097S

Identification of three novel genetic variations associated with electrocardiographic traits (QRS duration and PR interval) in East Asians. *J.E. Lim[1], K-W. Hong[2], J.W. Kim[3], Y. Tabara[4], H. Ueshima[5,6], T. Miki[7], F. Matsuda[4], Y.S. Cho[8], Y. Kim[2], B. Oh[1].* 1) Kyung Hee University, Seoul, South Korea; 2) Center for Genome Science, Korea Centers for Disease Control & Prevention, Division of Epidemiology and Health Index, Chungcheongbuk-do, South Korea; 3) Department of Internal Medicine, Inje University Ilsan Paik Hospital, Gyeonggi-do, South Korea; 4) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 5) Center for Epidemiologic Research in Asia, Shiga University of Medical Science, Otsu, Japan; 6) Department of Health Science, Shiga University of Medical Science, Otsu, Japan; 7) Department of Geriatric Medicine, Ehime University Graduate School of Medicine, Toon, Japan; 8) Department of Biomedical Science, Hallym University, Chuncheon, Gangwon-do, South Korea.

The electrocardiogram has several advantages in detecting cardiac arrhythmia - it is readily available, noninvasive, and cost-efficient. Recent genome-wide association studies have identified single nucleotide polymorphisms that are associated with electrocardiogram measures. We performed a genome-wide association study using Korea Association Resource data for the discovery phase (phase 1, n = 6,805) and 2 consecutive replication studies in Japanese populations (phase 2, n = 2,285; phase 3, n = 5,010) for QRS duration and PR interval. Three novel loci were identified: rs2483280 (PRDM16 locus) and rs335206 (PRDM6 locus) were associated with QRS duration, and rs17026156 (SLC8A1 locus) correlated with PR interval. PRDM16 was recently identified as a causative gene of left ventricular noncompaction and dilated cardiomyopathy in 1p36 deletion syndrome, which is characterized by heart failure, arrhythmia, and sudden cardiac death. Thus, our finding that a PRDM16 SNP is linked to QRS duration strongly implicates PRDM16 in cardiac function. In addition, C allele of rs17026156 increases PR interval (beta±se, 2.39±0.40 ms) and exist far more frequently in East Asians (0.46) than in Europeans and Africans (0.05 and 0.08, respectively).

2098M

Identify genetic risk factors for coronary collaterals in a genetically diverse population. *Z. Liu[1], L. Wang[1,2], N. Vasudeva[2], P.J. Goldschmidt-Clermont[3], M.A. Pericak-Vance[1,2], D.M. Seo[3], G.W. Beecham[1,2].* 1) Human Genetics, John T. Macdonald Foundation, University of Miami, Miami, FL; 2) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 3) Division of Cardiology, Miller School of Medicine, University of Miami, Miami, FL.

Coronary artery disease (CAD) is the leading cause of death worldwide. Important risk factors for CAD have been identified, but they fail to explain why some occlusive CAD patients with fewer or no natural bypass collateral vessels have a lower rate of survival compared with patients with greater number of such vessels. Genetic components in combination with multiple other factors may play roles, but the genetic pre-determinant for developing coronary collateral vessels in patients is unclear, mostly due to study populations too small to represent all CAD patients. Under-represented populations, especially Hispanics, will provide additional information to study the genetic contribution to coronary collaterals formation. Toward this end, we ascertained 879 occlusive CAD patients from South Florida region and conducted a genome-wide association study. These patient samples represented the large diversity of populations in South Florida region, including Hispanic, African American and Caucasian populations. Each patient had coronary stenosis greater than 50% of at least one coronary artery. Among these occlusive CAD patients, 473 had collaterals and 406 did not. We genotyped patient samples on Affymetrix 6.0 platform and performed standard QC, followed by single genetic variant test using covariates of gender and principal components as previously identified. For single genetic variant test, SNPs rs16890371 (p = 5.58E-06), rs17497524 (p = 8.75E-06), rs10894957 (p = 8.86E-06) and rs4892107 (9.79E-06) showed top association with collateral binary trait. Furthermore, we performed gene-based test with VEGAS, and the top associated gene was C1QTNF9B (p= 2.00E-07), which passed Bonferroni corrected genome-wide significant threshold (2.8E-06); other top associated genes were SNAI3 (p= 5.31E-05), CTU2 (p = 1.53E-04), RNF166 (p = 1.56E-04) and PIEZO1 (p = 2.2E-04). In conclusion, we identified potential common SNPs and genes that are associated with collaterals in a genetically diverse population. We will further investigate genetic effects of rare variants in coronary collaterals. Studying the genetic components of coronary collaterals allows us to understand why there are differences between occlusive CAD patients of developing a sufficient collateral circulation, therefore make it possible to develop adjuvant treatment to increase CAD patient survival rate through manipulating collateral process.

2099S

Meta-analysis of Variants on the Exome Chip in 120,700 Individuals of European Ancestry Identifies Multiple Rare and Common Loci for Blood Pressure. *C. Liu[1], A.T. Kraja[2], J.A. Smith[3], J.A. Brody[4], N. Franceschini[5], A.C. Morrison[6], Y. Lu[7], S. Weiss[8], P.L. Auer[9], X. Guo[10], A.Y. Chu[11], J. Jakobsdóttir[18], J.C. Bis[4], W. Zhao[3], K. Tsosie[12], N. Amin[13], H. Mei[17], C. Newton-Cheh[14], W. Palmas[10], Y. Liu[15], R.J.F. Loos[7], T.L Edwards[12], U. Völker[8], M. Fornage[16], C.M. van Duijn[13], I. Borecki[2], G. Ehret[19,20], V. Guðnason[18,21], D. Chasman[11], D. Levy[1], CHARGE Plus Exome Chip Blood Pressure Consortium.* 1) Framingham Heart Study, NHLBI/NIH, Framingham, MA. 01702, USA; 2) Division of Statistical Genomics, Department of Genetics and Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63110, USA; 3) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA; 4) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98195, USA; 5) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; 6) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston TX 77030, USA; 7) The Charles Bronfman Institute for Personalized Medicine, Icahn Sinai School of Medicine at Mount Sinai, New York, NY 10029, USA; 8) Ernst-Moritz-Arndt-University of Greifswald, Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, Friedrich-Ludwig-Jahnstr. 15A, 17487 Greifswald, Germany; 9) School of Public Health, University of Wisconsin-Milwaukee, Milwaukee WI 53211, USA; 10) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA; 11) Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02215, USA; 12) Center for Human Genetics Research, Vanderbilt Epidemiology Center, Department of Medicine, Vanderbilt University, Nashville, TN 37212, USA; 13) Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, 3015 CN Rotterdam, the Netherlands; 14) Broad Institute of Harvard and MIT, Massachusetts General Hospital, Boston, MA 02142, USA; 15) Department of Epidemiology & Prevention, Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; 16) Institute of Molecular Medicine and School of Public Health Division of Epidemiology Human Genetics and Environmental Sciences, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 17) Center of Biostatistics & Bioinformatics, Department of Medicine, University of Mississippi Medical Center, MS 39216, USA; 18) Icelandic Heart Association, Kopavogur, Iceland; 19) McKusick-Nathans Institute of Genetic Medicine Johns Hopkins University School of Medicine Baltimore, MD 21205, USA; 20) Cardiology,Geneva University Hospitals,Rue Gabrielle-Perret-Gentil, 4,1211 Genève 14 Switzerland; 21) University of Iceland , Reykjavik, Iceland.

Introduction: High blood pressure (BP) or hypertension is a major risk factor for cardiovascular disease. Previous genome-wide association studies (GWAS) have identified > 60 genetic loci for BP. However, most of the published variants are non-coding and explain a small proportion of BP heritability. Goal: To investigate the role of coding variation on interindividual variability in BP (systolic, diastolic, mean arterial, and pulse blood pressure). Methods: We genotyped ~250,000 mostly coding variants in 120,700 individuals of European ancestry using the Infinium HumanExome BeadChip. Association analyses were conducted using the seqMeta software in each of 15 participating studies and results were combined by fixed effect meta-analysis. Array-wide statistical significance was defined as $p < 5 \times 10\text{-}7$. Conditional analyses were performed to identify independent signals at known or novel loci. Results: Our analyses identified new associations of common variants (minor allele frequency or MAF > 5%) at known and novel loci, in addition to low-frequency (MAF 1- 5%) and rare (MAF < 1%) variants at novel loci. Significant novel associations ($p < 5\times10\text{-}7$) were discovered for common variants in several genes with influences on kidney function, erythrocyte measures, and cardiomyopathy: RGL3 (missense), PCNXL3 (intronic), TBX2 (intronic), TNNT3 (intronic), PHF19 (upstream), RABEPK (missense), and PRDM16 (missense). Among low-frequency and rare variants, novel associations were found at NPR1 (MAF 0.013, missense, $p = 8\times10\text{-}7$ DBP) and DBH (MAF 0.005, missense, $p = 7\times10\text{-}7$ MAP). Npr1 knockout has previously reported to result in higher BP in a mouse model, and Npr1 is a receptor for atrial and brain natriuretic peptides which regulate blood volume and pressure. In addition, we replicated 38 loci ($p < 7.0\times10\text{-}4$) reported in prior GWAS and candidate gene studies. Pathway analyses of top variants ($p < 1\times10\text{-}4$) showed enrichment for the angiotensin system. Several significant variants are in high LD with variants that reside in regulatory marks, which may affect promoter activity or distal regulation. Conclusion/Discussion: We have identified novel associations of low-frequency and rare variants, as well as common variants in coding genomic regions for BP phenotypes that may further elucidate novel biology related to BP regulation. Our findings suggest there are additional associations to be discovered by leveraging data from coding regions and low frequency variants.

2100M

GWAS of Serum (25(OH) Vitamin D levels in a Punjabi Sikh Diabetic Cohort. *B.R. Sapkota[1], G. Priamvada[1], A. Bjonnes[2], P.R. Blackett[1], R. Saxena[2], D.K. Sanghera[1].* 1) Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA.

Vitamin D, 25-hydroxy-vitamin D [25(OH)D] deficiency is associated with multiple medical complications, including musculoskeletal, inflammatory, malignancy and cardiovascular risk. Earlier, we reported a high prevalence of 25(OH)D deficiency (<50nm/L) associated with a significant increase in cardiovascular risk factor including obesity, hypertension, and type 2 diabetes in Punjabi Sikhs from the Asian Indian Diabetic Heart Study (AIDHS); and 82% of diabetic patients were 25(OH)D deficient compared to 64% healthy controls. In this study, we investigated the relationship of serum 25(OH)D with cardiometabolic risk and performed the first GWAS and meta-analysis to identify gene variants influencing 25(OH)D deficiency in AIDHS. Our discovery GWAS of 1,616 individuals (842 cases and 774 controls) was followed by Stage 1 replication of 67 top signals ($P<10^{-5}$) in an additional Sikhs (n=2,386). On combined discovery and Stage 1 meta-analysis (n= 4,002), we identified a novel locus represented by *IVL* gene in association with serum 25(OHD) levels ($\beta = 0.10$, $p=3.1 \times 10^{-6}$) after adjustment with age, sex and type 2 diabetes status. These findings are currently being replicated in other independent larger datasets. Our results also confirmed a previously reported association with 25(OH)D represented by rs2282679 at the *GC* gene (4q12.13) with ($\beta = -0.13$, $p=3.0 \times 10^{-4}$) in Sikhs. The *IVL* gene, located on chromosome 1q21.3, encodes involucrin. Both *IVL* and *GC* genes are implicated in the synthesis of vitamin D. Genetic variation in the 25(OH) D pathway may have a significant impact in the observed deficiency of vitamin D, which could be synergistically contributing to increased cardiometabolic risk in this population. Taken together, the identification of a new locus in a vitamin D pathway gene for influencing serum 25(OH)D levels may have important clinical implications and should be replicated in other independent sample sets. This study was supported by NIH grants -R01DK082766 (NIDDK) and NOT-HG-11-009 (NHGRI), and VPR Bridge Grant (OUHSC).

2101S

A smoking stratified meta-analysis of peripheral arterial disease identifies associations and interactions with SNPs near genes implicated in nicotine dependence. *N.R. van Zuydam[1,2], M. de Andrade[3], G. Thorleifsson[4], E. Vlachopoulou[5], E. Ahlqvist[6], V. Salomaa[7], S. Gretarsdottir[4], H. Colhoun[2], I.J. Kullo[3], GoLEAD and SUMMIT.* 1) WTCHG, Oxford University, Oxford, Oxfordshire, United Kingdom; 2) Medical Research Institute, University of Dundee, UK; 3) College of Medicine, Mayo Clinic, Rochester MN, USA; 4) deCODE Genetics, Reykjavik, Iceland; 5) University of Helsinki, Finland; 6) Lund University, Malmö, Sweden; 7) National Institute for Health and Welfare, Helsinki, Finland,.

Peripheral arterial disease (PAD), a morbid complication of nicotine addiction, affects more than 200 million people worldwide. Smoking increases the risk of PAD up to 10 fold and there is evidence that there may be different pathways contributing to the disease in smokers compared to non-smokers, in whom the major risk factor is diabetes. There are two established loci for PAD from genome wide association studies: a genome wide significant (p<5e-8) signal near *CHRNA3* and another signal (p<5e-7) in the 9p21.3 region. The lack of additional associations may be due to the highly polygenic architecture of PAD. To identify variants specifically associated with PAD in smokers and non-smokers we performed a smoking status stratified meta-analysis and also performed a smoking interaction analysis to identify variants that may interact with smoking status to modify the risk of PAD. We combined summary statistics in a fixed effects meta-analysis for 2,356,286 SNPs from 4,544 PAD cases and 30,404 PAD controls - of whom 3,535 cases and 20,212 controls were smokers. We also performed an interaction analysis of allelic effects in smokers vs. non-smokers. The top SNP in smokers was rs1051730 (p=7.2e-7), near *CHRNA3*, that is established for PAD and is associated with nicotine dependence and smoking quantity. The previously reported signal in the 9p21.3 locus, represented by rs10757269, was also associated with PAD in smokers (OR=1.18, p=2.5e-6). The smoking interaction analysis highlighted signals that had large effect sizes in smokers but little or no effect in non-smokers. Rs2076156 (OR= 2.4, p=1.1e-4, phet=3.5e-6), near *BIK*, rs12593396 (OR=1.46, p=4.2e-6, phet=1.3e-5), near *ADAMTS17* and rs11214800 (OR=1.3, p=2.6e-4, phet= 2.7e-5), near *HRT3A* were associated in PAD in smokers and showed evidence for interaction with smoking status. The expression of *BIK* is directly affected by cigarette smoking, while SNPs near *ADAMTS17* have been shown to interact with smoking status to affect blood pressure and SNPs in *HRT3A* have been associated with smoking quantity. Our analyses stratified on smoking status highlight biologically plausible signals related to smoking and nicotine dependence. Such analyses may help to elucidate differences in disease aetiology in smokers compared to non-smokers.

2102M

A Comprehensive 1000 Genomes-based GWAS of Coronary Artery Disease. *H.-H. Won[1,2,3,4], A. Goel[5], M. Nikpay[6] on behalf of the CARDIoGRAMplusC4D Consortium.* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 2) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA; 3) Department of Medicine, Harvard Medical School, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 5) Division of Cardiovascular Medicine, Radcliffe Department of Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 6) Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Canada.

Enhanced coverage of genetic variation provided by a 1000 Genomes imputation training set may allow detection of additional common and uncommon genetic variants associated with coronary artery disease (CAD). The CARDIoGRAMplusC4D consortium has assembled 60,801 cases and 123,504 controls for a GWAS of CAD with 1000 Genomes imputed data. Forty-eight case-control studies contributed to the meta-analysis, 77% of European ancestry, 13% and 6% of South and East Asian ancestry with smaller numbers of Hispanic and African Americans. Cases were identified with an inclusive CAD phenotype (e.g. MI, ACS, chronic stable angina). After filtering very low frequency and poorly imputed variants, 9.4 million variants (91% SNPs, 9% indels) of which 29% were lower frequency variants ($0.005 < MAF < 0.05$) were available for meta-analysis. Significant associations ($p < 5 \times 10^{-8}$) with CAD were detected for 2,163 variants (8% indels) with a low false discovery rate (q-value $< 2.2 \times 10^{-4}$). These variants mapped to 42 loci, of which 33 had been previously reported with GWAS levels of significance. Key SNPs in *LPA* and *APOE* with low MAF (0.008 - 0.03) and moderate imputation quality (0.80 - 0.93) that had eluded previous GWAS showed strong associations. Nine novel loci were identified with imputation qualities ranging from 0.78 to 0.99 and MAF from 0.02 to 0.45. Conditional association analyses identified 88 variants in GWAS significant loci that explained 13% of CAD heritability; 85 variants in 73 loci (q-value < 0.05) explain a further 7% of CAD heritability. Thirteen percent of the jointly associated variants were indels compared to 9% of all the variants in the meta-analysis ($p = 0.04$). The variants preferentially clustered within or close to genes ($p = 1.3 \times 10^{-6}$); only two variants directly disrupt protein function (APOE and ZC3HC1). Ninety-two percent of the associated variants were common ($MAF > 0.05$) suggesting that low-frequency, high-penetrance variants explain little of the heritable risk of CAD. Our findings demonstrate the value of using a global imputation training set to integrate data from different ancestry groups and enhance coverage of low allele frequency and incompletely tagged variants.

2103S

Exome-wide association study of blood lipid levels and positive selection of lipids associated genes in Asian population. *H. Zhang[1], Y. Yang[2], W. Zhou[1], M. Yang[1], O.L. Holmen[3], K. Holmen[3], P. Sham[4], K.S.L. Lam[5], K. Mohlke[6], E. Chen[1], W. Gao[7], Y. Zhang[2], M. Xu[7], S. Ganesh[1,9], C.J. Willer[1,8,9].* 1) Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, Michigan, USA; 2) First Hospital, Peking University, Beijing, China; 3) HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger, Norway; 4) Centre for Genomic Sciences, Jockey Club Building for Interdisciplinary Research; State Key Laboratory of Brain and Cognitive Sciences, and Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; 5) Department of Medicine, The University of Hong Kong, Hong Kong SAR, China; 6) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 7) Third Hospital, Peking University, Beijing, China; 8) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA; 9) Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA.

Cardiovascular disease (CVD) is the leading cause of death globally. Circulating blood lipids are heritable, treatable, risk factors for cardiovascular disease. Over the past six years, genome-wide association studies (GWAS) have identified 157 loci associated with plasma lipid levels. However, most of the GWAS were conducted on individuals of European descent, and so the degree to which knowledge gained from these studies is applicable to other populations has not been systematically investigated.

To have a better understanding of the global genetic variants associated with blood lipid levels, we conducted an exome-wide association study in samples from the Chinese population. In this study, we tried to answer three questions. (1) Whether the known lipids associated genes/variants were associated with blood lipids levels and have similar effect size in Asian samples? (2) Are there any novel genes/variants, especially low-frequency and rare variants, that are associated with blood lipid levels? (3) Whether the lipids-associated genes were affected by positive selection during evolution?

In the analysis, many of the known lipids associated loci/variants were also associated with blood lipids in our samples, and have similar effect size. Heritability analysis indicated that low-frequency and rare Asian-specific variants contribute a considerable proportion of the heritability, however, we could not identify any novel genes or variants that reached exome-wide significance. The rare variant, p.Leu3548Ile in *DNAH17* approached exome-wide significance ($P = 3 \times 10^{-7}$) but we could not identify sufficient copies of this variant in three additional East Asian studies to assess association in an independent cohort. We also identified several Asian-specific variants in known lipids associated loci (*APOA5*, *APOB* and *PCSK9*) that also reached genome-wide significant level. Using data from the 1000 Genomes Project, we found that lipid genes have a higher Fst between Asian and European populations than expected by chance. By estimating Fay and Wu's *H*, we also find these lipid genes tend to have a higher proportion of derived alleles in the Asian population. These suggested some lipids genes diverged quickly in Asian populations and may be driven by positive selection during evolution.

2104M

Genome-wide association analysis for chronic venous disease. *D. Ellinghaus[1], A. Fiebig[1], E. Ellinghaus[1], F. Lai[1], P. Krusche[2], N. Frings[2], C. Schreiber[3], S. Nikolaus[3], C. Gieger[4], W. Lieb[5], P. Rosenstiel[1], S. Schreiber[1,3], A. Franke[1].* 1) Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Germany; 2) Capio Mosel-Eifel-Clinic, Bad Bertrich, Germany; 3) Department of General Medicine, University Hospital Schleswig-Holstein (UKSH), Campus Kiel, Germany; 4) Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 5) PopGen Biobank, Section Epidemiology, Christian-Albrechts-University of Kiel, Germany.

Chronic venous disease (CVD) is one of the most common vascular abnormalities, especially in Northern and Western Europe, with a prevalence of up to 20% in Northern and Western Europe. According to the CEAP guidelines, CVD comprises those clinical entities that are characterized by visible venous ecstasies but which are not associated with an identifiable mechanism of venous dysfunction. CVD represents a summary term comprising patients with either simple varicose veins (VV) or chronic vein insufficiency. The underlying etiology and pathophysiology of CVD are still poorly understood. A genetic component has been proposed for many years, strongly indicated by reports on familial clustering and twin studies. The narrow-sense heritability of VV and CVD has recently been estimated to equal 18.5%; and 17.3%, respectively, in a large sample of affected nuclear families from Germany. So far, no gene has been identified as a susceptibility factor for PVV. To further our understanding of the genetic etiology of CVD, we undertook the hitherto first genome-wide association study (GWAS) for this multifactorial condition. In total, 2,265 patients were recruited through local outpatient services in Germany. We performed genome-wide SNP genotyping of 323 unrelated German CVD cases and 4,619 healthy control individuals using Affymetrix SNP arrays, followed by replication in 1,942 cases and 3,101 controls from Germany. We identified two new susceptibility loci with genome-wide significant evidence of association ($p<5\times10^{-8}$), as well as one additional suggestive locus with $p<2\times10^{-7}$. From these analyses we are able to draw insight into the biological underpinnings of CVD.

2105S

Novel genetic determinants associated with blood lipid concentration changes and coronary artery disease in European adults. *T.V. Varga[1,2,3], A. Kurbasic[1], Y. Chen[1], G. Hindy[4], S. Gustafsson[5], M. Aine[6], P. Eriksson[6], D. Shungin[1,7,8], A. Ali[1], C.-A. Schulz[4], P. Nilsson[9], G. Hallmans[10], O. Melander[11], I. Barroso[12,13,14], P. Deloukas[14,15,16], L. Lind[17], E. Ingelsson[5,18], M. Orho-Melander[4], F. Renström[1,10], P.W. Franks[1,8,19], CARDIoGRAMplusC4D Consortium.* 1) Department of Clinical Sciences, Lund University, Malmö, Skåne, Sweden; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 3) The Broad Institute of Harvard and MIT, Boston, MA, USA; 4) Department of Clinical Sciences, Diabetes and Cardiovascular Disease - Genetic Epidemiology, Skåne University Hospital, Malmö, Sweden; 5) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 6) Department of Clinical Sciences, Division of Oncology and Pathology, Skåne University Hospital, Lund University, Lund, Sweden; 7) Department of Odontology, Umeå University, Umeå, Sweden; 8) Department of Public Health & Clinical Medicine, Umeå University Hospital, Umeå, Sweden; 9) Department of Clinical Sciences, Lund University, Skåne University Hospital, Malmö, Sweden; 10) Department of Biobank Research, Umeå University, Umeå, Sweden; 11) Department of Clinical Sciences, Hypertension and Cardiovascular Diseases, Skåne University Hospital, Malmö, Sweden; 12) NIHR Cambridge Biomedical Research Centre, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; 13) University of Cambridge, Metabolic Research Laboratories Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; 14) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 15) Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia; 16) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 17) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 18) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 19) Department of Nutrition, Harvard School of Public Health, Boston, MA, USA.

Background: Large scale genome-wide association studies (GWAS) have identified over a hundred loci associated with blood lipid levels, and related cardiovascular traits in a cross-sectional setting. However, very few genetic association studies have focused on long-term changes in blood lipid levels. Methods and Results: Participants from the Northern Swedish GLACIER Study were genotyped with the MetaboChip array. Main effects of variants mapping to previously established lipid loci (n=9,898 SNPs) were estimated with linear regression models by fitting the lipid levels at follow-up as the dependent variables and conditioning on the corresponding baseline lipid measure (N=3,492 for total cholesterol change (ΔTC) and N=2,209 for triglyceride change (ΔTG)). We sought replication in three other Swedish studies (MDC, PIVUS, ULSAM; total N~4,000) and conducted an *in silico* look-up for the top ranking lipid-change associated variants in the CARDIoGRAMplusC4D Consortium (N~190,000) to explore whether these variants also associate with coronary artery disease (CAD). In total, 227 variants associated with either ΔTC or ΔTG in GLACIER ($P<0.01$). Of these 227 variants, 189 were only associated with longitudinal changes and not with baseline lipid levels ($P>0.05$). Four single nucleotide polymorphisms (SNPs) in or near two loci, *apolipoprotein A-V* (*APOA5*) and *spalt-like transcription factor 3* (*SALL3*) reached genome-wide significance ($P<5\times10^{-8}$) and one SNP in the proximity of *apolipoprotein B* (*APOB*) reached study-wide significance ($P<5\times10^{-6}$) for TG change in GLACIER. 27 variants (in 11 loci) had statistically significant ($P<0.05$) pooled effects for lipid changes after meta-analyzing the 4 cohorts' effect estimates; seven of these loci have not been discovered in previous cross-sectional meta-analyses. A novel genetic association with CAD was observed for rs2000999 in *haptoglobin-related protein* (*HPR*), which was associated with ΔTC ($P=5.1\times10^{-3}$) and CAD ($P=3.6\times10^{-4}$). Conclusion: In the GLACIER Study, we found 4 genome-wide significant hits for TG changes. Furthermore, through meta-analysis, we identified 11 regions associated with TC and TG changes and one novel locus, *HPR*, for CAD.

2106M

Identification of a predictive/prognostic genetic signature in Chagas Cardiomyopathy: A systems biology approach. *C. Chevillard[1,8], A.F Frade[2,3,9], L.R.P. Ferreira[2,3,9], S. Cabantous[1,9], L. Laugier[1,9], P.C. Teixeira[2,3], B.M. Ianni[2], C.W. Pissetti[4], B. Saba[5], L.H. Tzu Wang[5], A. Kuramoto[2,3], L.G. Nogueira[2,3], P. Buck[2], F. Dias[6], M. Baron[2,3], I. Navarro[2,3], A. Schmidt[6], E. Donadi[6], J.A. Marin-Neto[5], M. Hirata[5], M. Sampaio[5], A. Fragata[5], E.A. Bocchi[7], A.N. Stolf[2], A.I. Fiorelli[2], R.H. Barros Santos[2], V. Rodrigues[4], A.C. Pereira[1,2], J. Kalil[2,3,7], E. Cunha-Neto[2,3,7,8].* 1) INSERM, U906, Université Aix-Marseille, Faculté de Médecine, 27 bd Jean Moulin 13385 Marseille cedex 06 France; 2) Heart Institute (InCor), University of São Paulo School of Medicine, Av. Dr. Enéas de Carvalho Aguiar, 44 Bloco 2 9° andar- 06504-000 - São Paulo, SP-Brazil; 3) Institute for Investigation in Immunology (iii), INCT, São Paulo SP, Brazil; 4) Laboratory of Immunology, Universidade Federal do Triângulo Mineiro, 40 Frei Paulino, 48036-180 Uberaba, MG-Brazil; 5) Instituto de Cardiologia Dante Pazzanese (São Paulo) Avenida Dr. Dante Pazzanese, 600 Vila Mariana - 05013-909 - São Paulo, SP-Brazil; 6) School of Medicine of Ribeirão Preto (FMRP), University of São Paulo, Av. Bandeirantes, 4900 - Monte Alegre 15059-900 - Ribeirão Preto, SP -Brazil; 7) Division of Clinical Immunology and Allergy, University of São Paulo School of Medicine, 06504-000 São Paulo, SP -Brazil; 8) have an equal contribution; 9) have an equal contribution.

Chagas disease, due to Trypanosoma cruzi, occurs exclusively in the Americas, particularly in poor, rural areas of Mexico, Central America, and South America. An estimated 300,000 new cases and 50,000 fatilities occur per year. Chronic Chagas disease cardiomyopathy (CCC) is an inflammatory cardiomyopathy that affects approximately 30% of infected individuals. Familial aggregation of CCC suggests that there might be a genetic component to disease susceptibility. We hypothesize that expression of pathogenetically relevant genes and proteins in the myocardial tissue of CCC patients is controlled by genetic polymorphisms. Systems biology is an approach that aims to model and discover interactions and emergent properties of complex biological systems, which is addressed using quantitative measures and by rigorous integration of “omics” data. On heart biopsy samples, we performed a gene expression analysis and found 4142 differentially expressed genes between CCC patients and controls. Among these genes, there are immunoglobulins, cytokines, chemokines, integrins. There are also collagen, caspase and transcription factors. Several lncRNA were also differentially expressed such as MIAT that been previously associated to an increased risk of myocardial infarction. Several pathways were associated to disease such as: 1) Calcium Regulation in the Cardiac Cell 2) Striated Muscle Contraction 3) Arrhythmogenic right ventricular cardiomyopathy. The main genes were confirmed by single gene qPCR. Using IPA software, we performed a network/pathway analysis. We identified 5 main networks associated to heart function, apoptosis, and necrosis. In order to explain the gene expression variations we looked for miRNA on the same heart samples. We found 39 miRNAs differentially expressed between CCC and controls. There are good relationships between miRNA expression levels and targeted gene expression levels. For some of the gene described in our main networks the gene expression variation may be explained by the methylation mean of the promoter region. Statistical analyses were conducted to merge proteomic, gene expression miRNA and methylation data. On the main gene candidates, case control studies were set up to identify common variants associated to severe disease. On two nuclear families including several cases and asymptomatic controls, we are performing exome sequencing to identify rare variants shared only by the severe cases.

2107S

Exome sequencing identifies interacting cytoskeletal genes with mutations in congenital heart disease. *A. Manickaraj[1,2], L.C.A. D'Alessandro[1,2], K. Brand-Arzamendi[1,2], D. Manase[1,2], O.J. Buske[3], S. Al Turki[4,5], Z.Q. Lu[6], P. Sharma[6], R. Errol[1,2], P. Patel[1,2], C. Kinnear[1,2], M. Brudno[2], A.O. Gramolini[7,9], M.E. Hurles[4], I.C. Scott[7,8,9], S. Mital[1,2,7,9].* 1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Peter Gilgan Centre for Research and Learning, 686 Bay Steet, Toronto, Ontario, M5G 0A4, Canada; 2) Department of Pediatrics, The Hospital for Sick Children, University of Toronto, 555 University Avenue, Toronto, Ontario, M5G 1X8, Canada; 3) Department of Computer Science, University of Toronto, Toronto, Ontario, M5S 3H5, Canada; 4) Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK; 5) Department of Pathology, King Abdulaziz Medical City, Ar Rimayah, Riyadh 14611, Saudi Arabia; 6) Department of Physiology, University of Toronto, Toronto, Ontario, M5G 1L6, Canada; 7) Program in Developmental and Stem Cell Biology, The Hospital for Sick Children, Peter Gilgan Centre for Research and Learning, 686 Bay Steet, Toronto, Ontario, M5G 0A4; 8) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, M5S 1A8, Canada; 9) Heart & Stroke/Richard Lewar Centre of Excellence, University of Toronto.

The genetic etiology is known in less than 20% of congenital heart disease (CHD). The objective of our study was to identify genes enriched for novel and rare damaging variants in pathways crucial for cardiovascular development. We performed whole exome sequencing (Agilent Sureselect V3, Illumina HiSeq) in 6 Caucasian families, with14 affected members with CHD, 9 non-affected members and 147 unrelated Caucasian CHD subjects. Gene and protein expression data of early expressing cardiac genes were obtained and genes with novel/rare variants on highly conserved locations in these genes that were predicted damaging were prioritized. Further analysis of these 56 prioritized genes using DAVID (GO-term enrichment analysis) identified strong functional interactions amongst 9 of these genes for extracellular matrix (ECM) function. Novel/rare pathogenic variants, inherited and de novo, in these 9 genes were seen in 12.7% of the CHD cases. 3 cases showed co-segregation of mutations in more than one gene. Expression knockdown of individual genes in zebrafish embryos caused abnormal cardiac phenotypes. More severe phenotypes were seen with concurrent knockdown of more than one gene supporting a multigenic etiology for CHD. Our study identifies novel variants in functionally interacting ECM genes that contribute individually or in combination to CHD.

2108M

Transcriptome-wide single-cell gene expression and genetic variation analyses of metabolic stress response in macrophages reveal functionally relevant genetic cues in atherosclerosis. *S. Sauer[1], C. Fischer[1], M. Metsger[1], S. Bauch[1], M. Boettcher[1], P. Grote[1], M. Kliem[1], A. Ibanez[2], R. Steinhoff[2], R. Zenobi[2].* 1) Max-Planck-Institute for Molecular Genetics, Berlin, Germany; 2) ETH Zurich, Zurich,. Switzerland.

Metabolic and cardiovascular diseases represent major health problems these days. Recent unbiased genome-wide (epi-) genetic variation and transcriptomics studies led to a number of potentially causative factors that require further study. Beyond possibly gaining mechanistic insights, mapping of disease-associated or correlated factors to molecular networks and functional pathways resulted in a comprehensive, holistic view of complex physiological homeostasis. Here, we show the potential of applying single-cell transcriptomics to discover characteristics and so far unexplored genetic modules involved in atherogenic and inflammation processes. Using single-cell RNA-sequencing and related approaches we analysed metabolically stressed macrophages and lipopolysaccharide-treated inflamed cells. Thereby, we could define and calculate the degree of noise and flexibility in cellular transcription of genetic information, suggesting genetic buffering in stressed cells to counteract environmental stress. Furthermore, by linking genetic modules observed by co-expression events in cell sub-populations with genetic variation data derived from unbiased association-studies we could validate and functionally integrate causative factors and pathways involved in atherosclerosis processes. For example, we identified transcriptional modules, which include well known functional categories (such as genes involved in fatty acid and cholesterol metabolism, immune response, cytokine activity, NOD-like receptor signaling pathway) and previously uncharacterized co-regulated gene groups (such as members of the ZNF gene family that are associated with inflammation, and other unexpected genes including various non-coding RNAs). Importantly, our results hint to networks based on the concerted action of several transcription factors targeting specific sites in the (variable) genome, which can be exploited to develop rationale individualized therapeutic intervention. In general, integrating genetic variation and powerful functional genomics analyses provides valuable insights in the mechanisms of genetic interaction to better understand disease-processes and susceptibility, and provides avenues to develop tailor-made disease-treatments.

2109S

Identification of pathway/genes associated with Bicuspid Aortic Valve in three Caucasian Cohorts. *M. Heydarpour[1], J.G. Seidman[2], S. Prakash[3], D. Milewicz[3], Y. Bosse[4], G. Limongelli[5], S.C. Body[1], the Bicuspid Aortic Valve Consortium (BAVCon).* 1) Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) University of Texas, Houston, TX; 4) University of Laval, Quebec, Canada; 5) Monaldi Hospital, Naples, Italy.

OBJECTIVE: Bicuspid aortic valve (BAV) is observed in 0.6-1% of humans and is associated with premature aortic stenosis, aortic aneurysm and other congenital cardiac anomalies. BAV has been associated with high-penetrance rare mutations in a few families, but the majority of BAV is sporadic with unidentified inheritance. The embryologic molecular mechanisms causing BAV are unknown. The objective of this study was to identify pathways associated with sporadic BAV. METHODS: We genotyped 456 adult US Caucasians with BAV (339 males, 117 females) using Illumina Omni2.5. 5,216 Caucasians from three dbGaP cohorts were used as controls and typed with either Omni2.5 or Omni5 arrays. Quality control of the genotype data were performed using PLINK. After QC, we examined 603,499 SNPs mapped to 500kbp regions covering 16,775 genes. We used i-GSEA4GWAS and ICSNPathway software (Zhang et al. 2010, 2011) to identify gene sets associated with BAV. This method integrates gene set enrichment, linkage disequilibrium analysis, functional SNP annotation and pathway-based analysis. Permutation test and False Discovery Rate (FDR) were used for multiple testing corrections. RESULTS: Of 1,843 gene sets selected using canonical pathways, 45 were associated with BAV (FDR<0.05). Four pathways - Small GTPase mediated signal transduction (SGMST), Ras protein signal transduction (RPST), Cell recognition (CR), and Regulation of G-protein coupled receptor protein signalling (RGPCRPS), were identified as candidate pathways for BAV. The top pathway SGMST contained 44 significant genes (FDR<0.001) of which, the SNP rs2842895 in the regulatory region (promoter) had high-LD (>0.8) with rs4585612. Both SNPs were present in the region of gene RREB1. Two other SNPs (rs3360 and rs9806942) near OPCML and RGS11 were also identified as candidate causal SNPs, respectively. CONCLUSION: We identified potential candidate genes for BAV in the SGMST, RPST, CR, and RGPCRPS pathways. Both genes RREB1 and OPCML have known role in cell-contact/cell-adhesion. The top candidate gene RREB1 is a zinc finger protein involved in Ras/Raf-mediated cell differentiation by enhancing calcitonin expression. RREB1 is essential to reduce cell-cell adhesion when epithelial cells are undergoing the dynamic changes for cell shape (Melani et al., 2008). RREB1 may be biologically related to BAV but further replication and biological investigations are required to understand this potential new mechanism leading to BAV.

2110M

The multi-tissue cis-eQTL landscape in coronary artery bypass grafting patients: the STockholm Atherogenesis & Gene Expression (STAGE) study. *K. Nguyen[1], C. Molony[1], R. Dobrin[2], H. Zhou[1], L. Chen[1], E. Schadt[2,3], D. Reilly[1], J. Björkegren[2,3,4].* 1) Department of Genetics & Pharmacogenomics, Early Development and Discovery Sciences, Merck Research Laboratories, Boston, MA; 2) Department of Genetics & Genomic Sciences, Institute of Genomics & Multiscale Biology, School of Medicine, Mount Sinai Hospital, New York, NY; 3) Clinical Gene Networks AB, Karolinska Science Park, Stockholm, Sweden; 4) Cardiovascular Genomics Group, Division of Vascular Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden.

Genome-wide association studies (GWAS) have identified well over a hundred loci showing genome-wide significance for association with cardiovascular and coronary artery disease (CAD). However, with the exception of few examples such as the *SORT1* locus, the mechanisms underlying these associations are not well understood. With the aim to better comprehend the relevance of genetic variation for CAD, we here dissect the consequences of single nucleotide polymorphisms (SNPs) on gene expression in carotid plaques and other CAD-relevant tissues. Blood and biological specimen from 100 patients who underwent coronary artery bypass grafting (CABG) surgeries were collected for genome-wide genotyping and expression profiling. We performed expression quantitative trait loci (eQTL) analysis using Matrix_EQTL of 7M *1000G*-imputed variants and 40K protein-coding probe sets in individual tissues and across 7 tissue types: atherosclerotic arterial wall (AAW, n=63), internal mammary artery (IMA, n=74), liver (n=72), skeletal muscle (n=74), whole blood (n=94), subcutaneous (n=60) and visceral (n=82) fat. We identified 370,871 unique eQTLs in 7,194 genes (FDR≤0.05), present in at least 1 tissue. Initial screening against the GWAS catalog among Caucasians yielded 489 GWAS SNP-anchoring eQTLs; in 60% of these (n=302), reported genes do not contain the corresponding eQTL gene. This shows that the biological functions of association loci are often obscure, and that eQTL studies are informative resources in unveiling additional insights in the association of complex phenotypes. To understand the roles of the identified eQTLs in coronary plaques in the 100 CABG patients in STAGE, we first restricted our attention to the top ~5% (FDR≤5×10^{-10}, n=19,333) that are putatively functional (contain SNPs in coding/regulatory or ncRNAs regions, n=7,157) and are not present in population negative controls (GTEx v.1) (335 genes in 7,076 eQTLs). Gene ontology enrichment analysis of the top cis-genes confirms the importance of metabolic/biosynthetic processes of lipids, lipoprotein and fatty acids (in liver, muscle and whole blood), and antigen processing and peptidase activities in the context of inflammatory response (in liver and visceral fat). This work provides an improved understanding of the underlying genetic mechanisms and multi-tissue eQTL landscape in CAD patients, paving the path forward to informed molecular validations for the disease biology.

2111S

Metabolomic profiling identifies markers of cardiac atrial septal defects. *H. Wang[1,4], D. Craig[3], C. Haynes[3], W. Kraus[3], S. Shah[2,3].* 1) Department of Surgery, Duke University Medical Center, Durham, NC; 2) Department of Medicine, Duke University Medical Center, Durham, NC; 3) Duke Molecular Physiology Institute, Duke University, Durham, NC; 4) Department of Computational Biology and Bioinformatics, Duke University, Durham, NC.

Atrial septal defect (ASD) is among the most common congenital heart malformations in adults. Although genetic variants have been found to be associated with ASD, the molecular etiology of the disease is not fully understood. Metabolic profiling of peripheral blood has been used in cardiovascular research to discover new risk factors and pathways associated with disease. We performed metabolomic profiling of adult patients in the CATHGEN biorepository at Duke University. Targeted, quantitative profiling was performed of 68 metabolites using flow injection tandem mass spectrometry (MS). Two groups of patients were profiled: 40 ASD cases and 40 non-ASD controls matched on sex, race and renal function. Principal component analysis and factor analysis identified a total of eight latent factors corresponding to highly correlated metabolite groups. Elevated levels of three factors were associated with ASD: long-chain dicarboxylacylcarnitines (factor 2, p = 0.045), short-chain dicarboxylacylcarnitines (factor 3, p = 0.01) and branched-chain amino acids (factor 6, p = 0.047). In multivariate logistic regression, factors 2 and 3 were independently associated with ASD, after adjusting for baseline medical conditions such as history of myocardial infarction, number of diseased coronary arteries and history of diabetes (factor 2, odds ratio 6.6; 95% CI 1.6 - 36.9, p = 0.02; factor 3, OR 2.19; 95% CI 1.14 - 4.68, p = 0.03). Using these two factors as predictors, probit regression with cross-validation using random resampling achieved a mean accuracy of 0.72, sensitivity of 0.70, precision of 0.76 and F1 score of 0.72. Both of these metabolite factors have been implicated in CAD and incident cardiovascular events, although neither had been profiled in ASD or other congenital heart diseases. Variances in these metabolites may reflect similar differences in mitochondrial fatty acid oxidation, modification in peroxisomal metabolism or endoplasmic reticulum stress that may occur with ASD.

2112M

A Gene Network approach to Rare Variant analysis. *T.G. Richardson[1], N.J. Timpson[1], C. Campbell[2], T.R. Gaunt[1].* 1) MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; 2) Intelligent Systems Laboratory, University of Bristol, Bristol, United Kingdom.

Background: Current endeavours in rare variant analysis are typically underpowered when investigating signals from individual genes. We undertook a novel approach to rare variant analysis by utilising knowledge regarding interactions and relationships between sets of genes. Methods: Using the resource STRINGdb we identified 263,666 human protein-protein interactions that had experimental evidence and applied the ClusterOne algorithm to construct 2,099 Gene Networks. The variant effect predictor (VEP) was used enrich our analysis by identifying 116,621 unique variants in the UK10K sequence data predicted to be nonsynonymous. Variants within the same network were collapsed together and analysed with lipid traits using the sequence kernel association test (SKAT) after applying a threshold of 5% and 1% minor allele frequency. Results: Three networks provided evidence of association after correcting for multiple comparisons and were evaluated further using single gene analyses and permutation tests. The most statistically robust evidence found that variants from within 3 genes (PDE5A, MRVI1 and NPR1), whose products were known to interact, collectively contributed to an observed signal with LDLc levels (P = 5.30 x 10-6). Conclusion: We have undertaken a novel approach to rare variant analysis which examines the combined effect of variants from across genes which have been shown to interact according to experimental evidence. We identified networks which provide evidence of association with lipids traits which suggest further analysis of polygenic effects across networks and pathways should prove vital in better understanding the genetic architecture of complex disease.

2113S

WES reporting of mutations from cardiovascular "actionable" genes in clinical practice: a key role for UMD knowledgebases. *A. PINARD[1,2], N. HANNA[3,4], C. GUIEN[1,2], L. FAIVRE[5,6], G. JONDEAU[7,4], C. BOILEAU[3,4], C. BEROUD[1,2,8], G. COLLOD-BEROUD[1,2].* 1) Inserm UMR_S910, Marseille, France; 2) Aix Marseille Université, GMGF, 13385, Marseille, France; 3) Department of Genetics, Bichat University hospital, 46 rue Henri Huchard, 75018 Paris France; 4) Inserm U1148, Bichat University hospital, 46 rue Henri Huchard, 75018 Paris; 5) Centre de Référence Anomalies du Développement et Syndromes Malformatifs et FHU TRANSLAD, CHU Dijon, Dijon, France; 6) EA 4271 Génétique des Anomalies du Développement, Université de Bourgogne, Dijon, France; 7) Cardiology Department, Bichat University Hospital, 46 rue Henri Huchard, 75018 Paris; 8) APHM, Hopital d'Enfants de la Timone, Département de Génétique Médicale et de Biologie Cellulaire, 13385, Marseille, France.

High throughput next generation sequencing (NGS) such as whole exome sequencing (WES) are being rapidly integrated into clinical practice. The use of these techniques leads to unexpected findings for which decisions about the reporting to the patient need to be made. In the USA, the American College of Medical Genetics and Genomics (ACMG) recently published recommendations for the reporting of incidental findings in clinical WES and WGS for 58 "actionable" genes. Among these, seven are involved in Marfan Syndrome And Associated Pathologies (MSAAP) including familial thoracic aortic aneurysms and dissections and Loeys-Dietz syndromes associated with mutations in *FBN1, TGFBR1* and *2, ACTA2, SMAD3, MYH11* and *MYLK*. MSAAP are autosomal dominant disorders and are recommended to be reported in clinical practice because medical flollowup and treatment can improve survival and quality of life. In an effort to standardize information regarding mutations in the *FBN1* gene, we developed in 1995 a locus specific database with the generic system called Universal Mutation Database (UMD). Subsequently, databases for *TGFBR2* and *FBN2* genes were been created. All are now recognized as international references. To be exhaustive and facilitate NGS analysis, we have now developed databases for the *ACTA2, SMAD3, MYH11* and *MYLK* genes. They contain all known mutations collected from literature and through direct collaborations with diagnostic laboratories: *FBN1* (3200 entries), *TGFBR2* (300 entries) and the new *TGFBR1* (125 entries), *ACTA2* (203 entries), *SMAD3* (61 entries), *MYH11* (44 entries) and *MYLK* (13 entries). Each mutation is annotated at gene, protein and clinical levels. Several tools are also available for studying and extracting data of interest as well as complex algorithms to predict: pathogenicity of missense variations; consequences of variations on splicing signals, or to search for genotype-phenotype correlations. These databases are updated regularly and curated by experts. They are accessible at: http://www.umd.be/. Databases of variants are now essential in the context of NGS as more and more diagnostic laboratories worldwide are using these technologies without specific expertise for each gene. Relying on UMD knowledgebases, they could rapidly collect relevant information for data interpretation, report more accurate results and save time.

2114M

Sequencing of 2,000 Norwegians to evaluate genetic architecture of lipid and MI-associated variants. *C. Willer[1,2,3], W. Zhou[1], O. Holmen[4], H. Zhang[1], J. Chen[1], D. Hovelson[2], M. Boehnke[5], G. Abecasis[5], K. Hveem[4].* 1) Int Med & Human Genetics, University of Michigan, Ann Arbor, MI; 2) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 4) HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger, Norway; 5) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

Genome-wide association scans have been successful at identifying novel loci and genes associated with blood lipid levels and coronary artery disease. However, these approaches have not adequately tested the role of low frequency variants. To attempt to identify low frequency variants with large effect on risk of myocardial infarction or impact on blood lipid levels, we performed low-pass whole genome sequencing of 2,000 individuals from Nord-Trondelag, Norway. We selected individuals with medical-record confirmed MI with early-onset (< 58 years for men and < 68 years for women). For each MI case, we selected one sex- and birth-year matched healthy control from 70,300 potential controls by excluding individuals with a variety of cardiovascular and metabolic conditions. We opted to perform low-pass whole genome sequencing for the following reasons; i. increased power relative to higher-depth whole genome sequencing in fewer samples, ii. utility of whole genome sequence data for imputation into GWAS samples relative to exome-sequence data and iii. to assess low frequency non-coding variation. Our first data freeze of 1,237 individuals (626 MI cases and 611 controls) did not identify any novel genes or genetic variants associated with lipids or MI, but did identify association at several known loci. Furthermore, we identified eight low-frequency non-coding variants (freq < 5%) with $P < 5\times10^{-7}$ (including chr4:76188636 near PARM1 with $P = 7\times10^{-8}$ with total cholesterol) that were not previously evaluated in large-scale GWAS studies and may represent novel low frequency large-effect loci.

2115S

Whole exome sequencing highlights the importance of the CRELD1 interactome in atrioventricular septal defect in Down syndrome. *C.M. Ackerman[1], H. Li[2], D. Klinedinst[2], R. Polk[2], S. Blackshaw[2], H. Corbitt[1], R.H. Reeves[2], C.L. Maslen[1].* 1) Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR; 2) Department of Physiology and the Institute for Genetic Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD.

Atrioventricular septal defect (AVSD) is a congenital heart defect (CHD) frequently associated with Down syndrome (DS). Children with DS account for 65% of all cases of AVSD and have a 2000-fold increased risk of the defect. However, since half of all children with DS have a normal heart, trisomy 21 alone is insufficient to cause AVSD. Missense mutations in *CRELD1* are associated with AVSD in both euploid and DS populations and since disease-causing mutations tend to cluster in the same biochemical pathways, we hypothesized that AVSD-associated mutations would occur in genes within the CRELD1 interactome.

To define a CRELD1 interactome, we used a protein array to identify directly interacting proteins. We then assessed whole exome sequence data from cases and controls to test if rare variants in genes encoding proteins that interact with CRELD1 contribute to the risk of AVSD in DS. The study cohort was comprised of individuals with DS and a complete AVSD (cases) and individuals with DS and a normal echocardiogram (controls), including individuals of European Ancestry and African Americans. Missense variants were characterized as damaging, benign, or unknown based on PolyPhen2 and MutPred analyses.

Of the 33 genes we evaluated, 18 genes had at least one damaging rare variant. The majority of these genes have not been implicated in the pathogenesis of CHD, with the exception of *PTPN11*, a gene in which we previously reported mutations in individuals with non-syndromic AVSD. We detected 23 damaging variants in cases and only 3 in controls ($p<0.0001$), with the majority being missense variants. Other case-specific variants identified included one recurrent frameshift mutation in *RAB25* and a premature termination codon in *EDIL3*. There was no significant skewing of variants between genders, and inclusion of African American-specific variants did not falsely elevate the percentage of case-specific deleterious variants.

We observed that almost 17% of all cases of DS with AVSD had a damaging variant in a *CRELD1* pathway gene. The fact that we see mutations in the same genes in both euploid and DS with AVSD supports our hypothesis that the same genes involved in the pathogenesis of AVSD in DS, also contribute to AVSD in the euploid population. Further functional analysis of these *CRELD1* pathway variants will provide insight into the mechanisms that underlie heart defects during development.

2116M

Gene-centric association tests applied to cardiovascular disease using whole genome sequencing. *MAA. Almeida[1], J. Peralta[1,2], JW. Kent[1], TM. Teslovich[3], G. Jun[3], C. Fuchsberger[3], AR. Wood[4], A. Manning[5], TM. Frayling[4], P. Cingolani[6], TW. Blackwell[3], R. Sladek[7], TD. Dyer[1], AG. Commuzie[1], HHH. Goring[1], L. Almasy[1], MC. Mahaney[1], DM. Lehman[8], JE. Curran[1], G. Abecasis[3], R. Duggirala[1], J. Blangero[1].* 1) Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, WA, Australia; 3) University of Michigan, Ann Arbor, MI, USA; 4) University of Exeter, Exeter, United Kingdom; 5) Broad Institute, Boston, MA, USA; 6) McGill University, Montreal, Canada; 7) Montreal Diabetes Research Institute, Montreal, Canada; 8) University of Texas Health Center at San Antonio, San Antonio, TX, USA.

The advent of whole genome sequencing provides a unique opportunity to improve our understanding the genetic variation underlying complex diseases. As part of the T2D-GENES Consortium, we have directly sequenced 590 individuals (and accurately imputed another 448 members) from 20 large Mexican American pedigrees aiming the investigation of rare genetic variants contribution to the type 2 diabetes development. Those individuals are part of the SAFS (San Antonio Family Study) and have been extensively phenotyped during the 25 years of this project. The immense number of identified SNV imposes new statistical and analytical barriers that require the development of alternative approaches for screening of potential causal genes and pathways. We employed a variance component-based single degree-of-freedom test using an empirical gene-specific genetic relationship matrix (GRM) as the focal covariance kernel. The empirical gene-specific GRM (the GSGRM) utilizes any set of chosen variants identified in a gene or gene-pathway of interest and provides the correlation of the dosage vectors between individuals. The efficacy of this new focal kernel for explaining phenotypic variance is tested by the use of a single-degree freedom likelihood ratio test. Gene pathway definitions were obtained from the latest CytoScape database release and a GSGRM was estimated for each one of 1191 genes that constitute 9 CVD (Cardiovascular Diseases) related pathways. Each empirical kernels was tested against a set of CVD-related traits collected in the SAFS cohort. As might be expected, our top association was the LPA gene and the concentration of the apo(a) protein coded by this gene (p-value = 2.1×10-56); this is a known case of cys-regulation and serves as a positive control of our approach. We also observed a strong association between the gene PCSK9 and CVD itself (p-value = 8.2×10-5). The PCSK9 gene product plays a central role in cholesterol and fatty acid metabolism and is a major target for the new generation of anti-lipid drugs. Each variant in the gene was independently tested and two non-coding variants achieved genome-wide significance and a non-synonymous variant, rs11583680, located in first exon of this gene (p-value = 1.6×10-4) was also suggestive. Our results suggest that a pathway-filtered simple gene-based test may be useful for reducing the search space and aid the detection of specific functional variants influencing complex disease-related phenotypes.

2117S

Somatic Activating GNAQ Mutations are Frequent in Capillary Malformations (Port-Wine Stains). *M. Amyere[1], N. Revencu[2], A. Dompmartin[3], R. Healers[1], L. Boon[1,4], M. Vikkula[1].* 1) Laboratory of Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Brussels, Belgium; 2) Centre for Human Genetics, Cliniques Universitaires Saint-Luc, Université catholique de Louvain, Brussels, Belgium; 3) Department of Dermatology, Université de Caen Basse Normandie, CHU Caen, France; 4) Centre for Vascular Anomalies, Division of Plastic Surgery, Cliniques Universitaires Saint-Luc, Université catholique de Louvain, Brussels, Belgium.

Vascular malformations are localized defects of vascular development that are classified into capillary, venous, arterial, lymphatic, and combined anomalies. Capillary Malformation (CM) is the most common affecting cutaneous capillary vessels in 0.3% of newborns. These lesions can be localized or diffuse. CM most often occurs as an isolated and sporadic feature. Syndromic forms include Sturge-Weber syndrome, Klippel-Trenaunay syndrome and Parkes Weber syndrome. There is often associated soft tissue and/or bony overgrowth. In the autosomal dominant capillary malformation-arteriovenous malformation (CM-AVM), CMs increase in number with age. A third of the CM-AVM patients have an associated fast-flow anomaly, most often located in the head and neck region. CM-AVM is caused by haploinsufficiency, most likely combined with a tissular second-hit, p120-RasGAP, the protein product of RASA1. A series of Sturge-Weber syndrome and sporadic CMs were shown to harbor a non-synonymous somatic single-nucleotide variant in GNAQ, encoding guanine nucleotide binding protein. We assessed 33 fresh-frozen lesions by both Sanger sequencing on cDNA and allele-specific PCR on genomic DNA, for the presence of the c.548G>A (p.Arg183Gln) hot-spot mutation. Sixteen tissues with CMs; 5 lesions with Sturge-Weber-Syndrome; 4 lesions with CMs with overgrowth; 3 lesions with capillaro-veineuse malformation, 2 lesions with CM-AVM; 2 lesions with CM with dilated-vein and one lesion with hyperkeratotic cutaneous capillary venous malformation. A somatic mutation was detected in 18 lesions (54,5%); 5/5 (100%) of Sturge-Weber Syndrome lesions and 13/16 (81.2%) CMs. This mutation was not detected in the fifteen other lesions with other CM phenotypes (45.5%). This could be explained by the fact that normal cells can mask low-number mutant cells in the resected lesion. Targeted massive parallel sequencing of the complete coding sequences of GNAQ and RASA1 was therefore performed on the 16 negative lesions using Ion AmpliSeq Panel on PGM. Vertical coverage varied from 400x to 1550x. We did not find further evidence for the presence of the GNAQ hot-spot mutation, considering 1% as the limit for background. Whole exome sequencing is now performed in these negative tissues to identify additional causative genes. This should improve genetic classification and lead to better diagnosis, evaluation of genotype/phenotype correlation, and development of targeted therapies. (miikka.vikkula@uclouvain.be).

2118M

Molecular-genetic factors in normal and pathological angiogenesis. *I.I. Dimova[1], V.G. Djonov[2], A. Makanya[2], R. Hlushchuk[2], V. Stefanovic[3], M. Polenakovic[4], M.G. Mihailova-Hristova[1], R. Vazharova[5], L. Balabanski[5], S. Ivanov[5], D.I. Toncheva[1].* 1) Medical University Sofia, Sofia, Bulgaria; 2) Institute of Anatomy, University of Bern, Switzerland; 3) Faculty of Medicine, University of Nis; 4) Macedonian Academy of Sciences and Arts; 5) Laboratory Genomics, Malinov Hospital.

Angiogenesis is involved in many physiological processes, but also is a hallmark in the pathology of many diseases (cancer, ischemia, atherosclerosis, inflammatory diseases), in wound healing and in tissue regeneration. We aimed in studying the molecular-genetic mechanisms of angiogenesis in normal and pathological conditions. Chicken area vasculosa was used as a model for investigation of essential molecular pathways - Notch, ephrinB2/EphB4 and SDF-1/CXCR4 signaling. Glomerular sclerosis and interstitial fibrosis are characteristic angiogenic pathological changes in kidneys of patients with Balkan Endemic Nephropathy (BEN). DNA from blood samples of 22 patients was analyzed by NGS, searching for genes of predisposition. We demonstrated that Notch inhibition disturbed vessel stability and led to pericyte detachment followed by extravasation of mononuclear cells. The mononuclear cells contributed to formation of transluminal pillars with sustained intussusceptive angiogenesis resulting in a dense vascular plexus without concomitant vascular remodeling and maturation. Inhibition of ephrinB2 or EphB4 signaling induced some pericyte detachment and resulted in up-regulation of VEGFRs but with neither an angiogenic response nor recruitment of mononuclear cells. Notably, Tie-2 receptor was down-regulated, and the chemotactic factors SDF-1/CXCR4 were up-regulated only due to the Notch inhibition. Using NGS we found genetic variants with deleterious/damaging effects in three genes - CELA1, HSPG2 and KCNK5. These mutant genes in BEN patients encode proteins involved in basement membrane/extracellular matrix and vascular tone, tightly connected to process of angiogenesis. The studies contributed to elucidation of molecular mechanisms of normal angiogenesis and genetic factors predisposing to abnormal angiogenesis in glomerular sclerosis.

2119S

Genetic Variation of Scavenger Receptor Class B Type I (*SCARB1*) and Plasma Lipid Traits: An Association Study in a Nigerian Population. *V. Niemsiri[1], X. Wang[1], M.M. Barmada[1], C.H. Bunker[2], F.Y. Demirci[1], M.I. Kamboh[1].* 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Graduate School of Public Health, Pittsburgh, PA; 2) Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.

Abnormal lipid-lipoprotein levels are known to be the risk factors for cardiovascular disease. One of the genes involved in lipid metabolism is scavenger receptor class B type I (*SCARB1*), which has the main function in a selective uptake of cholesteryl esters in reverse cholesterol transport. Using Sanger sequencing, we examined 13 exons and exon-intron boundaries of *SCARB1* in 95 African Blacks from Nigeria having extreme high-density lipoprotein cholesterol (HDL-C) levels. Sequencing analysis revealed a total of 83 variants (MAF ≥5%, n = 32; MAF <5%, n = 51). Common tagSNPs (MAF ≥5%) and uncommon/rare variants identified by sequencing (n = 78), plus 69 additional common tagSNPs from HapMap project covering the entire gene and 2 additional relevant variants from the literature were genotyped in the entire Nigerian sample (n = 788). A total of 137 successfully genotyped variants were then analyzed for their associations with various lipid traits. The most significant associations were observed with HDL-C levels. Single-site analysis identified 20 common *SCARB1* variants significantly (P <0.05) associated with one or more lipid traits, of which the most significant was the rs11057851 SNP associated with HDL-C levels (P = 0.0043). Optimal sequence kernel association test (SKAT-O) revealed also a nominally significant association of rare variants (MAF ≤1%, n = 23) with HDL-C levels (P = 0.0475). In summary, our results suggest the genetic contribution of *SCARB1*—both common and rare variations—to lipid-lipoprotein levels (especially to HDL-C) and highlight the importance of this gene in lipid metabolism in humans.

2120M

Identification of novel genetic mutations causing familial hypercholesterolaemia among Saudi Arabian population. *F.A. AL-ALLAF[1,2], M. Athar[1,2], Z. Abduljaleel[1,2], A. Bouazzaoui[1,2], M.M. Taher[1,2], R. Own[1], A.F. AL-ALLAF[3], I. AboMansoor[1], Z. Azhar[1], F.A. BA HAMMAM[1,2], H. Abalkhail[4], A. Alashwal[4], S.S. Siddiqui[5].* 1) DEPARTMENT OF MEDICAL GENETICS, FACULTY OF MEDICINE, UMM AL-QURA UNIVERSITY, MAKKAH, Saudi Arabia; 2) Science and Technology Unit, Umm Al-Qura University, Makkah, Saudi Arabia; 3) Faculty of Medicine, Alfaisal University, Riyadh, Saudi Arabia; 4) King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia; 5) FACULTY OF DENTISTRY, UMM AL-QURA UNIVERSITY, MAKKAH, Saudi Arabia.

Familial hypercholesterolemia (FH) is a major risk factor for the development of Coronary heart diseases (CHD). The disease is hereditary in an autosomal dominant manner. Mutation(s) and/or deletion(s) in the LDL-receptor (LDLR) gene, in the apolipoprotein B-100 (ApoB) gene, or in the proprotein convertase subtilisin kexine 9 (PCSK9) gene are the most known causative mutations. The most accurate and unequivocal method for FH diagnosis is by molecular genetic testing of suspected cases and therefore this study aims to identify the genetic defects causing FH in Saudi population and to develop a diagnostic test for detecting such mutations. Forty four truly identified FH subjects were genetically screened for mutations in the LDLR, ApoB and PCSK9 genes through direct PCR-sequencing using ABI 3500 genetic analyzer. We identified four common mutations in coding sequences of the LDLR gene, one mutation in ApoB gene and three mutations in PCSK9 gene. Among the identified LDLR gene mutations, two have been reported previously and two are novel. In addition, a base substitution in the splice acceptor site of LDLR intron 11 and a second mutation was also observed in LDLR intron 11. We have also identified a single previously reported heterozygous mutation in ApoB gene. Similarly we found three mutations in PCSK9 gene which are also reported earlier. Moreover, two heterozygous mutations c.658-7C>T and c.799+3A>G in the PCSK9 gene intron 4 and 5 respectively, have also been identified. This knowledge is important for optimizing cholesterol lowering therapies and mutational analysis diagnostic test. In addition, these data contribute to the understanding of the molecular basis of FH in Saudi Arabia.

2121S

Whole Exome Sequencing of 350 LVOTO Cases Reveal Novel Candidate Genes for Congenital Cardiovascular Malformations. *J. Belmont[1,3], A. Li[2], N. Hanchard[1,3], M. Azamian[1], S. Fernbach[1], G. Zapata[3], P. Hernandez[3], D. Parekh[3], W. Franklin[3], D. Penny[3], C. Fraser[4], R. Gibbs[1], E. Boerwinkle[2].* 1) Dept Molec & Human Gen, Baylor Col Med, Houston, TX; 2) Div Epidemiology, School of Public Health, U Texas Health Sci Center, Houston, TX; 3) Dept Pediatrics, Baylor Col Med, Houston, TX; 4) Dept Surgery, Baylor Col Med, Houston, TX.

BACKGROUND: Congenital cardiovascular malformations (CVMs) occur in 5-8/1000 live births. Left Ventricular Outflow Tract Obstruction (LVOTO) defects, which include Hypoplastic Left Heart Syndrome (HLHS), Aortic Valve Stenosis (AS), Coarctation of the Aorta (CoA), Interrupted Aortic Arch Type A (IAAA) and Shone Complex, comprise 15-20% of severe CVMs. Genetic contributions to the causes of LVOTO defects are complex including chromosomal abnormalities, genomic disorders, and syndromes caused by single gene mutations. Familial clustering of cases and an increased risk of LVOTO in first degree relatives are consistent with single gene and oligogenic inheritance. The fact that most cases are sporadic also suggests a potential role for de novo mutations. METHODS AND RESULTS: We performed whole exome sequencing of 350 unrelated LVOTO cases without extracardiac malformations. Loss-of-function (LOF) variants (premature stop, splice, and frameshift indel) were prioritized using stringent criteria (depth >30x for nucleotide substitutions, >60x for indels) and compared to those observed in over 5000 ethnically-matched controls from the ARIC cohort sequenced using the same Illumina HiSeq platform. Novel LOF variants, not observed among controls, were highly overrepresented among genes with putative involvement in cardiac development, including 49 genes with overlapping cardiac phenotypes in model organisms (zebrafish, mouse); two known Mendelian genes previously implicated in congenital heart malformation (NF1, TBX20), and 1 gene previously hypothesized to be a candidate cardiac developmental gene. Targeted re-sequencing of these variants within cases and direct family members is currently ongoing to determine mode of inheritance and to validate genotypes. CONCLUSION: Our study implicates LOF mutations in a broad spectrum of known and postulated cardiac genes in the pathogenesis of LVOTO and illustrates the value of integrating an appropriately-matched control group, model organism bioinformatics, and a traditional family-based approach for discovering genes implicated in rare human disorders.

2122M

Low-frequency coding variation in *DNAH11* is associated with Sudden Cardiac Arrest among African Americans. *J.A. Brody[1], F. Asher[2], N. Bihlmeyer[2], X. Zhao[3], A. Mak[4], A.S. McCallion[2], T. Lumley[5], C.M. Sitlani[1], K. Rice[6], B.M. Psaty[1,7,8], R. Lemaitre[1], P.Y. Kwok[4], F.D. Kolodgie[3], D.S. Siscovick[1,7,9], N. Sotoodehnia[1,10].* 1) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) CVPath Institute, Inc., Gaithersburg, MD; 4) Cardiovascular Research Institute and Institute for Human Genetics, University of California, San Francisco, CA; 5) Department of Statistics, University of Auckland, Auckland, New Zealand; 6) Department of Biostatistics, University of Washington, Seattle, WA; 7) Departments of Medicine and Epidemiology, University of Washington, Seattle, WA; 8) Group Health Research Institute, Seattle, WA; 9) New York Academy of Medicine, New York, NY; 10) Division of Cardiology, University of Washington, Seattle, WA.

Background: Sudden cardiac arrest (SCA) is a major public health concern, particularly among African Americans where risk of cardiac arrest is higher than that of the general population, and survival is poor. While environmental factors clearly contribute to SCA risk, familial aggregation studies and molecular genetics studies of inherited arrhythmias suggest that genetic factors confer susceptibility to SCA in the general population. **Methods**: We genotyped >240,00 common and low-frequency, nonsynonymous and splice-site variants with the Illumina HumanExome chip. Our study sample of African Americans consisted of 402 SCA cases and 4673 controls from the two cohort studies, Atherosclerosis Risk in Communities study and the Cardiovascular Health Study, and one case-control study of SCA; the Cardiac Arrest Blood Study (CABS). Associations were modeled using Cox-proportional hazards models for the cohort studies or logistic regression for the case-control study and were combined using fixed effects meta-analysis. Common variants (minor allele frequency [MAF] >1%) were tested individually and rare and common variants were jointly modeled within gene using the Sequence Kernel Association Test (SKAT). We then examined our findings among 2405 cases and 2431 controls of European ancestry from the CABS study. **Results**: Gene-based SKAT analyses identified Dynein, axonemal, heavy chain 11 (*DNAH11*) to be a novel gene associated with SCA among African Americans. The findings for *DNAH11* (p= 1.16E-6, cumulative MAF=249%) reached significance after Bonferroni correction for the number of genes examined (N=17,574). The gene test result was primarily driven by two missense variants (rs147478795, His2795Asp, MAF=1.6% and rs74667361, Cys2763Arg, MAF=2.0%) in high linkage disequilibrium (r2>.8). Rs147478795 (OR=4.34; 95% CI=2.43-6.25; p=7.95E-6) is likely to impact protein function as it was predicted to be deleterious by all four bioinformatic prediction algorithms examined. Among European descent individuals, rs147478795 is rare and was found in 1 case and no controls. The gene-based SKAT result showed no evidence of association in those of European ancestry (p=0.84, cumulative MAF=238%). Other mutations in *DNAH11* cause ciliary dyskinesia type 7, which is characterized by situs abnormalities including dextrocardia. **Conclusions**: Through analyses of low-frequency coding variation, we have discovered a novel gene associated with SCA among African Americans.

2123S

Whole genome sequencing and analysis of three hypoplastic left heart syndrome trios using DNA from buccal swabs collected through the National Birth Defects Prevention Study. *K.J. Buckingham[1], M. Beightol[2], J.D. Smith[2], J.X. Chong[1], C.T. Marvin[1], B.W. Paeper[2], K. Patterson[2], L.D. Botto[3], M.L. Feldkamp[3], D.A. Nickerson[2], J. Shendure[2], M.J. Bamshad[1,2,4], University of Washington Center for Mendelian Genomics.* 1) Department of Pediatrics, University of Washington, Seattle, WA, USA; 2) Department of Genome Sciences, University of Washington, Seattle, WA, USA; 3) Department of Pediatrics, University of Utah, Salt Lake City, UT, USA; 4) Seattle Children's Hospital, Seattle, WA, USA.

Although identifying genes underlying Mendelian conditions has transformed our understanding of the genetic basis of human birth defects, the etiology for ~75% of birth defects remains unknown. To discover novel genes for isolated birth defects, we conducted a pilot study in which we performed whole genome sequencing (WGS) on three parent-child trios in which the newborn was diagnosed with hypoplastic left heart syndrome (HLHS). The samples and phenotypic information were collected as part of the National Birth Defects Prevention Study (NBDPS). As one of the largest studies of birth defects in the U.S., the NBDPS is potentially a rich resource for gene discovery. However the relatively low amount of DNA available for each participant (~1ug on average) presents a limitation in utilizing this cohort. To this end, we tested a protocol to generate high-quality WGS libraries from 225ng of genomic DNA extracted from NBDPS buccal swab specimens. Samples were sequenced to ~30X mean coverage without major modification to the WGS pipeline. We first analyzed the exome subset within the WGS data and identified novel biologically plausible candidate genes in each trio. In Trio 1 (HLHS, aortic atresia, mitral valve stenosis) we identified *de novo* missense mutations in *CABLES2* and *MVP*. *MVP* encodes major vault protein and directly interacts with PTPN11 protein. Mutations in *PTPN11* cause Noonan syndrome, which has been associated with HLHS in some cases. In Trio 2, (HLHS, atrial septal defect, mitral valve atresia), under a dominant model (a paternal uncle had HLHS), we found a frameshift mutation in *CMYA5*, which encodes cardiomyopathy associated 5 protein. A coding variant in *CMYA5* has been associated with variation in left ventricular wall thickness and cardiac hypertrophy. In Trio 3 (HLHS, aortic atresia, mitral valve stenosis, bilateral iris and retinal colobomas, partial agenesis of the corpus callosum), we identified a novel variant in *SOS1*, variants in which cause Noonan syndrome. While this finding suggests variants in *SOS1* could cause isolated CHD, this same variant was also found in a putatively unaffected parent for whom no phenotypic data were available. Next steps in the analysis of these trios will focus on non-coding variants and structural variants including CNVs. Collectively, our results demonstrate the feasibility of using samples of low quantity DNA combined with whole genome sequencing strategies for discovery of new genes underlying birth defects.

2124M

Rare variant association analysis for plasma lipids in coronary artery disease susceptibility loci. *H.K. Chheda[1], P.P. Palta[1], E.T. Tikkanen[1,6], S.M. McCarthy[2], V.S. Salomaa[3], R.D. Durbin[2], A.P. Palotie[1,4,5], T.A. Aittokallio[1], S.R. Ripatti[1,2,6].* 1) Institute for Molecular Medicine, Finland, Helsinki, Uusimaa, Finland; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom; 3) National Institute for Health and Welfare, Department of Chronic Disease Prevention, Helsinki, Finland; 4) Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) Department of Medical Genetics, University of Helsinki and University Central Hospital, Helsinki, Finland; 6) Hjelt Institute, University of Helsinki, Helsinki, Finland.

Coronary artery disease (CAD) is the leading cause of deaths worldwide. Genome-wide association studies have identified multiple common variants associated with CAD and related traits. Some of these loci mediate their effects via risk factors such as lipids or hypertension. However, most of the common loci identified so far have a relatively small effect on the disease susceptibility. The exact molecular mechanisms behind most of these loci are also unknown. Therefore, we aimed at identifying rare and low frequency variants associated with plasma lipids and other metabolites in the known CAD loci. We analyzed low pass whole genome sequenced 1217 Finnish individuals from the FINRISK cohort with deep phenotypic data for CAD risk factors including metabolomic profiles using different platforms. We undertook the analysis using sequence kernel association test (SKAT) for variants with minor allele frequency less than 1% or 5%. We performed SKAT using two approaches; window-based (20 or 50 variants in each window) and genomic region- based (3kb regions). Our preliminary results show a higher burden of CAD loci on plasma lipid concentrations using both approaches. In the known CAD loci, we observed a combined proportion of 20-25% windows associated with high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), total cholesterol (TC) and triglycerides (TG). One such region upstream of HDAC9 gene on chromosome 7 showed significant association with TG (window-base approach: P= 9 x 10-5, genomic region-based approach: P= 2 x 10-4). Fine-mapping of this region identified 3 independent low frequency variants showing association with TG: rs150895469 (frequency: 0.8%, P=2 x 10-5); var_7_18328999 (frequency: 0.2%, P=7 x 10-4); rs117533313 (frequency: 0.1%, P=4 x 10-3). The lead SNP (rs150895469) was also imputed in 9946 individuals in an independent cohort and shown to be nominally associated with CAD (HR=1.92, 95% CI 1.06-3.49, P=0.03). We will are currently extending these analyses for other risk factors and datasets. The preliminary findings have provided us with higher-resolution insights into the genetic architecture of CAD-related traits, which may later lead to the identification of rare/low frequency causal variants for CAD. In conclusion, this study will help explain the molecular mechanisms behind the loci associated with CAD.

2125S

Identification of potentially pathogenic mutations in 9 candidate genes for bicuspid aortic valve using next-generation sequencing. *N. Dargis[1], M. Lamontagne[1], N. Gaudreault[1], L. Sbarra[1], C. Henry[1], P. Pibarot[1], P. Mathieu[1], Y. Bosse[1,2].* 1) Quebec Heart and Lung Institute, Quebec City, Quebec, Canada; 2) Department of molecular medicine, Laval University.

BACKGROUND : Bicuspid aortic valve (BAV) is the most frequent congenital heart disease. Affected individuals are at greater risk of developing aortic valve stenosis as well as other valvuloaortopathies. Despite the high prevalence of BAV, its etiology and genetic origins remain elusive. To improve our understanding of the genetic components contributing to the development and progression of BAV, we sought to identify and assess the pathogenicity of all genetic variants located in 9 candidate genes for BAV (*AXIN1, EGFR, ENG, GATA5, NKX2-5, NOS3, NOTCH1, PDIA2* and *TGFBR2*) in affected French Canadians using next-generation sequencing and in silico analyses. **METHODS :**Genomic DNA from 48 French Canadians affected with BAV confirmed at surgery and having no concomitant aortic insufficiency and/or aortic aneurysm was extracted from blood buffy coat. Targeted DNA sequencing was performed using the Ion Torrent™ Personal Genome Machine™ (PGM™) from Life Technologies™. We designed an Ion AmpliSeq™ custom panel of primers covering 1,000 bp of the promoter region, the entire coding sequence and both 3' and 5' untranslated regions (UTR) for each of the nine genes. Novel variants were validated by conventional Sanger sequencing and the pathogenicity of all variants was evaluated with the Combined Annotation Dependant Depletion (CADD) method. **RESULTS :** Sequencing revealed a total of 217 genetic variants, 38 of which were novel. Ten new variants were detected in the *NOTCH1* gene, including one novel missense mutation present in one patient (p.G152S, located in exon 4). This variant was evaluated as functionally damaging and deleterious. **CONCLUSION :** We identified 38 novel genetic variants within 9 candidate genes for BAV in 48 subjects affected with the disease. One novel mutation in *NOTCH1* is predicted to be damaging and deleterious. Further analyses will be required to determine the functional impact and evaluate the frequencies of these variants in French Canadians and other populations. Identifying causal mutations is essential to increase our knowledge of the molecular mechanisms underlying this frequent cardiac pathology, which could ultimately translate into earlier diagnosis, screening for families at risk, predicting disease progression and personalizing treatments.

2126M

Whole Exome Sequencing Reveals Rare, Truncating Variants in Nuclear Envelope Genes are Present in a Large Subset of Cardio-Genetic Patients. *G.T. Haskell[1], B.C. Jensen[3,4], D.S. Marchuk[1], C. Skrzynia[1], C. Bizon[2], K. Wilhelmsen[1,2], K.E. Weck[1,5], J.P. Evans[1], J.S. Berg[1].* 1) Department of Genetics, UNC-Chapel Hill, Chapel Hill, NC; 2) Renaissance Computing Institute, Chapel Hill, NC; 3) Division of Cardiology, UNC-Chapel Hill, Chapel Hill, NC; 4) McAllister Heart Institute, UNC-Chapel Hill, Chapel Hill, NC; 5) Department of Pathology and Laboratory Medicine, UNC-Chapel Hill, Chapel Hill, NC.

The NCGENES (North Carolina Genomic Evaluation by Next-Generation Exome Sequencing) clinical trial is applying whole exome sequencing (WES) in a bedside to bench approach, in order to evaluate the use of genome-scale sequencing to identify genomic determinants of a variety of genetic disorders, including hereditary cardiac conditions. Eighteen patients suspected of having a genetic form of cardiac disease have been enrolled, and WES sequence variants analyzed against a diagnostic list of ~75 known cardiac disease genes. The diagnostic yield for clearly positive results in this cohort was 16.7 percent. In patients with negative or uncertain diagnostic results, we analyzed other rare, predicted deleterious variants that might account for their conditions. Given that mutations in the nuclear envelope (NE) gene LMNA are a major genetic risk factor for severe cardiac disease, we were particularly interested in evaluating whether deleterious variants in other NE genes might contribute to cardiac disease as well. We used bioinformatics tools to systematically annotate a number of variant characteristics, including frequency in population databases and our own in-house WES database, as well as in silico pathogenicity predictions. We identified novel or extremely rare truncating variants in NE genes in 5 out of 14 patients. Three of these are nonsense or splice site variants in genes encoding nucleoporins, and one alters the splice donor site between intron 43 and exon 44 in SYNE1, a gene associated with Emery-Dreifuss Muscular Dystrophy, that is also important for mechanotransduction and normal heart function in mice. The SYNE1 variant was identified in a patient with dilated cardiomyopathy (DCM) who underwent heart transplant at age 15, and is present in his father, who also has advanced DCM. Many of the WES-identified NE genes are expressed in the developing as well as the adult heart, and we are evaluating whether these NE genes play a role in early cardiac development. Together, these data suggest the hypothesis that integrity of the nuclear envelope is particularly important for cardiac function and that alterations in this class of genes may be an important cause of genetic cardiac disease. Genome-scale sequencing technologies provide an especially attractive strategy for surveying clinically relevant variants in cardiac disease, and provide a rich source of data for discovery-based investigations aimed at identifying novel disease-associated variants.

2127S

Initial analysis of exome sequence from 410 individuals with familial or simplex dilated cardiomyopathy. *D.J Hedges[1,2], A. Morales[1,2], D. Kinnamon[1,2], D. Wheeler[1,2], J. Shendure[4], M. Bamshad[4], D. Nickerson[4], R. Hershberger[1,2,3], University of Washington Center for Mendelian Genomics.* 1) Dorothy M. Davis Heart and Lung Research Institute, The Ohio State University College of Medicine, Columbus OH; 2) Division of Human Genetics, The Ohio State University College of Medicine, Columbus, OH; 3) Division of Cardiovascular Medicine, The Ohio State University College of Medicine, Columbus, OH; 4) Department of Genome Sciences, University of Washington, Seattle WA.

PURPOSE: Dilated cardiomyopathy (DCM) is a leading cause of heart failure. Large DCM families have facilitated DCM gene discovery, and rare variants in 40 genes have been implicated to cause familial DCM. However, the extent that simplex DCM cases also have a genetic basis remains uncertain. To further understand the genetic architecture of DCM, we performed whole exome sequencing (WES) on 410 individuals from familial and simplex pedigrees to identify novel and known DCM-causing variants. Here we report our initial analysis of variants in established DCM genes. METHODS: WES was conducted on 410 affected individuals from 281 DCM pedigrees, 105 with simplex DCM (history reported as negative) and 177 with familial DCM (confirmed by medical records). Secondary sequence processing, variant calling, and downstream filtering and prioritization were performed with GATK (v.1.6) and GEMINI. All single nucleotide variants (SNVs) and indels predicted to alter protein sequence were selected from 40 published DCM loci. Further prioritization included: the absence or frequency <0.5% in the 1000 Genomes or Exome Sequencing Project datasets 2) a GERP score of >2, or region conserved (Phastcons). Where affected family members' samples were available, variants were prioritized by sharing among affected cases. Prioritized variants were adjudicated as per ClinVar: pathogenic (P), likely pathogenic (LP), or variant of uncertain significance (VUS), and confirmed by capillary sequencing. RESULTS: Following clinical adjudication of 177 familial pedigrees, 89 (51%) had one or more variants identified; classifying probands by their highest ranked adjudicated variant yielded: 5(3%) P, 40 (23%) LP, and 44 (25%) VUS. Of 105 simplex cases, 43 (41%) had one or more variants identified, and stratifying as above yielded 1 (1%) P, 5 (5%) LP, and 37 (35%) VUS). Estimates of probands (including nonfamilial) with adjudicated variants in known DCM genes were concordant with earlier studies: TTN (truncations) (20%), LMNA (4%), TNNT2 (4%), MYH7 (3%); DSP variants (8%) were increased compared to earlier reports. Interestingly, 88 cases (21%) carried two or more adjudicated variants in different genes: 2 (n=63), 3 (n=21), 4 (n=7) across the 40 DCM genes. CONCLUSION: These results suggest a proportion of simplex DCM cases have a genetic basis and that DCM has a more complex genetic basis than is commonly appreciated, all of which may impact clinical genetic testing.

2128M

Common variants of LDLR & PCSK9 genes associated with the risk and severity of Coronary Artery Disease in Iranian patients. *S.Hamid. Jamaldini[1], M. Babanejad[1], R. Mozaffari[1], N. Nikzat[2], KH. Jalalvand[2], A. Badiei[2], H. Sanaati[1], F. Shakerian[1], M. Afshari[3], K. Kahrizi[2], H. Najmabadi[2].* 1) Cardiogenetics Research Center, Shahid Rajaie Cardiovascular Medical & Research Center, Tehran, Iran; 2) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; 3) Research Center for Modeling of Health, Kerman University of Medical Sciences, Kerman, Iran.

Coronary artery disease (CAD) is the leading cause of mortality in the world and it is expected to be the first cause of death worldwide by 2020. Genome-wide association studies (GWAS) have identified several genetic variants associated with CAD in LDLR and PCSK9 genes. This study was evaluated the possible association of common polymorphisms at LDLR and PCSK9 genes with the risk and severity of CAD in the Iranian patients. Sequencing of all 18 exons of the LDLR gene and promoter region of the PCSK9 gene was performed in 202 Iranian patients angiographically confirmed CAD and 130 healthy controls. Sullivan's scoring system was used for determining the severity of CAD in cases. Our results showed that homozygote genotypes of rs1122608 (P<0.0001), rs4300767 (P<0.005) and rs10417578 (p<0.007) SNPs have strong protective effects on the CAD. In addition, we found that rs1122608 (GT or TT) was at higher risk of three vessel involvement compared to single vessels affecting (P=0.01). The only change in the coding region was c.1620 G>A in exon 11 of LDLR gene, which leads to a change in amino acid production (Gly>Glu). Moreover, according to the known databases this alteration probably is dangerous and benign, respectively. The results from our case-control study and other studies might be explained by genetic heterogeneity in the susceptibility of CAD and ethnic differences between Asians and Caucasians. Further investigation on other SNPs in these genes is warranted to validate our findings in Iranian population.

2129S

Targeted oligonucleotide-selective sequencing of 101 genes from 147 patients with dilated cardiomyopathy in Finland. *J.W. Koskenvuo[1,2], O. Akinrinade[3], L. Ollila[4], J. Tallila[2], H. Koillinen[5], M. Gentile[2], M. Kaartinen[4], T. Ojala[6], S. Myllykangas[2,7], T. Heliö[4], T.-P. Alastalo[2,3].* 1) University of Turku, Turku, Turku, Finland; 2) Blueprint Genetics, Helsinki, Finland; 3) Childrens Hospital Helsinki, University of Helsinki, Helsinki, Finland; 4) Department of Cardiology, University Hospital Helsinki, Finland; 5) Department of Genetics, University Hospital Helsinki, Finland; 6) Department of Pediatric Cardiology, University Hospital Helsinki, Finland; 7) Institute of Biomedicine, University of Helsinki, Helsinki, Finland.

The genetic basis of dilated cardiomyopathy (DCM) is recognized but still not widely utilized in diagnostics. Hundreds of mutations in coding exons of over 50 genes have been reported to associate with isolated DCM. We applied oligonucleotide-selective sequencing and custom data analysis and interpretation pipelines to identify pathogenic base substitutions and insertions and deletions in 101 genes associated with cardiomyopathies. Our sequencing panel covered coding exons, exon-intron boundaries and known variants in intronic regions of target genes. We genetically profiled 147 well-characterized DCM patients with average clinical follow-up of 8.7 years. On average, 99.1% of the targeted regions were covered >15x. We identified pathogenic or likely pathogenic variants in 50% of the familial and 27% of the sporadic cases. These included truncating TTN mutations affecting all transcripts that were identified in 21% of the familial and 13% of the sporadic cases. Family segregation analysis revealed that all tested truncating TTN mutations co-segregated with the DCM disease. We observed a trend of less atrial fibrillation, conduction abnormalities and pacemaker/ICD/CRT related therapies in patients with TTN mutations compared to patients with other mutations or patients without identified mutations. Other major findings in the analysis were mutations in genes encoding the nuclear lamina, the desmosome and the sarcomere. Mutation positive DCM patients and especially those with a LMNA mutation had significantly more atrial fibrillation when compared to mutation negative individuals. In conclusion, our analysis revealed high diagnostic yield in the familial form of DCM. It further confirms that truncating TTN mutations are a significant contributor to familial DCM. Authors disclosure: None: OA, LO, HK, MK, TO and TH. Minor: JT, MG, SM*, TPA* and JWK*. *Founders of Blueprint Genetics, which offers gene diagnostics for cardiomyopathies.

2130M

Association of rare variants in *LDLR*, *HMGCR*, *NAT2*, *ABCA1*, and *APOA1* with plasma lipid levels; initial results from the eMERGE PGx project. *I.J. Kullo[1], S.J. Bielinski[2], D.S. Carrell[3], S. Stallings[4], J. Pathak[2], A. Gordon[5], L. Rasmussen-Torvik[6], K.F. Dohney[7], S. Volpi[8], M.D. Ritchie[9], J.C. Denny[10], C.G. Chute[2], D.M. Roden[4], D.A. Nickerson[5], G.P. Jarvik[11], D.R. Crosslin[10].* 1) Division of Cardiovascular Diseases, Department of Medicine, Mayo Clinic, Rochester, MN; 2) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 3) Group Health Research Institute, Center for Health Studies, Seattle, WA; 4) Vanderbilt University Medical Center, Nashville, TN; 5) University of Washington, Seattle, WA; 6) Northwestern University Feinberg School of Medicine, Chicago, IL; 7) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore MD; 8) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 9) Center for Systems Genomics, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA; 10) Departments of Biomedical Informatics and Medicine, Vanderbilt University, Nashville, TN; 11) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA.

Linking genome sequence data to the broad range of phenotypes in the electronic medical record (EMR) is a potential means of obtaining insights into the pathogenicity of rare variants. The Electronic Medical Records and Genomics (eMERGE) pharmacogenomics (PGx) project is using the next-generation PGRN-Seq capture reagent to sequence 84 drug response genes in 9,000 individuals across the network. In a preliminary analysis of the first 894 participants, we examined the association between rare variants in five candidate genes (*LDLR*, *HMGCR*, *NAT2*, *ABCA1*, and *APOA1*) that are implicated in lipid metabolism and plasma lipid levels (total cholesterol, LDL-C, HDL-C, and triglycerides) abstracted from the EMR at Group Health. For each trait, we ascertained the maximum level in the EMR for each participant. Of the 354 variants found in these 5 genes we identified 24 variants with evidence of high or moderate effects through *in silico* analyses using SnpEff and SeattleSeq annotation tools. For each functional variant, we assessed the differences in lipid levels between genotype groups. We describe the effect of one variant (C2177T) in *LDLR* on LDL-C levels in this cohort. The variant (rs45508991) leads to a THR/ILE amino acid change and was present in 12 participants (out of 894). Of these 12 patients, 25% were on statins vs. 42% in the remaining cohort. The mean LDL-C noted in the EMR in participants with the mutation was 128 mg/dL vs. 145 mg/dL in the remaining individuals. This was an unexpected finding. In review of the literature, ~15% of the *LDLR* mutation carriers have LDL-C levels below 75th percentile suggesting that certain mutations in *LDLR* may not affect receptor function. We believe that the C2177T mutation may belong to this 'non-pathogenic' category. Such information will be useful for variant annotation when interpreting *LDLR* sequence data. Our preliminary findings suggest that EMR-based studies may be helpful in confirming/assigning pathogenicity to rare variants identified by genome sequencing.

2131S

***TGFB3* pathogenic mutations cause MFS/LDS phenotypes and aortic aneurysms in 3 Japanese families.** *H. Morisaki[1,2], A. Ono[1,2], I. Yamanaka[1], R. Sultana[1], T. Oda[3], H. Tanaka[3], H. Sasaki[3], K. Minatoya[3], R. Matsukawa[4], T. Tsukube[4], N. Kubo[5], T. Morisaki[1,2].* 1) Bioscience and Genetics, National Cerebral and Cardiovascular Center Research Institute, Suita, OSAKA, Japan; 2) Medical Genetics, National Cerebral and Cardiovascular Center, Suita, OSAKA, Japan; 3) Vascular Surgery, National Cerebral and Cardiovascular Center, Suita, OSAKA, Japan; 4) Cardiovascular Surgery, Japanese Red Cross Kobe Hospital, Kobe, HYOGO, Japan; 5) Pediatrics, Urakawa Red Cross Hospital, Urakawa, HOKKAIDO, Japan.

To date, several genes have been identified to be responsible for hereditary aortopathies including systemic connective tissue disorders such as Marfan syndrome (MFS) and Loeys-Dietz syndrome (LDS). Some of these genetic mutations were found to cause high transforming growth factor (TGF)-β signaling. To identify new genes responsible for these diseases, 292 probands with MFS/LDS-like features negative in *FBN1*, *TGFBR1*, *TGFBR2*, *TGFB2*, or *SMAD3* mutations were subjected to exome sequencing. Three probands with family history of skeletal and/or aortic features resembling MFS/LDS were identified to have three novel mutations (p.N235fs, p.I322T, or p.R300W) in *TGFB3* gene. Case 1 was a 43y tall male with scoliosis, strabismus and a history of thoracic aortic dissection (Stanford A) at 41y with aortic root dilatation. He was treated with emergency total arch replacement followed by aortic root repair (David operation) at 43y. His mother was tall and slender, and diagnosed as MFS with a history of thoracic aortic dissection at 59y. Case 2 was a 56y tall male who had a Bentall operation for severe aortic root dilatation. His daughter and son were both tall and had chest deformities along with mild aortic root dilatation. Case 3 was a 3y female showing LDS features with arachnodactyly, hypertelorism, bifid uvula and translucent skin. Her father also had tall stature, hypertelorism and bifid uvula. Genetic analysis of family members revealed the *TGFB3* mutations segregated with phenotypes in all three cases. It was reported that *TGFB3* mutations cause familial arrhythmogenic right ventricular dysplasia and that *TGFB3* gene has an association with cleft lip and palate, hypertension, or ossification of the posterior longitudinal ligament of the spine. Murine model revealed that *Tgfb3* gene plays an important role in embryogenesis and palatogenesis. In addition, recently, it was reported that the patients with de novo *TGFB3* mutations showed MFS/LDS-like features with no or mild aortic involvement. This is the first report of *TGFB3* mutations found in familial cases with MFS/LDS-like features, and three of our patients required aortic root surgery. Based on these results, we conclude that *TGFB3* is essential not only for palatogenesis but also for systemic connective tissue disorders similar to MFS/LDS. Aortic aneurysms and dissections should be the important features of patients with *TGFB3* mutations.

2132M

Exome sequences and rare variant association for plasma lipids in over 18,900 individuals. *G.M. Peloso[1,2,3,4], J.A. Brody[5], J.R. Crosby[6,7], A. Isaacs[8], Y. Zhou[9,10], L.A. Lange[11], J.C. Bis[5], A.C. Morrison[6], A. Reiner[12,13], C. Kooperberg[12], C. van Duijn[8], S.S. Rich[14], C. Willer[15,16,17], C.J. O'Donnell[3,10,18], B.M. Psaty[5,13,19,20,21], J.G. Wilson[22], L.A. Cupples[9,10], S. Kathiresan[1-4], E. Boerwinkle[6,23] for the Lipids RVAS Working Group.* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA; 3) Department of Medicine, Harvard Medical School, Boston, MA; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 5) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA; 6) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; 7) Department of Biostatistics, Bioinformatics, and Systems Biology, The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, TX; 8) Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands; 9) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 10) National Heart, Lung, and Blood Institute (NHLBI) Framingham Heart Study, Framingham, MA; 11) Department of Genetics, University of North Carolina, Chapel Hill, NC; 12) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 13) Department of Epidemiology, University of Washington, Seattle, WA; 14) Center for Public Health Genomics and Department of Public Health Sciences, University of Virginia, Charlottesville, VA; 15) Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI; 16) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 17) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 18) Division of Intramural Research, NHLBI, US National Institutes of Health, Bethesda, MD; 19) Department of Medicine, University of Washington Medical Center, Seattle, WA; 20) Group Health Research Institute, Group Health Cooperative, Seattle, WA; 21) Department of Health Services, University of Washington, Seattle, WA; 22) Department of Physiology and Biophysics, the University of Mississippi Medical Center, Jackson, MS; 23) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

Genetic discovery using screens of common DNA sequence variants has identified many novel gene regions associated with plasma lipids but it has been challenging to pinpoint specific causal genes. Discovery using sequencing of protein-coding regions across the genome may overcome this barrier. Here, we test the hypothesis that exome sequencing in the population can identify novel genes associated with plasma levels of low-density lipoprotein cholesterol (LDL-C), triglycerides, high-density lipoprotein cholesterol (HDL-C), and/or total cholesterol. We sequenced the protein-coding regions (exome) of >17,000 genes from ~15,000 European ancestry (EA) and ~3,900 African American (AA) individuals and tested the association of plasma lipid levels with single rare coding variants or those aggregated within a gene. For gene-based tests, we implemented two methods - Sequence Kernel Association Test (SKAT) and burden - with various minor allele frequency thresholds. Functional classes of variants were aggregated into categories: loss-of-function (nonsense, splice-site and frameshift) and/or those predicted to be deleterious. seqMeta software was used for efficient gene-based meta-analysis, where each study computes association-specific statistics for each variant and genotypic covariance matrices within predefined gene regions. Score statistics and genotypic covariance matrices are combined across studies and used to construct the gene-based tests. With a sample size of 15,000 individuals, we had > 80% power to detect a gene with a cumulative minor allele frequency of 0.5% and a ½-SD unit effect at exome-wide statistical significance ($P < 2.5\times10^{-7}$). We found aggregated rare variant associations in many genes previously known to be associated with plasma lipids. For example, in EA individuals, LDL-C was associated with *LDLR* (beta=15 mg/dl; $p=2.7\times10^{-22}$ with Burden 0.1%), HDL-C was associated with *ABCA1* (beta=-4%; $p=2.4\times10^{-7}$ with Burden 1%), and triglycerides were associated with *APOC3* (beta=-37%; $p=6.4\times10^{-23}$ with Burden 1%). In AA individuals, LDL-C was associated with *PCSK9* ($p=6.7\times10^{-9}$ with SKAT 1%) and HDL-C was associated with *ABCA1* (beta=-8%; $p=4.8\times10^{-8}$ with Burden 1%). Surprisingly, despite analysis in over 18,000 individuals with exome sequences, our gene-based association analysis did not reveal any novel genes at exome-wide significance. These data highlight the challenges of rare variant association to identify novel genes for complex traits.

2133S

Rare mutations in cardiomyopathy genes are associated with takotsubo cardiomyopathy. *A.L. Siniard[1], P. Nakaji[2], J.J. Corneveaux[1], R.F. Richholt[1], R.P. Bruhns[1], E. Carlson[1], A. Courtright[1], K. Van Keuren-Jensen[1], J. Forseth[3], J.M. Zabramski[2], R.F. Spetzler[2], M.J. Huentelman[1], M.Y.S. Kalani[2].* 1) Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ; 2) Division of Neurological Surgery, Barrow Neurological Institute/Saint Joseph's Hospital and Medical Center, Phoenix, Arizona; 3) Arizona Pulmonary Specialists, Phoenix, Arizona.

Takotsubo cardiomyopathy (TC), defined as transient cardiac malfunction with characteristic left ventricular (LV) wall motion abnormalities in the setting of normal coronary vasculature, is a poorly understood complication of cerebral aneurysm rupture that most commonly affects post-menopausal women. Little is known about the etiology of this condition; however, the clinical approaches for management of TC are largely supportive and reactive to the presentation of symptoms. In this study we utilized next generation exome sequencing analysis in an effort to define genetic variation associated with and responsible for this phenomenon. We performed exome sequencing on eight female patients diagnosed with TC after aneurysmal subarachnoid hemorrhage. Unless the variant was already present in ClinVar, filtering of the identified variants was performed to limit our investigation to those candidates that passed freebayes-derived haplotypic scoring, were rare (<0.5% MAF in NHLBI GO ESP) or novel (not observed in publicly available databases including 1000 genomes, NHLBI GO ESP, and an internal TGen database), were determined algorithmically to have high potential impact on the associated protein function (Combined Annotation Dependent Depletion, CADD, score >15), were within regions of high species conservation (PhyloP > 50%), and exist in genomic loci previously associated with cardiac dysfunction. Additionally, all candidate variants were verified to be present using Sanger sequencing chemistry. Exome sequencing and analysis revealed that 7 out of 8 patients with TC after aneurysmal subarachnoid hemorrhage were heterozygote carriers of predicted deleterious mutations in established cardiomyopathy genes including MYOZ2, SDHA, ANKRD23, MYLK2, PKP2, DSP, and TTR. Our study suggests that patients with TC following cerebral aneurysm rupture harbor deleterious mutations in established cardiomyopathy genes. We postulate that these patients likely live in a compensated state of cardiac dysfunction that manifests only after the myocardium is stressed by the adrenergic surge associated with aneurysm rupture or the stress associated with the aggressive hemodynamic treatment of cerebral vasospasm. It is possible that the clinical knowledge of the presence of these mutations would change the management of patients presenting with TC.

2134M

Screening of the *PRKG1* gene in a British cohort of Thoracic Aortic Aneurysm and Dissection (TAAD) patients. *Y.B.A. Wan[1], J.A. Aragon-Martin[1], L.J. Collins[1], D.C. Guo[4], A. Sagger[2], M. Jahangiri[3], D. Milewicz[4], A.H. Child[1].* 1) Cardiovascular & Cell Sciences Research Institute, St. George's Hospital, University of London, SW170RE, United Kingdom; 2) Clinical Genetics Unit, St George's University of London, Cranmer Terrace, SW170RE, London, UK; 3) Department of Cardiothoracic Surgery, St George's Healthcare NHS Trust, London, UK; 4) Department of Internal Medicine, University of Texas, USA.

Background: A mutation in the *PRKG1* gene, which encodes for the type I cGMP-dependent protein kinase (PKG-I, [OMIM 176894]) that was published in 2013, causes decreased contraction of the vascular smooth muscle cells. This in turn can increase risk of TAAD.

Method: A total of 101 UK patients (80M: 21F, mean age 53±13) with known TAAD who did not fulfill the revised Ghent criteria for Marfan syndrome, and with no demonstrable mutations in *FBN1*, *TGFB2*, *TGFBR1/2*, *ACTA2* and *SMAD3* genes were recruited to this study. These patients were screened for the whole gene and in particular the published variant c.530G>A (p.R177Q) in exon 3 of *PRKG1*, by bi-directional Sanger sequencing. The methodology was taken from the original paper (Recurrent gain-of-function mutation in PRKG1 causes thoracic aortic aneurysms and acute aortic dissections. Guo DC et al, Am J Hum Genet. 2013).

Results: Our TAAD cohort of 101 British samples did not carry the c.530G>A (p.R177Q) mutation in exon 3 of the *PRKG1* gene. However, 2 heterozygous synonymous mutations (c.477C>T/p.T159T and c.993T>C/p.V331V; NM_001098512) were found in 2 unrelated probands. These mutations were not reported on genome databases (1000 genomes and Exome Variant Server). In-silico (mutation t@ster) predicted these to be disease causing. Splice site tools (BDGP and ESE finder) were used to assess the pathogenicity of these mutations and the causativity remains to be proven.

Conclusion: This study was negative for the recurring mutation found in 3 families from a USA cohort. Family segregation studies will be performed for the two synonymous mutations in the two families to demonstrate that they could be linked to the disorder. Samples should then assayed to determine if it increases the activity of the kinase before the variant can be considered to be disease-causing. Further screening of novel single nucleotide variations through next generation exome sequencing data in our British cohort of TAAD patients is currently underway.

You may contact the author at **P0900001@sgul.ac.uk**.

2135S

Exome sequencing identifies a homozygous splice site variation in ZBTB17 in arrhythmogenic ventricular dysplasia. *J. Wang[1,2], H. Lee[3], C. Silverstein[4], S. Nelson[1,3].* 1) Human Genetics, UCLA, Los Angeles, CA; 2) Medicine, UCLA, Los Angeles, CA; 3) Pathology and Laboratory Medicine, UCLA, Los Angeles, CA; 4) Medicine, Western University, Pomona, CA.

Background: Arrhythmogenic right ventricular dysplasia/cardiomyopathy (ARVD/C) is a genetic disorder of the myocardium characterized by progressive fibrofatty replacement that predisposes to ventricular tachycardia and sudden death. Although it classically affects the right ventricle, left and biventricular forms are increasingly recognized. A mutation in one of the 8 known genes, primarily encoding desmosomal proteins, is found in fewer than half of patients. We identified a family with onset of dilated cardiomyopathy in the 2nd or 3rd decade in four out of four siblings with progression to cardiac transplantation so far in three. Gross and histopathologic examination of the three explanted hearts demonstrated the characteristic fibrofatty replacement of ARVC/D involving predominantly the left ventricle. The family reported a high degree of consanguinity. Methods: To identify the molecular basis of these findings in the family, Clinical Exome Sequencing was carried out in the index patient. Exome variants were filtered for protein-altering rare heterozygous variants and homozygous variants to identify causative mutations. Mean oversampling of bases was 107X and at least 93.9% of the ~33.4 Mb of protein coding sequence defined by RefSeq was read to at least 10X depth. Illumina HumanOmniExpress BeadChip assay was performed on the living affected siblings to filter the candidate variants. Results: Within the shared linkage regions, three rare homozygous variants, predicted to cause missense, essential splice site, and premature stop codon changes, were identified in the index case and confirmed to be shared among the three affected siblings. Among the genes affected by the candidate variants, *ZBTB17* resides in a region that was previously implicated in dilated cardiomyopathy by association studies in European and Han Chinese populations. Conclusions: *ZBTB17* (zinc finger and BTB domain containing 17), also known as Miz-1, encodes a zinc finger protein involved in the regulation of c-myc. Our findings implicate *ZBTB17* as a new gene for ARVD/C and provide new evidence of a critical role for *ZBTB17* in cardiac function. The significance of the three variants shared by the affected siblings is to be established experimentally using animal models.

2136M

Identification of Novel Genes and Variants Associated with Hypertrophic Cardiomyopathy in Panel Negative Patients Through Targeted Sequencing of 450 Genes and Rare Variant Association Testing. *M.T. Wheeler[1,2,3], J.R. Homburger[1,4], F.E. Dewey[1,2,3], A. Pavlovic[1,2,3], H. Chaib[1,4], J.R. Priest[2,3,6], R.L. Goldfeder[1,5], D.M. Waggott[1], O. Soyinka[1], E.A. Ashley[1,2,3,4,5].* 1) Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA; 2) Stanford Center for Inherited Cardiovascular Disease, Stanford University School of Medicine, Stanford, CA; 3) Cardiovascular Institute, Stanford, CA; 4) Deparment of Genetics, Stanford University School of Medicine, Stanford, CA; 5) Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA; 6) Department of Pediatrics, Stanford University School of Medicine, Stanford, CA.

Hypertrophic Cardiomyopathy (HCM) is a genetic disorder of the heart muscle leading to increased heart wall thickness and risk of sudden death. HCM is typically inherited in an autosomal dominant fashion. Clinical genetic testing for HCM patients sequences 8 to 17 genes to identify a causal variant. In about half of HCM patients who undergo clinical genetic testing, no known causal genetic variant is found. To better understand the genetic basis of HCM in these panel-negative patients, we performed targeted sequencing of 450 cardiac genes covering 3.3 Mbp in 58 unrelated HCM patients without known causal variants. We used Real Time Genomics (RTG) alignment and variant calling pipeline together with sequence to medical phenotypes (STMP) to align, call, and prioritize variants by comparison to publically available and cardiac-specific annotation data. We identified 126 extremely rare or novel exonic nonsynonymous and premature termination variants. Sanger sequencing validated novel stop variants in GJA5, CSRP3, and ATP1A4. To prioritize additional genes that may contribute to development of HCM, we performed gene-delimited rare variant association testing using the SKAT-O test. We compared our HCM patients against a cohort of 424 exome sequenced individuals without HCM. We identified new genes significantly associated with HCM case status including ILK (p = 9.9639 x 10^-18) and ADAMSTS9 (p = 1.8570 X 10^-5). To control for the potential confounding effect of ancestry, the results were replicated using a subset of the data set that contained only individuals of European ancestry. ILK has previously been implicated in dilated cardiomyopathy, suggesting that it is a strong biological candidate for affecting HCM. Through both functional annotation and burden testing we identify new candidate genes and variants that may play an important role in the genetic basis of HCM.

2137S

Whole exome sequencing of a family trio to identify potential genetic modifiers of *LMNA* arrhythmia and cardiomyopathy. *M. Zaragoza, S. Hakim, V. Hoang.* Pediatrics- Genetics & Genomics, University of California, Irvine, School of Medicine, Irvine, CA.

We identified a novel *LMNA* splice site mutation (c.357-2A>G) as the primary mutation associated with Sick Sinus Syndrome (SSS), Dilated Cardiomyopathy (DCM), and Sudden Cardiac Death (SCD) in a large family. To identify potential secondary factors, genetic "suppressors" or "enhancers" that may result in clinical variability within this family, we then focused on two affected members: a 71-year-old female and her 39-year-old daughter. The mother had bradycardia and atrial fibrillation at 49, sinus arrest at 52, and DCM at 61 while her daughter presented at 36 with bradycardia and ventricular arrhythmias. To identify potential modifiers associated with this phenotypic range, we conducted whole exome sequencing (WES) on the affected mother and daughter and the unaffected father. This produced 42 Gb of sequencing data in which we identified ~1 million unique variants for each parent and ~700,000 for the daughter. We then focused on the variants that were not shared by mother and daughter located in over 700 genes for ion channels, cardiac muscle contraction, *LMNA*-associated proteins (nuclear envelope proteins, transcription factors, chromatin-associated proteins), and *LMNA*-associated signal transduction (MAPK, mTOR, TGF Beta pathways). We validated the potential modifier variants by Sanger sequencing. We found seven potential "suppressors" including rare missense variants in the Lysosomal-associated membrane protein 2 (*LAMP2*), Mitogen-activated protein kinase 3 (*MAPK3*), and Retinoblastoma-like 2 (*RBL2*) genes. We found six potential genetic "enhancers" including a novel missense variant in the Unc-51-like kinase 3 (*ULK3*) gene and rare missense variants in the RNA binding motif protein 20 (*RBM20*) and Spectrin repeat containing nuclear envelope 1 (*SYNE1*) genes. In conclusion, we used WES to identify novel and rare variants in *LMNA* pathway-associated genes that serve as potential modifiers of SSS, DCM, and SCD. Gene expression and functional studies are planned. These studies may provide insight into the molecular mechanisms for the intrafamilial variability in *LMNA*-associated arrhythmia and cardiomyopathy. By taking advantage of the endogenous methods to alter disease; we hope that molecular targets may be discovered to develop new preventive strategies and treatments.

2138M

Timothy syndrome type 2 associated CACNA1C mutation G402S in a teenager with normal development presenting with ventricular fibrillation. *TP. Alastalo[1], A. Hiippala[1], J. Tallila[2], S. Myllykangas[3], JW. Koskenvuo[4].* 1) Padiatric Cardiology, University Hospital Helsinki, Helsinki, Finland; 2) Blueprint Genetics, Helsinki, Finland; 3) Institute of Biomedicine, University of Helsinki; 4) Research Center of Applied and Preventive Cardiovascular Medicine, University of Turku, Finland.

Timothy syndrome (TS) is a multiorgan disorder with prolonged QTc interval, congenital heart defects, syndactyly, typical facial features and neurodevelopmental problems. Ventricular tachyarrhythmia is the leading cause of death at early age. Classical TS type I results from a recurrent de novo CACNA1C mutation, G406R in exon 8A. An atypical form of TS, type 2, is caused by G406R and G402S mutations in the alternatively spliced exon 8. Only one individual for each exon 8 mutations has been described to date. In contrast to multiorgan disease caused by G406R either in exon 8A or 8, the G402S carrier manifested only an isolated cardiac phenotype with LQT and cardiac arrest. This isolated phenotype was suggested to result from somatic mosaicism. We describe a teenage patient resuscitated from ventricular fibrillation and treated with ICD. She has no other organ manifestations, no syndactyly, normal neurodevelopment and her QTc is <480ms. There is no family history of arrhythmias or sudden deaths. Targeted oligonucleotide-selective sequencing (OS-Seq) of channelopathy genes revealed a de novo substitution, G402S in exon 8 of CACNA1C. This is the third reported case of TS type 2. Direct sequencing of blood and oral mucosa derived DNA showed an identical mutation peak suggesting ubiquitous expression in different tissues. The phenotype of our patient and the previously described patient show an isolated arrhythmia disease with no other clinical manifestations of classical TS.

2139S

Rare potentially pathogenic variants in the congenital arrhythmia syndrome disease genes *SCN5A* and *KCNH2* are detected frequently but rarely associated with arrhythmia phenotypes in electronic health records. *S.L. Van Driest[1], S. Stallings[1], W.S. Bush[1,2], A. Gordon[3], D.R. Crosslin[3], G.P. Jarvik[3], D.S. Carrell[4], S.J. Bielinski[5], J.E. Olson[5], Z. Ye[5], I.J. Kullo[5], N.S. Abul-Husn[6], S.A. Scott[6], B.A. Castillo[7], J. Connolly[7], H. Hakonarson[7], L.J. Rasmussen-Torvik[8], S. Persell[8], M. Smith[8], T. Kitchner[9], J.R. Wallace[10], K.F. Doheny[11], R. Li[12], T.A. Manolio[12], T.E. Callis[13], D. Macaya[14], M.J. Ackerman[5], M.D. Ritchie[10], J.C. Denny[1], D.M. Roden[1].* 1) Vanderbilt University, Nashville, TN; 2) Case Western Reserve University, Cleveland, OH; 3) University of Washington, Seattle, WA; 4) Group Health Research Institute, Seattle, WA; 5) Mayo Clinic College of Medicine, Rochester, MN; 6) Icahn School of Medicine at Mount Sinai, New York, NY; 7) Children's Hospital of Philadelphia, Philadelphia, PA; 8) Northwestern University Feinberg School of Medicine, Chicago, IL; 9) Marshfield Clinic Research Foundation, Marshfield, WI; 10) The Pennsylvania State University, University Park, PA; 11) Johns Hopkins School of Medicine, Baltimore, MD; 12) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 13) Transgenomic, New Haven, CT; 14) GeneDx, Gaithersburg, MD.

The Electronic Medical Records and Genomics (eMERGE) consortium is using the next-generation PGRN-Seq platform to sequence 84 drug response genes in a diverse cohort of 9,000 individuals. Two of the 84 genes, *SCN5A* and *KCNH2*, encode ion channel subunits and harbor rare variants causing long QT syndrome and other arrhythmia phenotypes. Pathogenic variants in both are recommended for reporting as actionable incidental findings in recent American College of Medical Genetics and Genomics guidelines. In the initial 2,022 individuals sequenced, we identified 122 potentially pathogenic, rare (minor allele frequency < 1%), non-synonymous (nonsense, missense or splice site) variants. Five are currently designated as pathogenic for cardiac arrhythmia phenotypes in the ClinVar database. The 122 variants were submitted to 3 genomics laboratories with ion channel expertise (2 commercial and 1 academic) for assessment of pathogenicity. The number of variants assigned as pathogenic or likely pathogenic by each lab varied (lab 1: 16 variants; lab 2: 24; lab 3: 17). In all, 40 variants were designated as pathogenic or likely pathogenic by ≥ 1 lab, 9 by ≥ 2 labs, and only 4 by all 3 labs (rs199473166, rs137854618, rs199473603, and rs137854602, causing SCN5A I848F, D1275N, T1304M, and R1512W, respectively). To determine arrhythmia phenotypes for the 48/2,022 individuals with ≥ 1 of the 40 putatively-pathogenic variants, we reviewed electronic health record (EHR) data and ECG tracings (available in 33/48). None of the 48 had a family history of arrhythmia in the EHR. One individual with rs199473618 (SCN5A V1532I) had atrial fibrillation, a common arrhythmia. One individual with rs150264233 (SCN5A S1904L) had multiple instances of prolonged corrected QT interval (471-540 ms, normal < 440 ms), as well as normal corrected QT interval on many other ECGs. Four individuals had evidence of bundle branch block. No additional arrhythmia phenotypes were identified. We conclude that rare variants of undefined pathogenicity are often identified in these two arrhythmia syndrome genes. Expert curation of these variants resulted in inconsistent classifications. Among individuals with potentially pathogenic variants, arrhythmia phenotypes were rare. These data highlight the difficulty of rare variant interpretation and support a highly conservative approach to returning results of uncertain effect as "pathogenic," even in genes known to play a critical role in human disease.

2140M

Identification of histone modifier genes in 22q11DS patients with conotruncal heart defects by whole-exome sequencing. *T. Guo[1], J. Chung[1], D. McDonald-McGinn[2,6], W. Kates[3], T. Wang[4], R. Shprintzen[5], B. Emanuel[2,6], B. Morrow[1].* 1) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 2) Division of Human Genetics, Children's Hospital of Philadelphia and Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; 3) Department of Psychiatry and Behavioral Sciences, and Program in Neuroscience, SUNY Upstate Medical University, Syracuse, New York; 4) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York; 5) Velo-Cardio-Facial Syndrome International Center, Department of Otolaryngology and Communication Science, SUNY Upstate Medical University, Syracuse, New York; 6) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

The 22q11.2 deletion syndrome (22q11DS, velo-cardio-facial/DiGeorge) is the most common microdeletion disorder, occurring in 1:4,000 births. The majority of patients have a similar deletion on chromosome 22q11.2. Approximately 60% of patients have varied conotruncal heart defects (CTDs), while 40% have normal cardiac anatomy. We hypothesized that novel (not previously reported) or rare (<1%) functional coding variants, could serve as genetic modifiers contributing to the risk of CTDs in 22q11DS patients. Genomic DNA samples from 186 22q11DS subjects including 91 cases with CTDs and 95 controls without CTDs were selected for whole exome sequence (WES) capture using the NimbleGen SeqCap EZ Exome Library V3.0, followed by Illumina HiSeq 2000 sequencing yielding 100-bp paired-end reads (NHLBI; Nickerson, U. Washington). A total of 77,836 coding variants were identified from the WES data. Among these, 10,621 are novel mutations not observed in the general population, and are predicted in silico to be pathogenic. Among these novel mutations, 98.8% are private with 127 in more than one sample. Novel mutations in FES (Feline sarcoma virus oncogene) and ZNF429 occurred in 3 cases but 0 controls. Novel mutations in ZNF559-ZNF177 (readthrough) and PDE4A (Phosphodiesterase 4A) occurred in 0 cases but 3 controls. Gene-based analysis using collapsing methods found three genes (JMJD1C, SPTBN5, and ZNF559-ZNF177) significantly associated with CTDs (gene based p=0.01). Of particular interest, JMJD1C had separate missense mutations in 6 of 91 cases. JMJD1C is a histone demethylase, playing a role in chromatin modification. When we searched for other mutations in genes associated with chromatin modification, we found two missense mutations in KDM3B and two others in KDM7A. Both genes encode lysine histone demethylases, supporting the idea that missense mutations in histone demethylases may contribute to CTDs in 22q11DS. Interestingly, de novo rare functional variants in histone modifying genes were found in WES data in the NIH congenital heart disease, PCGC cohort (PMID: 23665959). Among the 26 genes reported from the PCGC, we found a mutation in RNF20 (E3 ubiquitin ligase; histone H2B) in one of our cases. Overall, the two studies implicate histone-modifying genes either as causative or as modifiers in the etiology of CTDs. In addition, it expands the number of such genes that may be important for heart development in humans.

2141S

Use of high-throughput genetic sequencing to screen for copy number variation in hypertrophic cardiomyopathy. *C. Dalageorgou[1], L.R. Lopes[1], C. Murphy[2], P. Syrris[1], W.J. McKenna[1], V. Plagnol[2], P.M. Elliott[1].* 1) Institute of Cardiovascular Sciences, University College London, London, United Kingdom; 2) UCL Genetic Institute, University College London, London, United Kingdom.

The role of copy-number variation (CNV) as a cause or modifier of the hypertrophic cardiomyopathy (HCM) phenotype is poorly studied. This lack of knowledge is important as most diagnostic genetic laboratories have used Sanger sequencing when screening patients with the disease and are unable to detect CNV. The aim of this study was to use high-throughput sequence (HTS) data combined with a read-depth strategy, to screen for CNV in cardiac sarcomeric protein genes in a large consecutive cohort of patients with HCM. Five-hundred-and-five unrelated HCM patients were genotyped using a HTS approach for 41 cardiovascular genes. A previously validated read-depth strategy (ExomeDepth) was used to call CNVs from the sequencing data. Detected CNVs in 17 prioritized HCM associated genes were then validated by comparative genomic hybridization array. Eleven CNVs were identified in total. Four CNVs identified in 4 patients (0.8% of the cohort) were confirmed by aCGH: one large deletion in *MYBPC3*; one large duplication in *MYBPC3*, one large deletion in *PDLIM3* and one duplication of the entire *TNNT2* gene. The findings have immediate implications for the genetic testing and counselling of HCM patients. A patient considered as genotype-negative can in fact harbour a structural variation not detected by direct sequencing. As HTS gradually substitutes Sanger sequencing in the clinical context, bioinformatic strategies should be used to call CNVs, in parallel to single nucleotide variants and small insertion-deletions.

2142M

Gene expression profile of bicuspid and tricuspid aortic valves with and without calcification by RNA sequencing. *S. Guauque-Olarte[1], A. Droit[2], F. Hadji[1], N. Gaudreault[1], J.G. Seidman[3], S.C. Body[3], P. Pibarot[1], P. Mathieu[1], Y. Bossé[1,4].* 1) Centre de recherche Institut universitaire de cardiologie et de pneumologie de Québec, Laval University, Quebec, Canada; 2) Centre de Recherche du CHU de Québec, Laval University, Quebec, Canada; 3) Harvard Medical School, Boston, Massachusetts, USA;; 4) Department of Molecular Medicine, Laval University, Quebec, Canada.

Aortic valve stenosis (AS) is a cardiovascular disease that can be fatal in the absence of treatment. Individuals with a bicuspid aortic valve (BAV), the most common cardiovascular congenital abnormality, develop symptoms of AS 10 to 15 years earlier compared to those with a tricuspid valve. The molecular mechanisms leading to the premature development of AS in BAV patients are unknown. The objective of this study was to identify genes differentially expressed between calcified BAV and tricuspid valves with (TAVc) and without (TAVn) calcification using RNA sequencing. Ten human calcified BAV and nine TAVc were collected from male patients who underwent aortic valve replacement at the IUCPQ. All valves had the same degree of fibro-calcific remodeling. Ten TAVn were obtained from male patients who underwent heart transplantation. mRNA levels were measured using the Illumina HiSeq 2000 system. The alignment of paired-end reads and the comparison of gene expression among three groups of valves were performed using the TopHat-Cufflinks protocol. A q-value <0.05 was used as a significant threshold. Genes with FPKM <10 in two tissues compared were excluded. IPA was used to identify molecular functions enriched for genes differentially expressed. Thirty-one genes were up-regulated and 33 were down-regulated in BAV compared to TAVc, including 37 genes with fold change >2. The atherosclerosis (q=0.01), calcium signaling (q=1.95E-06), and phospholipases (q=7.76E-03) pathways were significantly enriched with differentially expressed genes. A total of 283 genes were up-regulated and 340 were down-regulated in BAV compared to TAVn, counting 288 genes with fold change >2. Compared to TAVn, 189 and 233 genes were up- and down-regulated in TAVc, respectively, including 248 genes with fold change >2. Tens of genes were newly identified to be differentially expressed among these groups of aortic valves. This is the first study analyzing the transcriptome of aortic valves by using RNA sequencing. AS causes significant changes in the expression profile of BAV and TAVc compared to normal valves. Pathways biologically relevant for AS were significantly enriched with differentially expressed genes. The role of these genes and pathways in the pathogenesis of AS in BAV patients requires further investigation. These results contribute to the search of new therapeutic targets with the potential to avoid or slow the development and/or progression of AS in patients with BAV and TAV.

2143S

Family-control analysis based on Hamming distance successfully prioritizes exome sequence variants in familial cardiomyopathy. *A. IMAI[1], J. Ott[2,3], J. Majewski[4], Y. Sakata[1], Y. Asano[1], S. Takashima[1,5].* 1) Cardiovascular Department, Osaka University Graduate School of Medicine, Suita, Japan; 2) Laboratory of Statistical Genetics, Rockefeller University, NY, US; 3) Institute of Psychology, Chinese Academy of Sciences, Beijing, China; 4) McGill University and Genome Québec Innovation Centre, Quebec, Canada; 5) Department of Medical Biochemistry, Osaka University Graduate School of Medicine, Suita, Japan.

Detecting causative genes for cardiomyopathy among hundreds of candidate variants is still challenging, especially in small pedigrees where conventional linkage analysis is underpowered or even uninformative. To overcome this problem, we developed a novel approach combining family data with population controls. The idea is that, due to linkage disequilibrium, a set of variants surrounding a disease locus should have different frequencies in affecteds and controls. Thus, we measure the distance between sets of variants in affected and unaffected individuals, where “distance” refers to the Hamming distance. At any variant surrounding a candidate variant, we distinguish between presence and absence of that variant. For a set of N such variants, the Hamming distance is the number n of variants differing between two individuals in terms of presence/absence. We then work with the Hamming distance ratio, HDR = n/N. We form pairs of all individuals and distinguish those pairs containing an affected family member (group 1) from pairs consisting only of controls (group 2). At a given candidate variant and surrounding variants, we calculate HDR for each pair of individuals and assess the difference in mean HDR values between groups 1 and 2 by a t statistic. In two hypertrophic cardiomyopathy families, known pathogenic mutations were previously detected: c.173G>A (p.R58Q) in the MYL2 gene (rs104894369), and c.746G>A (p.R249Q) in the MYH7 gene (rs3218713). Using current state-of-the-art approaches, we applied the usual filtering methods to select ~600 candidate variants from all variants obtained in exome sequencing. For one affected individual in a given small autosomal dominant family and 41 control individuals, we applied our approach. Not knowing how many variants should be included in our HDR calculation, we defined 8 regions (30-100KB) around a given candidate variant. For each region, a t statistic was computed whose maximum over the regions represents the test statistic, T, for a given candidate variant. We ranked candidate variants by their T statistic. For the two cardiomyopathy families, we were able to narrow down known disease variants to be up to the top 3% of all candidate variants. By contrast, HDR ranked 535th out of 631 variants for a TNNI3 *de novo* mutation in a restrictive cardiomyopathy pedigree. Our new statistical method for prioritizing disease regions will be useful for affected individuals in small autosomal dominant pedigrees.

2144M

Rare and common variation in *SCN10A* is associated with Brugada Syndrome. *E. Petropoulou[1], B.P. Prins[1], E. Savio-Galimberti[8], T. Yang[8], L. Crotti[4], J. Jabbari[3], M.J. Ackerman[7], R. Redon[6], P. Guicheney[5], M.S. Olesen[3], D. Darbar[8], C.R. Bezzina[2], E.R. Behr[1], Y. Jamshidi[1], Brugada Gene Discovery Group.* 1) Human Genetics Research Centre, St George's University of London, London, United Kingdom; 2) Academic Medical Center (AMC), Amsterdam, The Netherlands; 3) Laboratory for Molecular Cardiology, The Heart Centre, Copenhagen University Hospital , Copenhagen, Denmark; 4) Department of Molecular Medicine, Section of Cardiology, University of Pavia, Pavia, Italy; 5) ICAN Institute for Cardiometabolism and Nutrition, Université Paris 06, Pierre et Marie Curie, Paris, France; 6) Centre Hospitalier Universitaire (CHU) Nantes, France; 7) Mayo Clinic, Boston, MA, USA; 8) Vanderbilt University, Nashville, USA.

Introduction: Brugada syndrome (BrS) is a rare cardiac arrhythmia associated with mutations in *SCN5A*, which encodes the alpha subunit of the canonical cardiac sodium channel in ~20% of cases. BrS is associated with QRS prolongation, implying that slowed cardiac conduction is an important element of the phenotype. We hypothesized that genes in loci associated with QRS duration in the general population may harbor rare and/or common variation that contribute to the predisposition of BrS. **Methods:** A multicentre candidate gene study of 7 genes (*SCN10A*, *HAND1*, *PLN*, *CASQ2*, *TKT*, *TBχ3* and *TBχ5*) was carried out for 156 Caucasian *SCN5A* mutation-negative patients with BrS. The coding regions of the genes were sequenced, and variants categorised as pathogenic or benign when ≥3 bioinformatics tools (Polyphen2, SIFT, Grantham and GERP) predicted pathogenicity. Variant allele frequencies were compared with publicly available data (4,300 European Americans) from the NHLBI GO Exome Sequencing Project (ESP). Co-segregation analysis of putative pathogenic mutations was performed where DNA was available. Voltage-clamp experiments in ND7/23 cells were performed for the V1073A variant in *SCN10A*, which encodes the Nav1.8 sodium channel. **Results:** 49 variants were identified in the 7 genes; 23 were common single nucleotide polymorphisms (SNPs; minor allele frequency >1%). 9 variants were novel with 7 predicted as potentially pathogenic. Although co-segregation analysis was performed for 4 of the novel putative pathogenic variants, only G1299A in *SCN10A* showed co-segregation with disease alongside another novel *SCN10A* variant, I671V. There was a strong association of the common *SCN10A* variant V1073A (T>C) with BrS [frequency of T allele in BrS: 65.1% vs 40.1%, $p = 3.54\times10^{-19}$, odds ratio (95%) = 3.05 (2.37-3.91)] and voltage-clamp experiments showed a “gain-of-function” phenotype for this variant. **Conclusion:** Our study suggests that rare coding variants in the genes screened are not a major cause of BrS. The *SCN10A* common variant V1073A associated with BrS, whilst unlikely to explain the occurrence of BrS alone, confirms previous studies identifying an important role for common variation in *SCN10A* for BrS. Furthermore, we observed low specificity for predictions of likely pathogenicity when using a combination of in-silico prediction tools alone. This demonstrates that the performed in-vitro electrophysiology studies represented a valuable addition to predict pathogenicity.

2145S

Whole genome sequencing association study of ECG conduction traits. *B.P. Prins[1], F. Petropoulou[1], E. Zeggini[2], N. Soranzo[2], T.D. Spector[3], Y. Jamshidi[1], UK10K Consortium Cohorts Group.* 1) Cardiogenetics Lab, Cardiovascular and Cell Sciences Institute, St George’s University of London, London, United Kingdom; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom; 3) Department of Twin Research & Genetic Epidemiology, King's College London, London, United Kingdom.

The UK10K project, funded by the Wellcome Trust, has completed whole genome sequencing (WGS) to ~6.7x coverage of 1,754 individuals from the TwinsUK study. Variants discovered through WGS, along with those from 1000 Genomes were imputed into the full cohort with genome-wide genotype data, increasing the sample size to 4,329, for which for 2,327 samples ECG measurements were available. We performed discovery analyses testing 9.2 million variants (MAF>1%) in a meta-analysis of sequenced and imputed variants for association with ECG conduction traits (PR interval, QT interval, QRS duration, and heart rate) and with the extremes of these traits. We identified 2 variants ($p < 5 \times 10^{-8}$) in known ECG trait-loci; a novel variant for PR interval and *SCN10A* (rs6599255, intronic, MAF 41.6%), and a known variant for QT interval and *NOS1AP* (rs12143842, 6 kb upstream, MAF 24.6%). We also identified 133 variants across 131 known and novel loci suggestively associated ($p < 1\times10^{-5}$) across the four ECG traits, some of which appear to be pleiotropic in nature. 60 variants were uncommon (<5% MAF), and 73 common (≥ 5% MAF). We are replicating these in independent cohorts with 1000 Genomes and UK10K imputed data, followed by direct genotyping of non-imputable variants. We performed molecular and cellular function enrichment analyses on a set of genes harbouring the identified variants, for which intergenic variants were assigned to their nearest gene and combined with genes already harbouring identified variants. We found significant enrichment ($p < 5\times10^{-2}$) for genes involved in sarcolemma function, calmodulin binding, voltage-gated channel activity and cytoskeletal protein binding. Using publicly available data from the ENCODE project on chromatin states, epigenetic modifications and sequence motifs, we show that some of these variants are likely to have functional consequences affecting regulatory elements. Currently rare variant-, and combined common-rare variant analyses are being performed using SKAT, where we collapsed variants in 57,820 gene sets based on GENCODE v19 definitions to test whether rare variation may be associated with ECG traits, and to dissect the contribution of common and rare variants to each ECG trait. These results will provide insights into how both rare and common variation are likely to contribute to the understanding of the genetic architecture of ECG traits, and are likely to explain a substantially larger fraction of the missing heritability.

2146M

Replication of a single nucleotide polymorphism variant in CETP gene associated with HDL level in the ClinSeq® Study. *H. Sung[1], T. Schwantes-An[1], B. Suktitipat[1], K. Lewis[2], D. Ng[2], S.G. Gonsalves[2], J.C. Mullikin[3,4], L.G. Biesecker[2], A.F. Wilson[1], National Institutes of Health Intramural Sequencing Center (NISC).* 1) Genometrics Section, Computational and Statistical Genomics Branch, NHGRI, NIH, Baltimore, MD; 2) Medical Genomics and Metabolic Genetics Branch, NHGRI, NIH, Bethesda, MD; 3) Comparative Genomics Analysis Unit, Cancer Genetics and Comparative Genomics Branch, NHGRI, NIH, Bethesda, MD; 4) National Institutes of Health Intramural Sequencing Center (NISC), NHGRI, NIH, Bethesda, MD.

ClinSeq® is a large-scale medical sequencing study designed to investigate associations of rare sequence variants with traits related to Coronary Artery Disease (CAD). The study currently includes more than 1000 non-smoking participants between the ages of 45 to 65 with normal to severe coronary artery calcification scores. About 200 CAD-related traits were measured at the NIH Clinical Research Center in Bethesda, MD. Whole-exome sequencing was performed with the Agilent SureSelect 38Mb and 50Mb capture kits for 387 and 325 individuals, respectively, at the NIH Intramural Sequencing Center. Single nucleotide variants (SNVs) common to both capture regions with a call rate > 98% and minor allele frequency (MAF) > 1% were used to check for cryptic relatedness and for misspecified population stratification by multidimensional scaling analysis - 635 unrelated European Americans (EAs) remained. For each capture kit with EAs only, SNVs with at least one homozygote and a call rate greater or equal to 50% were included. The two capture regions with only SNVs in common were merged, yielding 439,807 SNVs. Of these SNVs, 74% and 46% had MAFs < 0.01 and < 0.001, respectively. The SNVs with MAF < 0.01 were collapsed into a single derived variant for each genomic region defined by hotspot blocks. Collapsed variants were coded as the proportion of minor allele occurring within each region; common variants were coded as the number of minor alleles (scaled from 0 to 1). Tests of association of each SNV with High Density Lipoprotein (HDL) level was performed on untransformed, log-transformed and rank-inverse transformed HDL with simple linear regression, adjusting for age, sex, BMI and medication. The SNV rs1532625 in the intron of CETP gene was found to have a significant association for all untransformed and transformed HDL traits at the level of 3.76e-07 after Bonferroni correction, replicating the study by Reilly et al. For all untransformed and transformed HDL traits, an additional three SNVs were found significant in the intron of the CETP (rs7205804), RBBP6 (rs3815906) and MCM5 (rs713618) at the level of 1e-05 and one SNV in the exon of PNMAL11 (rs111356009) at the level of 1e-04. CETP and RBBP6 genes are known to be associated with HDL and PNMAL1 gene with triglycerides.

2147S

Assessment of the implication of rare coding variants in familial and sporadic Fibromuscular Dysplasia. *N. Bouatia-Naji[1,2], SR. Kiando[1,2], PF. Plouin[2,3], MC. Barlassina[4], D. Cusi[5], P. Galan[6], M. Lathrop[7], X. Jeunemaitre[1,2,8].* 1) INSERM UMR970 PARCC, Paris, FRANCE; 2) Université Paris-Descartes, PRES Sorbonne Paris Cité, Paris FRANCE; 3) AP-HP, Department of Hypertension, Hôpital Européen Georges Pompidou, Paris, FRANCE; 4) Filarete Foundation, Genomic and Bioinformatics Unit ; Dept. of Medicine, Surgery and Dentistry, University of Milano, ITALY; 5) Graduate School of Nephrology, University of Milano, Division of Nephrology, San Paolo Hospital, Milano, ITALY; 6) Unité mixte de recherche 557 Inserm/Inra/Cnam/Paris13, FRANCE; 7) McGill University and Génome Québec Innovation Centre, Montréal, Québec, CANADA; 8) AP-HP, Refferal Center for Rare Vascular Diseases, Hôpital Européen Georges Pompidou, Paris, FRANCE.

Fibromuscular dysplasia (FMD) is an arterial disease characterized by non-atherosclerotic stenosis reported mostly in renal and extra-cranial carotid arteries. FMD predisposes to hypertension, renal ischemia and stroke, the first cause of disability worldwide. The genetics of fibromuscular dysplasia is under-investigated because of the lack of large families and cohorts due to the rarity of the disease and the complexity of the imaging-based diagnosis. The causes of FMD are unknown and it occurs predominantly in females with a prevalence of ~4/1000 for the clinical forms. There are strong arguments in favour of the genetic origin of FMD, based on documented and reported family history, although the precise estimation of its heritability is missing. FMD is probably a typical complex genetic disease and it is challenging to investigate because it is also a rare disease. Exome sequencing in 16 familial cases of FMD (five sibs and two sib-trios) generated 3,971 genes with at least one rare (MAF<0.01) and predicted functional variant. Inter-families analyses showed that no gene could explain the shared FMD status in at least three out of seven families. We also analysed 15 known causative genes of vascular diseases and/or connective tissue syndromes (e.g. FBN1, TGFB2 and COL3A1). None of these genes presented rare and putatively functional variants that could explain the FMD status in any of the seven families (intra-family analyses). Then, we aimed to follow-up 22 genes that were identified as harboring at least four rare variants and the 15 candidate genes in a larger sample of 259 FMD unrelated cases and 698 controls using genotyping data generated by the exome chip, an array enriched for rare and predicted functional variants. Neither gene-based association analyses of rare variants (MAF<0.01) using SKAT-O nor single SNP association of common SNPs by logistic regression (MAF>=0.05) support a role of the 37 genes tested (Ntotal=982 SNPs, including 264 common SNPs). These findings support strong genetic heterogeneity for FMD and encourage more powerful and comprehensive genomic approaches, such as genome-wide association studies, to decipher the genetic basis of FMD.

2148M

High Throughput Sequencing and Bioinformatic Analysis in Familial Congenital Heart Disease. *D. Corsmeier[1], S. Fitzgerald-Butt[2,5], G. Zender[2], M. Mori[3], L. Kelly[3], K. Walters[4], V. Vieland[4,5], H. El Hodiri[3,5], V. Garg[2,5], K. McBride[2,5], P. White[1,5].* 1) The Biomedical Genomics Core & The Center for Microbial Pathogenesis, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) The Center for Cardiovascular and Pulmonary Research, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 3) The Center for Molecular & Human Genetics, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 4) Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 5) Department of Pediatrics, College of Medicine, The Ohio State University, Columbus, OH.

Congenital heart disease (CHD) is the most common type of birth defect, appearing in eight of one thousand live births. CHDs are a leading cause of death in infants and represent an increasing cause of adult morbidity and mortality. There is strong evidence for a large genetic component in the etiology of non-syndromic CHDs. As such, whole exome and whole genome sequencing will be invaluable in elucidating the underlying causal variants which lead to disease. In this study, high throughput sequencing approaches were utilized to identify genetic variants causing non-syndromic CHDs in families exhibiting Mendelian segregation of CHDs. Our subject cohort consists of over 800 affected cases and includes over 100 multiplex families. To date whole exome sequencing has been performed on several hundred of these samples using Agilent SureSelect Human All Exon V5 kits and the Illumina HiSeq platform. In selected cases for which preliminary results yielded no disease candidate mutations, more extensive sequencing methods were utilized including paired end and mate-paired whole genome sequencing. Bioinformatic analysis was performed using Churchill, the pipeline we developed for the discovery of human genetic variation. Churchill utilizes the Burrows Wheeler Aligner (BWA) and the Genome Analysis ToolKit (GATK) to perform sequence alignment and variant calling. The initial filtering and prioritization of variants, we used traditional heuristic methods based on existing variant databases (dbSNP, 1000 Genomes, NHLBI Exome Sequencing Project, HGMD). For bioinformatic and statistical validation, we explored other widely-used methods. These different methods yield highly consistent result sets from which we have been able to identify putative causal variants in these families and mark them for further validation and functional studies. These findings will provide a better understanding in the pathogenesis of CHD and can lead to improved diagnosis and novel preventive and therapeutic approaches.

2149S

Exome Sequencing Reveals Candidate Genes for Spontaneous Coronary Artery Dissection. *J.L. Theis, M.S. Tweet, J.M. Evans, R. Gulati, S.N. Hayes, T.M. Olson.* Mayo Clinic, Rochester, MN.

Spontaneous coronary artery dissection (SCAD) is an uncommon, idiopathic disorder that primarily affects young, otherwise healthy women. The pathophysiologic process involves an intimal dissection or intramural hematoma, which can obstruct blood flow and cause myocardial infarction and sudden death. A majority of patients has concomitant fibromuscular dysplasia (FMD), implicating a generalized arteriopathy. Postulating a genetic basis for sporadic SCAD, whole exome sequencing and comparative variant filtering of proband-parent trios was performed. The cohort for this pilot study was comprised of 28 unrelated patients with angiographically confirmed SCAD in the absence of a systemic connective tissue disorder (93% female; mean age 41 years). Exomes of each trio were uploaded into Ingenuity Variant Analysis™ software and filtered for rare (<1% minor allele frequency), predicted-deleterious coding variants. Modeling for de novo and recessive modes of inheritance yielded 150 unique candidate genes, none of which comprise commercial gene testing panels for thoracic aortic aneurysm and dissection. A short list of 8 prioritized candidate genes was generated based on presence in >2 trios (n=1), known coronary artery expression (n=2), or abnormal vascular phenotype associated with disruption of the murine ortholog (n=5). Each candidate gene harbored missense variants arising de novo (n=1) or inherited as recessive alleles (5, compound heterozygous; 2, homozygous). Two of these genes, each linked to SCAD associated with FMD, have been selected for mutation scanning of a cohort of 270 additional unrelated SCAD subjects by high throughput heteroduplex analysis. While our preliminary investigation has ruled out a common disease gene for SCAD, whole exome sequencing has identified plausible candidate genes, a first step toward deciphering molecular underpinnings of a disorder that may account for 40% of myocardial infarctions in women <50 years.

2150M

Feasibility of a Targeted Ion Torrent Next Generation Sequencing platform for Identification of Mutations Associated with Inherited Cardiac Arrhythmia Syndromes. *E. Burashnikov, R. Pfeiffer, G. Caceres, Y. Wu, C. Antzelevitch.* Molecular Genetics, Masonic Medical Research Laboratory, Utica, NY.

Introduction: Many unexplained sudden deaths are due to inherited cardiac arrhythmia syndromes such as Brugada Syndrome (BrS), Early Repolarization Syndrome (ERS), Short QT Syndrome (SQT) and idiopathic ventricular fibrillation (IVF). We sought to determine the suitability of screening a panel of 30 genes encoding sodium, calcium and potassium channels genes, including ATP-sensitive potassium channels, glycerol-3-phosphate dehydrogenase 1-like (GPD1L), genes previously associated with arrhythmia syndromes. We also included ankyrin 2 (ANK2), ryanodine receptor 2 (RYR2), ATP-binding cassette transporters (ABCC8, ABCC9), caveolin 3 (CAV3) and hes-related family gene (HEY2), as a potential causative contributors to these diseases. Methods: We employed an Ion Torrent Personal Genome Machine (Life Technologies) for Next Generation Sequencing and analyzed data with Ion Reporter Software (Life Technologies). Variants were annotated by Ingenuity Variant Analysis (Qiagen) and were confirmed by Sanger method on 3730 Automatic DNA Analyzer. Any variation with a frequency less than or equal to 0.5% was considered as a mutation, based on 1000 Genome Frequency. Libraries were constructed using Custom Ion Ampliseq to amplify all exons and intron borders of the targeted genes. Probands included in the study were predominantly SCN5A-negative. Results: Total of 121 probands were screened, among them 50 BrS, 42 ERS, 17 SQT and 12 cases IVF/VF. We identified mutations in 33 of 50 BrS probands screened, for a yield of 66%. Mutations in ANK2, calcium channel genes, SCN10A, SCN5A, and $I_{K\text{-}ATP}$ genes were found in 24, 20, 16, 8 and 18% of probands, respectively. In the ERS group, we identified mutations in 21 of 42 probands screened, for a yield of 50%. Mutations in ANK2, calcium channel genes, SCN10A, SCN5A, and $I_{K\text{-}ATP}$ genes were found in 17, 13, 14, 7 and 9% of probands, respectively. The SQT group had mutations in 13 of 17 probands screened, for a yield of 76%. Mutations in ANK2, CACNA1C, SCN10A, RYR2 and $I_{K\text{-}ATP}$ genes were found in 18, 24, 18, 24 and 12% of probands, respectively. In the IVF group, we uncovered mutations in 6 of 12 patients screened, for a yield of 50%. Mutations in ANK2, CACNA1C, SCN10A, SCN1B, RYR2, GPD1L and $I_{K\text{-}ATP}$ genes were 8% for each. Conclusions: Our study shows the feasibility of using a targeted Next Generation Sequencing approach for high throughput screening of SCD mutations. The study shows a yield of 50 to 76% for BrS, ERS, SQT and IVF probands.

2151S

Whole-Exome Sequencing in Multiplex Family Identifies Novel Risk Variants for Thrombotic Storm. *J.M. Vance[1], N.D. Dueker[1], A. Beecham[1], G. Beecham[1], P. Whitehead[1], S. Hahn[1], N. Hoffman[1], J.W. Lawson[2], D. Erkan[3], L.R. Brandao[4], A.H. James[5], M.J. Manco-Johnson[6], R. Kulkarni[7], C.S. Kitchens[2], T.L. Ortel[8], M.A. Pericak-Vance[1].* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) University of Florida, Gainesville, FL; 3) Hospital for Special Surgery, Weill Medical Center of Cornell University, Ithaca, NY 14850; 4) Paediatric Haemotology/Oncology, The Hospital for Sick Children, Toronto, ON; 5) Department. of OB/GYN, Duke University Medical Center, Durham, NC; 6) University of Colorado Denver and The Children's Hospital, Aurora, CO; 7) Pediatrics & Human Development, Michigan State University, East Lansing, MI; 8) Hematology, Duke University Medical Center, Durham, NC.

Thrombotic storm (TS) is a devastating and extremely rare disorder that occurs in a small subset of patients with venous thromboembolic disease. TS is characterized by patients exhibiting ≥ 2 of the following in a short period: 1) more than two acute arterial/venous thromboemboli, or thrombotic microangiopathy, 2) unusual location, 3) progressive/recent unexplained recurrence, or 4) refractory/atypical response to therapy. The causes of TS are currently unknown, but we hypothesized that rare genetic variants explain a proportion of the risk for TS. To identify candidate genetic variants we performed whole-exome sequencing (WES) on the only known TS multiplex family existing in the world. Within our multiplex TS family, 10 family members were WES including four sisters, 2 of which were TS affected and the other 2 unaffected. DNA was captured using the Agilent 50mb kit and sequenced using Illumina HiSeq2000. Alignment and genotype calling were performed with Burrows-Wheeler Aligner and Genome Analysis ToolKit. Using this WES data we filtered variants to identify those inherited under 3 different inheritance models; dominant, recessive, and co-dominant. For all analyses we included only those variants that met the following criteria; followed the specified inheritance model, found in the 2 affected sisters but not found in the 2 unaffected sisters, VQSLOD>-4, conserved, rare (minor allele frequency (MAF)<5%) and either missense, nonsense or splice site variants. Under a dominant model, we identified a novel variant located in zinc finger protein, FOG family member 1 (ZFPM1). ZFPM1 is an excellent TS candidate as the ZFPM1 protein interacts with GATA binding protein 1 (GATA1), a protein that plays a role in erythroid development. Previously identified mutations in GATA1 have been associated with X-linked dyserythropoietic anemia and thrombocytopenia. The variant we identified in our analysis was a missense variant predicted to be possibly damaging in PolyPhen-2. Under a recessive model, we identified two additional candidate variants; a stop-gained variant in keratin 37 (KRT37) and a missense variant in zinc finger protein 568 (ZNF568). In conclusion, through WES, we identified several loci that may play a role in the development of thrombotic storm, a complex disorder that has only recently been clinically defined and for whose cause is unknown.

2152M

Identification of rare exonic variants predisposing to early onset coronary heart disease in families from Northern Finland. *E. Nevala[1,2,10], S. Sebert[1,2], S. Kakko[2,3,4], M.J. Savolainen[2,3,4], M.R. Järvelin[1,2,5,6,7], A. Palotie[8,9], M. Männikkö[1,2,10], T. Kangas-Kontio[2,3,4].* 1) Institute of Health Sciences, Centre for Lifecourse Epidemiology, Faculty of Medicine, University of Oulu, Finland; 2) Biocenter Oulu, University of Oulu, Finland; 3) Institute of Clinical Medicine, Department of Internal Medicine, University of Oulu, Finland; 4) Clinical Research Center, Oulu University Hospital, Finland; 5) Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, United Kingdom; 6) Unit of Primary Care, Oulu University Hospital, Finland; 7) Department of Children and Young People and Families, National Institute for Health and Welfare, University of Oulu, Finland; 8) Broad Institute of MIT and Harvard, Boston, Massachusetts, USA; 9) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 10) Faculty of Biochemistry and Molecular Medicine, Univeristy of Oulu, Finland.

Coronary heart disease (CHD) is the cause of every fourth death in the working-age population in Finland and the leading cause of death in the whole population. In our study we have altogether 60 extended families with over 500 subjects from an isolated and homogenous population of Northern Finland. The probands were diagnosed with early onset CHD (<55 years). To identify rare variants affecting the disease risk 51 most interesting subjects from 20 different families were selected for whole-exome sequencing (WES) on the basis of the severity of the phenotype or family history. We studied which SNPs with low occurrence (minor allele frequency < 0.05 in 1000 Genomes European population) and harmful functional prediction according to Sift, PolyPhen and Condel algorithms are co-inherited with the disease in each of the families separately. We identified per family on average 1) one variant located in CHD susceptibility genes found from OMIM database, 2) five variants in genes located nearby to the variants that have been associated with CHD in GWAS studies, and 3) five variants located in the linkage regions identified in our previously published linkage analysis. Furthermore, certain variants were shared between the families, and several genes had alternative variants in multiple families. There is intense discussion on disentangling the role of rare and common genetic variants in the etiology of complex diseases. Many studies attempting to identify the rare variants associated with the disease are large population based association studies whereas our study is designed to recognize inherited risk alleles. The discovered novel variants are located on promising genes and a more detailed analysis might provide new insights to molecular pathways responsible for the pathogenesis of CHD. Our family-data-based findings are now being confirmed by Sanger sequencing in the remaining study population and also in a large Northern Finland Birth Cohort.

2153S

Retrotransposons in Nonsyndromic Conotruncal Heart Defects. *D. Webber[1,2], S. MacLeod[1,2], S. Erickson[1,2], M. Li[1,2], L. Murphy[1,2], C. Hobbs[1,2].* 1) Pediatrics Department, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR; 2) Arkansas Children's Hospital Research Institute, Little Rock, AR.

Introduction: Congenital heart defects (CHDs) are a leading cause of infant morbidity and mortality. Research on CHDs has focused primarily on genomic variants within coding regions and has yet to investigate DNA transposable elements, which occupy nearly half of the human genome and are implicated in a wide range of structural and developmental conditions. We identified and characterized retrotransposon insertion polymorphisms (RIPs) among families affected by nonsyndromic conotruncal heart defects (CTDs). **Methods:** Participants were enrolled in the National Birth Defects Prevention Study at the Arkansas site between 1997 and 2007. We used probe hybridization enrichment and sequencing to map the position of human specific retrotransposons, including L1Hs and AluYb8/9, throughout the genomes of 88 case families ($n = 263$) affected by conotruncal heart defects and 40 unaffected control families ($n = 120$). Paired-end sequencing data was aligned to the reference genome with the Burrows-Wheeler Aligner (BWA), and non-reference RIPs were identified using RetroSeq. RIPs reported in dbRIP or 1000 Genomes were filtered out of the dataset in order to enrich for novel retrotransposon insertions, likely to be under selective pressure or to be a product of recent retrotransposition. **Results:** A similar proportion of novel RIPs were identified in case families compared to controls. Among cases, there were 26 novel RIPs in mothers, 37 in fathers, and 30 in infants. Nearly half (14/30) of the RIPs affecting case infants were located within introns, and the majority (21/30) were members of the LINE-1 family. Eight RIPs were observed in more than one case family, including an Alu retrotransposon within the *GLUD1* intron that was present in 5 case families and in no controls. Modeling suggest that this novel insertion introduces two canonical splice acceptor sites that are poised to disrupt the adjacent exon, which encodes a GTP binding domain previously implicated in familial hyperinsulinemic hypoglycemia 6 (HHF6; OMIM #606762). These findings are strengthened by gene ontology analysis among case infants, which showed significant enrichment ($P = .003$) of intronic RIPs in genes (*GLUD1, SLC30A8, PARK2*) involved in "regulation of insulin secretion." **Conclusion:** Proper glucose homeostasis is critical to cardiogenesis, and maternal diabetes and obesity are well-established risk factors for CHD. Results suggest that rare RIPs may influence glucose metabolism and risk of CTDs.

2154M

Cardiovascular Disease in Pediatric Patients with Ciliopathies. *S. Bowdin[1,2], N. Karp[1].* 1) Hospital for Sick Children, Clinical Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Pediatrics, University of Toronto, Canada.

Ciliopathies are a heterogeneous group of disorders caused by mutations in single genes, which alter proteins of the cilium-centrosome complex (CCC). As the primary cilia can play multiple roles and are ubiquitously distributed throughout the body, abnormal ciliary proteins can lead to multiple organ malformations and dysfunctions, including congenital heart disease. Published information on cardiovascular disease in patients with ciliopathies remains scarce , although the presence of significant cardiac disease is of cardinal importance to the clinical outcome. We conducted a retrospective review of the hospital charts and echocardiograms/electrocardiograms of 72 pediatric patients diagnosed with a ciliopathy syndrome, all of whom were cared for at the Hospital for Sick Children, in Toronto, Canada between 1983-2013. Each diagnosis of a ciliopathy was made on the basis of clinical phenotype, when possible confirmed by molecular genetic testing. A total of 23 out of 72 ciliopathy patients (32%) were found to have an abnormal echocardiogram. Five out of six patients with Alstrom syndrome had dilated cardiomyopathy as well as other structural cardiac anomalies and ECG findings . Six out of 39 patients with Bardet-Biedl syndrome, and 7 out of 19 patients with Joubert syndrome and related disorders, had structural cardiac anomalies and/ or rhythm disturbances. Three out of 6 patients with nephronophthisis had hypertension but none had structural cardiac defects. Seven patients with Leber Congenital Amaurosis, 1 patient with Senior-Loken syndrome, and 1 patient with Oral-Facial-Digital Syndrome type 1 had no documented cardiovascular disease. This data suggests that consideration should be given to a cardiovascular evaluation for all patients with a known or suspected ciliopathy. In the future, collection of prospective data may enable screening recommendations to be tailored to the patient's specific ciliopathy diagnosis .

2155S

Genotype and Phenotype Characteristics of Filipino Families with Familial Hypercholesterolemia and Novel LDL-R Gene Mutation. *E.C. Cutiongco de la Paz[1], R.G. Sy[2], K.N. Hernandez[1], E.B. Llanes[2], F.E. Punzalan[2], C.P. Cordero[3].* 1) Institute of Human Genetics, National Institutes of Health, University of the Philippines Manila, Philippines; 2) Section of Cardiology, Department of Internal Medicine, College of Medicine-Philippine General Hospital, University of the Philippines Manila; 3) Institute of Clinical Epidemiology, College of Medicine, University of the Philippines Manila.

Familial hypercholesterolemia (FH) is a disorder of lipoprotein metabolism and is one of most common inherited disorders worldwide A diagnosis of FH is typically based on high plasma levels of total and LDL cholesterol, a family history of childhood-onset hypercholesterolemia, cholesterol deposition in extravascular tissues (e.g. tendon xanthomas, corneal arcus), and a personal and family history of premature CVD. However, there are no absolute predictive clinical criteria for FH diagnosis. Genetic testing is recommended in combination with clinical assessment to provide a more definitive diagnostic process. The LDL-R gene codes for the LDL receptor protein which is responsible for the uptake (and subsequent degradation) of serum LDL cholesterol via receptor-mediated endocytosis. Mutations in the LDL-R gene, can impair receptor synthesis or lead to the production of defective receptors. To date, 1,122 unique genetic variations have already been reported for the LDL-R gene and may be found in the online registry maintained by the Leiden Open (source) Variation Database (LOVD). Included here are the five novel mutations found by study involving selected Filipino FH patients: exon 3 G50R, exon 10 D147N, exon 13 I602V, intron 2 190+4 A>T 5' splice donor, and intron 8 1187-10 G>A 3' splice acceptor. Objectives: 1) To determine the presence of LDL-R gene mutation among first degree relatives of patients diagnosed with this genotype; 2) To describe the phenotype among these family members. Methodology: First degree relatives of the index patients identified with novel LDL-R mutations in the 2005 study of Punzalan et al. were recruited for the study. For each participant, clinical history and lipid profile were obtained and the presence/absence of the proband mutation determined via direct sequencing. The p values were calculated using the Kruskall-Wallis test. Results: A total of seven participants from four families were included, two (28.6%) males and five females (71.4%) with ages ranging from 16 to 60 years old and having a mean LDL level of 4.874 mmol/L. Calculated p values show significant association of elevated LDL- and total cholesterol levels as well as HDL ratio with the presence of an LDL-R mutation. Conclusion: Consistent with the known etiology of FH, elevated LDL cholesterol levels was found to be associated with mutations in the LDL-R gene.

2156M

Implantable cardioverter defibrillator (ICD) therapy in two founder arrhythmogenic right ventricular cardiomyopathy (ARVC) populations with mutations in TMEM43 (p.S358L) and PKP-2 (p.Q378X). When genotype matters. *K. Hodgkinson[1,3], P. Boland[2], A. Bullen[3], F. Curtis[3], A. Collier[3], S. Stuckless[1], T-L. Young[3], B. Fernandez[3], A.J. Howes[2], S. Connors[2].* 1) Clinical Epidemiology, Faculty of Medicine, Memorial University, Health Sciences Centre, St John's, NL, Canada; 2) Division of Cardiology, Faculty of Medicine, Memorial University, Health Sciences Centre, St. John's, NL, Canada; 3) Discipline of Genetics, Faculty of Medicine Memorial University, Health Sciences Centre, St. John's, NL, Canada.

Background: Arrhythmogenic right ventricular cardiomyopathy (ARVC) may lead to sudden cardiac death (SCD) due to ventricular tachycardia (VT) for which implantable cardioverter defibrillator (ICD) therapy is warranted. Determining when to provide primary ICD therapy in symptomless individuals is difficult. We assessed the efficacy of ICD therapy in two homogeneous founder cohorts with disease-causing mutations in PKP-2 (p.Q378X) and TMEM43 (p.S358L), the latter known for the significant difference in survival between mutation positive and negative individuals and the significantly worse disease in males than females. Affected status comprised mutation positive and/or clinically positive individuals. Unaffected status could only be defined as mutation negative. Methods: Genetic and clinical information was obtained from an existing database. Of 867 individuals at a priori 50% risk for TMEM43 (p.S358L), mutation status was known in 627 (398 affected). Of 117 individuals at a priori 50% risk for PKP-2 (p.Q378X), disease status was known in 61 (42 affected). The natural history to death in the PKP2 cohort was defined and outcomes following ICD in both cohorts assessed. Results: In 129 TMEM43 (p.S358L) mutation positive individuals with an ICD for primary prophylaxis, survival was significantly improved in both males ($p<0.001$, RR 9.31, 95% CI 3.34-26.00) and females ($p<0.001$, RR 3.57, 95% CI 1.34-9.54). In the PKP2 cohort, there was no survival difference between affected and unaffected individuals ($p=0.533$). Ten affected individuals had ICDs, nine for primary prophylaxis. One had VT requiring defibrillation. One SCD occurred after ICD explantation (implant infection). Six (67%) individuals had inappropriate ICD therapy encompassing 52 defibrillations. Discussion: In the missense TMEM43 (p.S358L) cohort there was a significant survival benefit following ICD. Individuals with the nonsense PKP-2 (p.Q378X) mutation had no difference in survival and thus no benefit from the ICD. However, they had an increased burden of inappropriate defibrillation. Conclusion: ICD therapy for individuals with the PKP-2 (p.Q378X) mutation should occur secondary to VT whereas primary ICD therapy is necessary for individuals with TMEM43 (p.S358L). This translational research supports the necessity for mutation-specific population studies to provide accurate natural history and clinical course information to help direct costs and lower patient burden.

2157S

Elevated risk of abnormal arterial remodeling in relatives of individuals with fibromuscular dysplasia. *A. Katz[1,3], M. Oberdoerster[2], S. Blackburn[2], D. Coleman[2], J. Stanley[2], A. Ruesch[1], J. Douglas[3], J. Li[3], S. Ganesh[1,3].* 1) Cardiovascular Medicine, University of Michigan, Ann Arbor, MI; 2) Vascular Surgery, University of Michigan, Ann Arbor, MI; 3) Human Genetics, University of Michigan, Ann Arbor, MI.

Fibromuscular dysplasia (FMD) is a rare disease of abnormal arterial remodeling, predominantly occurring in women (90%). While prior research has suggested a genetic etiology of FMD, little is known of vascular risk to family members of affected individuals. We hypothesized that compared to the general population, first degree relatives of individuals with FMD have a higher rate of diseases of abnormal arterial remodeling, manifesting as FMD with multi- or uni-focal stenosis, arterial aneurysm, and arterial dissection. Individuals with FMD were enrolled through care received at the University of Michigan and through self-referral to our study. Probands (N = 62) all had physician diagnosed FMD. Medical history of the family members was assessed for: (1) physician diagnosed FMD, (2) clinical signs of suspected but not formally diagnosed FMD, (3) aneurysm without evidence of FMD. Clinical signs of suspected FMD were stroke before the age of 55, hypertension diagnosed before age 30, renal artery stenosis, or sudden death without clear alternative etiology. Familial relative risk was calculated by comparing observed number of cases to expected number of cases among relatives of probands. Among all first degree relatives for which a medical history could be obtained (373; 120 parents, 153 full siblings, and 100 offspring), 2.7% had physician diagnosed FMD, 4.8% had clinically suspected FMD, and 3.5% had an aneurysm. Among female relatives, 5.5% (10/181) had physician diagnosed FMD, and 9.9% (18/181) had physician diagnosed or suspected FMD. Among probands' sisters, 6.3% (5/79) had physician diagnosed FMD. Using an estimate of 0.1% population prevalence of FMD, these data represent a crude unadjusted relative risk of 55 for female first degree relatives and 63 for sisters of FMD affecteds. Among probands' fathers, 11.7% (7/60) had been diagnosed with an abdominal aortic aneurysm (AAA). Using an estimate of 6% prevalence of AAA among men over 65 (from population screening studies), this represents a relative risk of 1.9. Our data indicate an elevated risk of arterial dysplasia in relatives of individuals with FMD, which supports the hypothesis that FMD is a heritable, systemic arteriopathy with variable clinical manifestations that may be specific to gender. These findings have potential clinical screening implications for family members of individuals with FMD. Further studies are needed to delineate the genetic basis and clinical presentations of FMD.

2158M

First 2 years of experience of an integrated multidisciplinary clinic for adults with aortopathies in a Canadian context. *A.M. Laberge[1,2], I. El Hamamsy[2], L. Robb[3], N. Poirier[2], P. Demers[2], F.P. Mongeon[3].* 1) Division of Medical Genetics, Dept of Pediatrics, CHU Sainte-Justine, Montreal, Canada; 2) Dept of Heart Surgery, Montreal Heart Institute, Montreal, Canada; 3) Dept of Medicine, Montreal Heart Institute, Montreal, Canada.

In 2012, the Montreal Heart Institute started an integrated multidisciplinary clinic for adults referred for suspicion of Marfan syndrome or other connective tissue disorders at risk of aortic disease. In this clinic, a specialized heart team (cardiologist specialized in adult congenital heart disease and heart surgeons specialized in aortic surgery) and a genetics team (medical geneticist, genetic counselor) see patients side-by-side. Patients with a family history of aortic disease or systemic features of Marfan syndrome are seen initially by both teams, unless a genetic diagnosis had previously been confirmed. Patients with presumed isolated aortic disease are seen first by the heart team, who determines if evaluation by the genetics team is needed. Special consultation corridors were established with a specialized orthopedist, ophthalmologist and physiatrist. A patient-centered, holistic approach is our standard of care. We report here our experience with this integrated multidisciplinary clinic over our first two years of activities. Between May 2012 and April 2014, 183 new patients were seen, from 146 different families. Reasons for referral included suspicion of Marfan syndrome (72), TAAD (56), suspicion of Loeys-Dietz syndrome (15), suspicion of Ehlers-Danlos syndrome (8), sudden death in the family (6) and various other reasons (26). All were seen by the heart team, 70 were seen by the geneticist for a dysmorphological exam. All had dedicated cardiovascular imaging in our center. Genetic tests were ordered for 35 patients. To date, genetic testing led to identification of mutations in FBN1 in eight patients and ACTA2 mutations in two families. Close links with the equivalent pediatric clinic and the prenatal genetics clinic at the affiliated mother and child university health center facilitated the evaluation of eight children of our adult patients. Most importantly, these close links allowed for the rapid evaluation of three pregnant women at risk of aortic disease and eight affected parents who had no active follow-up at the time their child was seen in the pediatric clinic. Our integrated multidisciplinary approach results in efficient access to specialized cardiac and genetic assessment and rapid management when required. Further follow-up is required to ascertain the value of such dedicated clinics in altering the natural history of these conditions.

2159S

Patterns of discordant phenotypes in familial congenital heart disease. *L.A. Larsen[1], S.G. Ellesøe[2], C.T. Workman[3], P. Bouvagnet[4], K. van Engelen[5], R. Hinton[6], V. Hjortdal[7], C.A. Loffredo[8], K. McBride[9], B.J.M. Mulder[10], A. Postma[11], L. Søndergaard[12], S. Brunak[2,3].* 1) Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark; 2) NNF Center for Protein Research, University of Copenhagen, Copenhagen, Denmark; 3) Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark; 4) Laboratoire Cardiogénétique, Hospices Civils de Lyon, Groupe Hospitalier Est, Bron, France; 5) Department of Clinical Genetics, Academic Medical Centre, Amsterdam, The Netherlands; 6) Division of Cardiology, The Heart Institute, Cincinnati Children's Hospital Medical Center; 7) Department of Cardiothoracic Surgery, Aarhus University Hospital, Skejby, Denmark; 8) Department of Oncology, Lombardi Cancer Center, Georgetown University, Washington, DC; 9) Center for Cardiovascular and Pulmonary Research, Nationwide Children's Hospital, Department of Pediatrics, Ohio State University, Columbus OH; 10) Department of Cardiology, Academic Medical Centre, Amsterdam, The Netherlands; 11) Department of Anatomy, Embryology & Physiology, Academic Medical Centre, Amsterdam, The Netherlands; 12) Department of Cardiology, Rigshospitalet, University Hospital of Copenhagen, Denmark.

Congenital heart disease (CHD) affects approximately 1% of live births. CHD may be caused by both genetic and environmental factors, but much of the molecular pathology remains to be elucidated. Observations of co-occurrence of specific heart defects in CHD families and in specific genetic syndromes suggest that some discordant heart defects may share susceptibility genes or pathways. Here, we have investigated the occurrence of concordant and discordant heart defects in a dataset of 4004 clinically diagnosed CHD cases, from 1175 families with at least two affected individuals per family (663 unpublished families, 512 previously published families). To investigate if specific discordant heart defects co-occur in families, we arranged the dataset into 62 anatomical subgroups of heart defects. Next, to determine the strength of association between pairs of anatomical subgroups, we calculated the log odds ratio of familial occurrence for each pair. Hierarchical clustering of the observed log odds ratios revealed specific patterns of occurrence of discordant heart defects within families, supporting a genetic relationship between specific cardiac phenotypes. The data show that defects typical for heterotaxy co-occur in families and suggest that left-sided defects rarely co-occur with right-sided defects. Targeted deletion of specific genes in mouse models often results in more than one type of heart defect, and for a specific subgroup of heart defects several overlapping susceptibility genes may exist in mouse models. We used published genotype-phenotype data from mouse models to investigate if the observed patterns of occurrence in families could be explained by shared susceptibility genes; we compared the pairwise discordance log odds ratios in families with pairwise subgroups of heart defects observed in mouse models, and we calculated the frequency of overlapping susceptibility genes between the pairs of phenotypes. The analysis revealed correlation between phenotype-pairs with high familial occurrence and phenotype-pairs with a large number of overlapping susceptibility genes in mice. For example; ventricular septal defects and interrupted aortic arch show high familial occurrence and in mouse models 14 susceptibility genes are overlapping between these two groups. In summary, we observed significant clustering of specific discordant CHD phenotypes within families and our data suggest that shared susceptibility genes may be an underlying mechanism.

2160M

Titin As a Gene for Conduction Defects With and Without Cardiomyopathy. *E. Smith[1], A. Mani[1,2].* 1) Program for Cardiovascular Genetics, Department of Internal Medicine/Section of Cardiology, Yale University School of Medicine, New Haven, CT; 2) Department of Genetics, Yale University School of Medicine, New Haven, CT.

Radical mutations in the gene Titin (TTN) have been associated with dilated cardiomyopathy. Mutations in this gene have been also reported to be associated with arrhythmogenic right ventricular cardiomyopathy, but not with isolated arrhythmia. We report three cases of sudden cardiac death due to ventricular fibrillation, where the underlying genetic causes have been narrowed down to a radical, protein altering TTN mutation. Case one is a young woman who had cardiac arrest due to ventricular fibrillation while driving a car. After resuscitation she underwent a single chamber implantable cardioverter defibrillator (ICD) placement. Post-arrest, she had several runs of non-sustained ventricular tachycardia and frequent bouts of premature ventricular excitations. Genetic testing by whole exome sequencing (WES) identified a TTN frameshift (p.G8858fs) mutation. She was later found to have mildly dilated left ventricle by echocardiography examination. The subject had family history of unspecified sudden death at young ages. Case two is a gentleman who had ventricular fibrillation at home. He had a history of paroxysmal atrial fibrillation. He too had a single chamber ICD implantation post-arrest. There is no history of cardiomyopathy or sudden death in the family. His genetic testing by WES revealed a TTN nonsense mutation (p.R18966X). He did not meet the criteria for dilated cardiomyopathy and instead had mild left ventricular hypertrophy. Case three is a woman who also had ventricular fibrillation while out with friends. Though she carried he diagnosis of hypertrophic cardiomyopathy post-arrest, serial echocardiograms did not corroborate this finding. An automatic implantable cardioverter defibrillator (AICD) was implanted post-arrest. Her family history was significant for young death in her maternal grandparents and great-grandfather. Genetic testing revealed a splice site substitution (c.28091-2A>C). These three isolated cases raise the possibility that radical TTN mutations may underlie malignant ventricular arrhythmia with and without cardiomyopathy, which constitutes a new phenotype for this gene. Further studies are needed to establish the association and to determine the mechanism.

2161S

Cystatin C and cardiovascular disease: a Mendelian randomization study. *S.W. van der Laan[1], T. Fall[2], J. van Setten[1,3], P.I.W. de Bakker[3,4], G. Pasterkamp[1], J. Ärnlöv[2], M.V. Holmes[5], F.W. Asselbergs[3,6,7,8] on behalf of the Cystatin C MR Consortium.* 1) Experimental Cardiology, UMC Utrecht, Utrecht, the Netherlands; 2) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 3) Department of Medical Genetics, UMC Utrecht, Utrecht, the Netherlands; 4) Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht, the Netherlands; 5) Penn Medicine, University of Pennsylvania Health System, United States of America; 6) Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom; 7) Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, the Netherlands; 8) Department of Cardiology, UMC Utrecht, Utrecht, the Netherlands.

Observational studies show a strong dose-rsponse association between circulating Cystatin C (encoded by *CST3*) and incident coronary heart disease (CHD), independent of traditional risk factors and renal function. This supports the hypothesis that circulating Cystatin C could represent a causal factor for CHD. However, residual confounding and reverse causality could be alternative explanations that are difficult to tease from observational studies. We sought to investigate the causal role of Cystatin C in CHD development by conducting a Mendelian randomization (MR) analysis using a common variant in the *CST3* locus.

We conducted a MR analysis in 43 studies including 219,219 individuals with 37,321 measures of Cystatin C and 41,162 CHD events. We used rs911119 (or a proxy, r^2>0.90) in the *CST3* locus (identified previously by GWAS) as a genetic instrument for MR analysis.

Cystatin C associated with risk of CHD in an observational analysis adjusted for age and sex (odds ratio [OR] 2.20; 95% confidence interval [CI]: 1.90, 2.57 per doubling of Cystatin C concentration; p=8.87×10^{-31}); additional adjustment for confounders (smoking, HDL-cholesterol, BMI, CKD-EPI, and systolic blood pressure) diminished the association (OR 1.60; 95%CI 1.34, 1.96 per doubling of Cystatin C concentration; p=9.09×10^{-7}). Rs911119 had a strong effect on circulating Cystatin C levels (-0.061; 95%CI -0.066, -0.057; p=4.49×10^{-149} per effect allele). However, the variant did not show significant association with risk of CHD (OR 1.01 (95%CI 0.99, 1.03; p=0.41).

In summary we replicated the association of Cystatin C with CHD risk and show a strong association of rs911119 with circulating Cystatin C. However, we find no evidence for a causal role of Cystatin C in the development of CHD.

2162M

Association of oxidative DNA damage with Folic acid metabolizing genes in Children with congenital septal defects. *S. Syed[1], K. Koneti[2], K. Kola[1], D. Dadala[3], G. Gundimeda[4], A. Akka[1], M. Mundluru[1].* 1) Institute of Genetics, Osmania University, Hyderabad, India; 2) Care Hospital, Banjara Hills, Hyderabad; 3) National Institute of Nutrition, Tarnaka, Hyderabad; 4) Indo american cancer hospital hyderabad.

Association of oxidative DNA damage with Folic acid metabolizing genes in Children with congenital septal defects Sunayana Begum Syed*, Nageswara Rao Koneti**, Srujana Kola*, Sujatha Dadala***, Sandhya Devi Gundimeda****, Jyothy Akka*, Hema Prasad Mundluru* * Institute of Genetics & Hospital for Genetic Disease, Osmania University, Begumpet, Hydreabad. **Care Hospital, Banjara Hills, Hyderabad. ***National Institute of Nutrition, Tarnaka, Hyderabad **** indo american cancer hospital, Hyderabad Background: Congenital Septal Defect (CSD) is the most common developmental anomalies, occurs 1 in 3000 live births with unknown disease etiology. The present study was designed to understand the possible etiological factors using FISH and comet assay for chromosomal and oxidative DNA damage. SNPs of folic acid metabolism and antioxidant genes MTHFR, MTRR, RFC and GST were studied by PCR-RFLP for the oxidative stress induced abnormal cardiac development during embryogenesis. Methods: The present study includes 162 children with CSD and their mothers and age matched controls (n=174). Blood samples were collected from Department of Pediatric Cardiology, Care hospital, Hyderabad with prior clearance from institutional ethics committee and written consent from parents. Lymphocytes were isolated for comet assay and FISH analysis. RBCs were used for the estimation of folic acid and Vitamin 12 levels. Inflammatory marker like C-reactive protein (CRP) was analyzed by ELISA for correlating with DNA damage. Genomic DNA was isolated for PCR based RFLP analysis of SNPs 677C>T, A66G, A80G, GSTM1 and GSTT1 genes. Results: Low RBC folate and high CRP levels were found in patients than the control group, demonstrating the severity of oxidative DNA damage. Fasting vitamin B12 levels in CSD patients and their mothers were significantly decreased than the control group. 22q11.2 micro deletions and significant DNA damage were observed in CSD patients. In MTHFR C677T the overall mutant T allele frequency was higher in patient. Similarly the frequencies of 'G/G' alleles of the genes MTRRA66G / RFC1A80G were also found to be 2 fold increased in CSD patients. Among the mothers of CSD children a 3 fold increase in frequencies of mutated 'T'/'G'/'G' allele was observed when compared to the controls. Conclusion: Elevated DNA damage, increased CRP levels and decreased folate, vitamin B12 indicates that oxidative stress plays an essential role in CSD.

2163S

Causal role of alcohol consumption in blood levels of lipids and hemostatic factors, and risks of coronary heart disease an ischemic stroke. *K. Vu[1], E. Boerwinkle[1,2], A.C. Morrison[1].* 1) School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA; 2) The Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA.

Background: Moderate alcohol use may reduce the risk of coronary heart disease (CHD) and ischemic stroke (IS) through an improved blood lipid profile and reduced hemostatic factor levels. However, the observed association may be confounded. Mendelian randomization (MR) can help to examine the causality of the association. Methods: The analysis involved 10,893 European American (EA) and 3,856 African American (AA) from the Atherosclerosis Risk of Communities (ARIC) study. Common and rare variants in alcohol metabolism genes, alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), and cytochrome P450 (CYP2E1), were selected as the genetic instruments (25 in EA, 28 in AA). A stringent process was applied to examine the genetic instruments against MR assumptions and possible violations including linkage disequilibrium and pleiotropy. The instruments were combined into unweighted genetic scores. Alcohol use was measured by gram per week (g/wk). Lipids included triglycerides (TG), and high-density lipoprotein (HDL) and low-density lipoprotein (LDL) levels. Hemostatic factors included fibrinogen, factor VII, factor VIII, and von Willebrand factor (vWF). To examine for a possible J-shape effect of alcohol use, a sensitivity analysis excluding heavy drinkers (> 210 g/wk) was conducted. Results: In EA and AA, log alcohol use causally increased the log HDL levels (EA: β = 0.038, 95% CI: 0.001, 0.076; AA: β = 0.065, 95% CI: 0.026, 0.104), and reduced log factor VIII (EA: β = -0.039, 95% CI: -0.074, -0.003; AA: β = -0.031, 95% CI: -0.061, -0.001). A causal reduction in CHD incidence was also observed in EA (risk difference (RD) = -9.5%, 95% CI: -15.8%, -3.3%) and AA (RD = -4.8%, 95% CI: -7.4%, -2.3%). Among EA only, alcohol use causally reduced LDL levels (β = -5.715, 95% CI: -11.348, -0.082), TG (β = -0.100, 95% CI: -0.172, -0.028), and factor VII (β = -0.038, 95% CI: -0.072, -0.003). In AA, alcohol use causally reduced IS incidence (RD = -3.2%, 95% CI: -4.5%, -1.9%). No evidence was found for a causal link between alcohol use and fibrinogen and vWF. Only 5.9% EA and 4.2% AA were heavy drinkers and the sensitivity analysis yielded similar results. Conclusions: This study supports the causal role of moderate alcohol use in a raised blood level of HDL, reduced blood level of factor VIII, and reduced risk of CHD among EA and AA. The analysis also suggests a race-specific causal role of moderate alcohol use in a reduction of LDL, TG, factor VII, and risk of IS.

2164M

Mendelian randomisation study of alcohol and cardio-metabolic risk factors. *I.Y. Millwood[1], L. Li[2,3], R.G. Walters[1], W. Mei[1], D. Bennett[1], Y. Guo[2], Z. Bian[2], R. Peto[1], R. Collins[1], S. Parish[1], R. Clarke[1], Z. Chen[1], China Kadoorie Biobank collaborative group.* 1) Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford; 2) School of Public Health, Peking University Health Science Center, Beijing, China; 3) Chinese Academy of Medical Sciences, Dong Cheng District, Beijing, China.

Background: To explore the causal relationship between alcohol and cardio-metabolic risk factors in a large Chinese population, a Mendelian randomisation approach was taken using alcohol metabolising gene variants *ADH1B*2* and *ALDH2*2*. **Methods:** The study includes 82,155 men and women aged 30-79 years from the China Kadoorie Biobank prospective cohort, recruited in 2004-2008 from ten regions across China, with baseline data on alcohol consumption, physical measurements, random blood glucose and *ADH1B* (rs1229984) and *ALDH2* (rs671) genotypes. **Results:** 10,929 men (34%) and 1014 (2%) women were weekly drinkers, consuming mean (SE) 40.9 (0.3) and 16.7 (0.6) g alcohol/day, respectively. Allele frequency of *ADH1B*2* was 0.69 and *ALDH2*2* was 0.20, and both varied by region. For each acetaldehyde increasing allele, alcohol intake was lower by 3.4 (0.5) g/day (*ADH1B*2*) and 14.0 (0.9) g/day (*ALDH2*2*) in men, and by 1.8 (0.8) g/day (*ADH1B*2*) and 5.1 (1.6) g/day (*ALDH2*2*) in women, adjusted for age and region. Table 1 shows the association of reported alcohol consumption (adjusted for covariates) and alcohol predicted from a weighted genetic risk score (GRS) combining *ADH1B*2* and *ALDH2*2* (adjusted for age and region), with outcomes, in male weekly drinkers (**P<0.05, **P<0.0001). Results in female weekly drinkers were generally consistent, but not significant.

Beta (SE) 10g/d	SBP (mmHg)	DBP (mmHg)	BMI (kg/m2)	Waist:hip ratio	Waist (cm)	Body fat (%)	Glucose (mmol)
Reported-alc	0.59 (0.06)**	0.40 (0.03)**	0.01 (0.01)	0.001 (0.000)**	0.10 (0.03)**	0.04 (0.02)*	0.02 (0.01)*
GRS-alc	1.90 (0.32)**	1.61 (0.19)**	0.30 (0.05)**	0.009 (0.001)**	1.21 (0.15)**	0.79 (0.97)**	0.18 (0.04)**

Conclusion: Alcohol has a strong causal effect on blood pressure, with 10 g/day (≈one drink) increasing SBP by about 2 mmHg, and a modest adverse effect on glucose and measures of overall and central adiposity. Stronger effects seen using gene variants could reflect lifelong consumption patterns, and the limitations of self-reported questionnaire data.

2165S

Heritability and linkage study on heart rate variability in an isolated Arab population. *L. Munoz[1], M.O Hassan[2], V.S. Voruganti[3,4], D. Jaju[2], S. Albarwani[2], A. Aslani[1], R. Bayoumi[2], S. Al-Yahyaee[2], A.G. Comuzzie[4], B.Z. Alizadeh[1], I. Nolte[1], H. Snieder[1].* 1) Department of Epidemiology, University Medical Center Groningen, University of Groningen, The Netherlands; 2) College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Sultanate of Oman; 3) Department of Nutrition and UNC Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC, USA; 4) Department of Genetics, Texas Biomedical Research Institute, Texas, USA.

Background: Heart rate variability (HRV) is a reliable, non-invasive, economical reflection of many physiological and psychological factors regulating the normal heart rhythm. It is often used as an index of cardiac parasympathetic system activity. In the general population decreased HRV is an independent risk factor for morbidity and mortality. Individual differences of HRV can be attributed to genetic factors and are found to be more pronounced in subjects during stress tests. However, most published studies on the genetic influence on HRV have focused on subjects of European and African-Americans descent. In Oman Family Study (OFS) we aim to (a) estimate and quantify the contribution of genes to the variance of HRV; (b) calculate the overlap in genetic and environmental influences on HRV at rest and under stress using uni-and bivariate analyses; and (c) perform a multipoint variance component-based linkage analysis for all HRV parameters. Methods: The OFS consists of five large, multigenerational pedigrees of Arab families living in Willayat (state) of Nizwa. The strengths of the OFS are: (1) geographically isolated which provides a more homogeneous environmental exposure, (2) similar socio-economic status and health-related habits, (3) highly consanguineous, and (4) authenticity and accessibility of genealogical records. HRV variables (SDNN, RMSSD, HF, LF, VLF, and TP) and heart rate (HR) were measured from beat-to-beat heart rate data at rest and during physical (cold pressor test, CPT) and mental (word conflict test, WCT) stress tests for 1274 participants. Solar software was used to calculate the heritability estimates and to perform the linkage analyses. Results: Univariate and bivariate heritability (h2) estimates at rest and for the two stress tests were all highly significant ($p<0.05$). Linkage is currently being analyzed for all HRV measurements. Conclusions: In the OFS, additive genetic factors significantly influence all HRV parameters and especially HR. Heritability estimates at rest and during physical stress were lower than during mental stress, in particular for SDNN and RMSSD. Bivariate heritability analysis showed evidence that the same genetic factors that influence HRV at rest, influence HRV during both mental and physical stress tests.

2166M

The Familiality of Extreme Lipid Values. *S. Knight[1,2], J.B. Muhlestein[1,2], J.F. Carlquist[1], J.L. Anderson[1,2], B.D. Horne[1,2].* 1) Intermountain Medical Center, Murray, UT; 2) Department of Medicine, University of Utah, Salt Lake City, UT.

Background: It is estimated that ~40% of the variability in blood lipid values is heritable. However, little is known about the familiality of extreme lipid values, thus, we examined the familiality of extreme total cholesterol, low density lipoprotein cholesterol (LDL-C), high density lipoprotein cholesterol (HDL-C), and triglycerides. Methods: The Intermountain Genealogical Registry was used for this study. This registry contains the genealogy for over 700,000 patients in the Intermountain Healthcare (IHC) system. The extreme trait cases were defined as individuals who had at least 3 high/low value and at least 1 extreme value - Total cholesterol (high value ≥240 mg/dL & extreme value ≥300 mg/dL), LDL-C (≥160 mg/dL & ≥190 mg/dL), HDL-C (≤40 mg/dL & ≤20 mg/dL), and triglycerides (≥200 mg/dL & ≥500 mg/dL). We further restricted these cases to those that linked to a pedigree that has a minimum of 3 generations and 12 ancestors. For each extreme lipid trait, we generated a genealogy index of familiality (GIF), which is 10,000 times the average pairwise kinship of coefficient of cases. The significance of the GIF is assessed empirically using 1000 set of controls IHC patients with 3 generations and 12 ancestors. Results: The total numbers of extreme cases for the GIF analyses were 395 for cholesterol, 534 for LDL, 441 for HDL and 790 for triglycerides. The GIF for cholesterol was 0.3638, which was not significant compared to the matched controls (Average control set GIF = 0.448, p=1.00). However, the GIF for LDL-C (GIF=0.5035), HDL-C (GIF= 0.5127), triglyceride (GIF=0.6358) were all significant (all pvalues<0.01). The greatest contribution to the GIF for triglycerides and second greatest for LDL-C came from case pairs that were related by 2 meioses (e.g., siblings), while the greatest for HDL-C was case pairs that were separated by 6 meioses. This may indicate shared environmental factors contribute more to the heritability of triglycerides and LDL, than contribute to HDL-C. Conclusions: Extreme values of individual lipid traits (LDL-C, HDL-C, triglycerides) were familial, although extreme total cholesterol levels might not have a familial component. These results also suggest that shared environment may play a greater role in extreme LDL-C and triglycerides values, than they do in HDL-C.

2167S

Characteristics of Aortic Disease Associated with ACTA2 mutations. *E.S. Regalado[1], D. Guo[1], S. Prakash[1], T.A. Bensend[1], K. Flynn[1], A. Estrera[2], H. Safi[2], D. Liang[3], J. Hyland[4], A. Child[5], G. Arno[5], C. Boileau[6], G. Jondeau[7], A. Braverman[8], R. Moran[9], T. Morisaki[10], H. Morisaki[10], R. Pyeritz[11], J. Coselli[12], S. LeMaire[12], D.M. Milewicz[1], Montalcino Aortic Consortium.* 1) Department of Internal Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA; 2) Department of Cardiothoracic and Vascular Surgery, University of Texas Health Science Center at Houston, Houston, TX, USA; 3) Department of Medicine, Stanford University Medical Center, Stanford, CA, USA; 4) Connective Tissue Gene Tests, Allentown, PA, USA; 5) Department of Cardiac and Vascular Sciences, St George's, University of London, London, UK; 6) AP-HP, Hopital Bichat, Centre National de Reference pour le syndrome de Marfan et apparentes, Paris, France; Universite Paris 7, Paris, France; AP-HP, Hopital Bichat, Laboratoire de Genetique moléculaire, Boulogne, France; INSERM, U1148, Paris, France; 7) AP-HP, Hopital Bichat, Centre National de Reference pour le syndrome de Marfan et apparentes, Paris, France; Universite Paris 7, Paris, France; AP-HP, Hopital Bichat, Service de Cardiologie, Paris, France; INSERM, U1148, Paris, France; 8) Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO, USA; 9) Cleveland Clinic, Cleveland, OH, USA; 10) Department of Bioscience and Genetics, National Cerebral and Cardiovascular Center Research Institute, Osaka, Japan; 11) Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA; 12) Division of Cardiothoracic Surgery, Michael E. DeBakey Department of Surgery, Baylor College of Medicine, the Texas Heart Institute, and Baylor St. Luke's Medical Center, Houston, TX, USA.

ACTA2 is the most frequently mutated gene causing familial thoracic aortic aneurysms and dissections and is responsible for up to 20% of cases. We characterized the aortic disease, management, and outcome associated with the first aortic event (defined as aortic dissection or aortic aneurysm repair) in a large case series of individuals (n=277) with 41 ACTA2 mutations. Aortic events occurred in 48% of the cohort, with the vast majority presenting with thoracic aortic dissections (88%) at variable ages ranging from 12 to 76 years and associated with a 25% mortality rate. Type A dissections were more common than type B dissections (54% versus 21%), but the median age of onset of type B dissections was significantly younger than type A dissections (27 vs. 36 years of age). Among individuals presenting with type A dissections, the maximum diameter ranged from 3.8- 9.5 cm, and one-third of these individuals had dissections at diameters <5.0 cm. Only 12% of aortic events were repair of ascending aortic aneurysms at diameters ranging from 4.2-6.5 cm (median 5.2 cm), with maximal diameter in the ascending aorta in half of the individuals. Overall cumulative risk of aortic event (i.e., penetrance) at age 85 years was only 0.76 (95% CI 0.64, 0.86). After adjustment for intra-familial correlation, gender, and race, mutations disrupting p.R179 and p.R258 were associated with higher penetrance of aortic events, whereas the mutations altering p.R185 and p.R118 showed significantly lower penetrance of aortic events. These findings demonstrate the following features of aortic disease caused by ACTA2 mutations differ from Marfan syndrome: higher rates of thoracic aortic dissections; higher prevalence of ascending aortic aneurysms involving the aortic root and/or aortic arch; decreased penetrance of aortic events for the majority of ACTA2 mutations; mutation-specific differences in penetrance and age of onset. The lack of syndromic features in ACTA2 patients most likely delays the diagnosis and proper management of aortic disease, therefore making these patients more likely to present with acute dissections. Integration of these data into genetic counseling and medical management as well as early identification of ACTA2 patients by genetic testing may reduce the extensive risk of aortic dissections. Future investigations will focus on identifying genetic and environmental modifiers in individuals with ACTA2 mutations and reducing the dissection rates in these patients.

2168M

Are short telomeres are cause or consequence of hypertension in spontaneously hypertensive mice? *C.L Chiu, N.L Hearn, D. Paine, N. Steiner, J.M Lind.* University of Western Sydney, School of Medicine, New South Wales, Australia.

Telomere length is widely considered as a marker of biological ageing. Clinical studies have reported associations between reduced telomere length and hypertension. However, the role of telomere length in the pathogenesis of hypertension, and whether reduced telomere length is a cause or consequence of disease, remains unknown. The aim of this study was to determine whether telomere shortening occurred prior to the onset of disease in spontaneously hypertensive mice. Spontaneously hypertensive Schlager mice (BPH/2J) and their normotensive controls (BPN/3J) were used in this study. Genomic DNA was extracted from kidney tissue of 4, 12 and 20 week old male BPH/2J and BPN/3J mice (n=10/group). Relative telomere length (T/S) was measured using quantitative PCR. Linear correlation estimates were performed to analyse telomere length over time within a strain. A general linear model with repeat measures testing was used to compare rate of telomere shortening between groups. A general linear model was used to compare relative telomere lengths and gene expression between groups. $P<0.05$ was considered significant. Age was inversely related to telomere length. In 4 week old pre-disease animals no difference in T/S was observed between BPH/2J and BPN/3J animals ($p=0.09$). The rate of telomere attrition between BPH/2J and BPN/3J was significantly different ($p=0.001$). At 12 and 20 weeks, established disease, BPH/2J animals had significantly shorter telomeres when compared to their age-matched controls (12 weeks $p<0.001$ and 20 weeks $p=0.004$). This is the first study to show that reduced telomere length occurs after the development hypertension, indicating that this is not the cause of hypertension in spontaneously hypertensive mice. Further studies are needed to determine the mechanisms which lead to the development of hypertension and the shortening of telomeres in these animals.

2169S

Infusions of dexamethasone loaded erythrocytes in ataxia teleangiectasia patients. *L. Chessa[1], V. Leuzzi[2], R. Micheli[3], M. Piane[1], D. D'Agnano[2], A. Molinaro[4], T. Venturi[2], A. Plebani[5], AR. Soresina[5], E. Fazzi[3], M. Marini[5], P. Ferremi Lealì[5], I. Quinti[6], FM. Cavaliere[6], MC. Pietrogrande[7], A. Finocchi[8], M. Magnani[9,10].* 1) Dept. Clinical Molecular Medicine, Sapienza University, Roma, Italy; 2) Dept. Pediatrics Child Neurology and Psychiatry, Sapienza University, Roma, Italy; 3) Child Neurology and Psychiatry, Spedali Civili and University, Brescia, Italy; 4) School Reproductive and Developmental Science, University of Trieste and Brescia, Italy; 5) Dept. Clinical Experimental Sciences, Spedali Civili and University, Brescia, Italy; 6) Dept. Molecular Medicine, Sapienza University, Roma, Italy; 7) Dept. Pediatrics, Milano University and Fondazione IRCCS Ca' Granda, Milano, Italy; 8) Dept. Pediatrics, Ospedale Pediatrico Bambino Gesù and University Tor Vergata, Roma, Italy; 9) Dept. Biomolecular Sciences, University Carlo Bo, Urbino, Italy; 10) Erydel S.p.A, Urbino, Italy.

Background. Ataxia Teleangiectasia (AT) is a rare devastating neurodegenerative disease presenting with early onset ataxia, oculocutaneous teleangiectasias, immunodeficiency, radiosensitivity, and proneness to cancer. In a previous phase II study we showed that 6 monthly infusions of autologous erythrocytes loaded with dexamethasone (EryDex©) were effective in improving the neurological impairment in young AT patients. Here we report the results of the extension of this study for an additional 24 month-period. Methods. After the end of first trial, 4 subjects continued to be treated with monthly EryDex infusions for further 24 months and their clinical outcome was compared with that of 7 age-matched subjects who had stopped the treatment after the first 6 infusions. The protocol included serial assessment of ataxia (by International Cooperative Ataxia Rating Scale) and adaptive behavior (by Vineland Adaptive Behavior Scales), and clinical and laboratory tests revealing treatment- and steroid-dependent adverse effects, if present. Results. Patients in the extended study experienced a continuous neurological improvement with respect to their pre-treatment status while control subjects showed a progressive neurological deterioration (according to the natural history of the disease) after the discontinuation of the treatment. The delivery system we adopted proved to be safe and well-tolerated and none of the side effects usually associated with the chronic administration of corticosteroids was observed during the entire trial. Conclusion. These preliminary promising results call for a wide-scale controlled study on protracted treatment of AT patients with dexamethasone loaded erythrocytes.

2170M

Inta-nasal DDAVP administration for the prevention of massive subcutaneous hematoma in dermatan 4-O-sulfotransferase 1 (D4ST1)-deficient Ehlers-Danlos Syndrome (DDEDS). *T. Kosho[1], M. Ishikawa[1], E. Kise[1], J. Takahashi[2], S. Yuzuriha[3], Y. Fukushima[1].* 1) Dept Med Gen, Shinshu Univ Sch Med, Matsumoto, Japan; 2) Department of Orthopedics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Department of Plastic and Reconstructive Surgery, Shinshu University School of Medicine, Matsumoto, Japan.

Dermatan 4-O-sulfotransferase 1 (D4ST1)-deficient Ehlers-Danlos syndrome (DDEDS), caused by recessive loss-of-function mutations in CHST14, is a recently delineated form of EDS [Dündar et al., 2009; Malfait et al., 2010; Miyake et al., 2010; Kosho et al., 2011], characterized by a unique set of clinical features consisting of progressive multisystem fragility-related manifestations (skin hyperextensibilty and fragility, progressive spinal and foot deformities, massive subcutaneous hematoma) and various malformations (facial features, congenital multiple contractures). Massive subcutaneous hematoma is the most serious complication for patients with DDEDS, which could cause significant reduction of ADL and QOL. In this report, we present experiences of three patients (7-years-old girl; 8-years-old girl; 13-years-old boy) suffering from recurrent serious hematomas and receiving intra-nasal spray of 1-desamino-8-D-arginine vasopressin (DDAVP), the only agent described to have efficacy in the control of bleeding in other types of EDS [Kimo and Becton, 1997]. First, clinical trial was performed before the therapy. Having 1 puff (300 micrograms) of DDAVP, vital signs were not changed excluding reduced urine output. Bleeding time was significantly decreased (3 to 1 minute; 4 to 2.2 minutes; 2.4 to 1 minute) with no change in PT, with mild decrease in APTT, and significant increase in activities of VIII factor and vWF. Then, they started to have intra-nasal DDAVP at an event that could cause a large hematoma. Although hematomas, typically, became large rapidly with severe pain, intra-nasal DDAVP relieved pain significantly as well as enlargement of hematomas in several minutes. The results showed efficacy of DDAVP for the prevention of massive hematoma and suggested the mechanism of this efficacy to be attributable not only to increased activities of VIII and vW factors but to rapid contraction effects of DDAVP for small arteries that might ruptured easily and not contracted normally.

2171S

Catatonia in Down syndrome; the overlooked, treatable cause of regression. *J.H. Miles[1], G. Yao[1], M. Ithman[1], A. Nassiri[2], N. Ghaziuddin[2].* 1) Dept Child Hlth, Med Gen Div, Thompson Ctr, Dept Psychiatry, Dept Biological Engineering, University of Missouri, Columbia, MO; 2) Dept Psychiatry, University of Michigan, Ann Arbor, MI.

Background: A significant minority of individuals with Down syndrome (DS) regress in motor, behavior, intellectual functioning, mood and activities of daily living during adolescence or as young adults. Catatonia, which produces these symptoms, is not considered in DS since clinicians who care for DS are generally unfamiliar with Catatonia outside Schizophrenia. Life threatening Catatonia associated with autonomic dysfunction can be triggered by neuroleptic medications used when patients are misdiagnosed. DS care guidelines don't reference Catatonia. **Objectives**: Alert physicians to Catatonia in DS and stimulate research on pathophysiology of catatonic regression. **Methods**: Five DS adolescents who experienced regression are reported. Extensive testing to rule out other causes of motor and cognitive regression was within normal limits. Based on the presence of multiple motor disturbances (slowing, grimacing, posturing), the adolescents were diagnosed with unspecified Catatonia and treated with anti-catatonic treatments (benzodiazepines/electroconvulsive therapy). One patient who had a Pupillary Light Reflex (PLR) study prior to developing Catatonia was followed with PLRs over his treatment course. **Results**: All five cases responded to initial treatment with a benzodiazepine, but required electroconvulsive therapy (ECT) to achieve remission. Longer duration Catatonia responded more slowly. All recovered baseline or near baseline functioning. A DS specific Diagnostic Protocol and Catatonia Improvement Scale were developed to standardize diagnosis and monitor treatment response over time. Changes in pupil constriction coincident with Catatonia implicate autonomic involvement. A survey of adolescents and young adults who attended one DS clinic as young children suggests a tentative prevalence of catatonic regression of 3%. **Conclusions**: We suspect Catatonia is a common, overlooked cause of unexplained deterioration in adolescents and young adults with DS. Being unaware of Catatonia in DS results in years of incapacitating illness and a risk of malignant catatonia, which may be induced by treatment with serotonergic or dopaminergic medications. Alerting geneticists to Catatonia in DS is crucial to prompt diagnosis, appropriate treatment, prevention of life threatening neuroleptic syndrome and discovering the frequency and course of this disorder. PLR evidence of autonomic dysregulation may suggest why individuals with DS are at risk for Catatonia.

2172M

To study anticancerous and antiproliferative effect of Zerumbone and Oridonin on HeLa cell lines. *N. Gill, A. Walia, V. Vanita.* Department of Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

Purpose: The aim of present study was to evaluate anticancerous properties of Zerumbone (sesquiterpene phytochemical from a type of edible ginger) and Oridonin (a natural diterpenoid purified from Rabdosia rubescens, a Chinese herb) on HeLa cell lines (cervical cancer human origin). **Methodology:** HeLa cells were grown in DMEM medium supplemented with 10% FBS, 100 IU/mL Penicillin, 100μg/mL Streptomycin and Gentamycin at 37 °C in humid environment containing 5% CO2. HeLa cells were treated with different concentrations of Zerumbone and Oridonin (100 μmol/L, 50 μmol/L, 25 μmol/L, 20 μmol/L, 10 μmol/L, 5 μmol/L, 2.5 μmol/L). Cell viability and toxicity was assessed using the 3-(4, 5-dimethylthiazol-(2)-yl)-2, 5-diphenyl tetrazolium bromide (MTT) assay so to ascertain the dose effective against HeLa cell line. Morphological analysis, nuclear condensation, and fragmentation of chromatin in cells treated with different concentrations of Zerumbone and Oridonin were studied by confocal laser staining microscope using DAPI (4', 6-diamidino-2-phenylinodole), a nuclear stain. **Results:** The concentration-dependent cytotoxic activity of Zerumbone and Oridonin has been observed in HeLa cells. Zerumbone and Oridonin at a concentration of 20 μmol/L, 10 μmol/L, 5 μmol/L, 2.5 μmol/L after 24 hours showed decreased HeLa cells viability. On morphological examination, in contrast to control cells, the cells exposed to the lowest concentration of 2.5 μmol/L Oridonin and Zerumbone provoked apoptotic changes, such as cell shrinkage, nuclear fragmentation, and formation of apoptotic bodies. **Conclusion:** In the present study, we demonstrated by MTT assay and confocal studies that Zerumbone and Oridonin significantly inhibited HeLa cell growth and induced apoptosis in a dose-dependent manner.

2173S

Subjects treated with Migalastat continue to demonstrate stable renal function in a Phase 3 extension study of Fabry Disease. *D.G. Bichet[1], D.P. Germain[2], R. Giugliani[3], D. Hughes[4], R. Schiffmann[5], W. Wilcox[6], J. Castelli[7], E.R. Benjamin[7], J. Yu[7], J. Kirk[7], N. Skuban[7], J. Barth[7] On behalf of the FACETS study investigators.* 1) Hôpital du Sacré-Coeur, University of Montreal, Montreal, Quebec, Canada; 2) Hôpital Raymond Poincaré (AP-HP), University of Versailles - St. Quentin en Yvelines (UVSQ), Garches, France; 3) Medical Genetics Service, HCPA/UFRGS Porto Alegre, Brazil; 4) Royal Free Campus, University College London, London, UK; 5) Baylor Research Institute, Dallas, TX; 6) Department of Human Genetics, Emory University, Georgia; 7) Amicus Therapeutics, Cranbury, NJ.

Objectives: Fabry disease is caused by more than 800 different mutations in GLA, the gene encoding the lysosomal enzyme α -galactosidase A (α-Gal A), resulting in accumulation of globotriaosylceramide (GL-3 or Gb3) and serious morbidity and mortality. Migalastat (1-deoxygalactonojirimycin HCl, AT1001) is an orally-administered investigational pharmacological chaperone that selectively binds and stabilizes α-Gal A, leading to increased cellular levels and greater lysosomal activity. Renal function in the Phase 3 FACETS study (Study AT1001-011) was assessed over 18-24 months of treatment with migalastat using several different measures of glomerular filtration rate (GFR). After completing the FACETS study, subjects were offered continued treatment in a separate long term extension study (Study AT1001-041) where additional long-term GFR data were collected. **Methods:** The FACETS study included a 6-month double-blind placebo-controlled period (Stage 1), a 6-month open-label period (Stage 2), and a 12-month extension period. Sixty-seven subjects (24 males, 43 females) were randomized and 48 subjects (17 males, 31 females) entered long term extension Study AT1001-041. Fifty of the 67 randomized subjects had GLA mutations amenable to treatment with migalastat based on an in vitro human embryonic kidney cell assay. Analyses of eGFR (CKD-EPI, MDRD) were performed in subjects with amenable GLA mutations across both studies. **Results:** As previously reported for the FACETS study, GFR in subjects with amenable GLA mutations remained stable over 18-24 months of treatment, with mean annualized eGFR changes of -0.30 ± 0.66 (CKD-EPI eGFR) and +0.79 ± 1.03 (MDRD eGFR) mL/min/1.73m2/yr. Preliminary analysis indicates that renal function has continued to remain stable with an additional average of 6 months of treatment with migalastat in Study AT1001-041, with mean annualized eGFR changes over 24-30 months of +0.29 ± 0.64 (CKD-EPI eGFR) and +1.22 ± 0.87 (MDRD eGFR) mL/min/1.73m2/yr. **Conclusions:** Treatment with migalastat was associated with stable renal function in patients with Fabry disease and amenable GLA mutations over 24-30 months.

2174M

Development of a cell-based reporter assay suited for small-molecule drug discovery in FGF23-inducible HEK293 cells stably expressing Klotho. *S. Diener[1], K. Schorpp[2], B. Lorenz-Depiereux[1], K. Hadian[2], T.M. Strom[1,3].* 1) Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Bavaria, Germany; 2) Institute of Molecular Toxicology and Pharmacology, Assay Development and Screening Platform, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Bavaria, Germany; 3) Institute of Human Genetics, Klinikum Rechts der Isar der Technischen Universität München, Munich, Bavaria, Germany.

Fibroblast growth factor 23 (FGF23) is a key regulator of phosphate homeostasis. It is of crucial importance in hereditary and acquired hypo- and hyperphosphatemic disorders. Moreover, FGF23 has emerged as a promising biomarker for the prediction of adverse clinical outcomes in patients with chronic kidney disease (CKD), as it might be related to mortality, cardiovascular abnormalities and disease progression. FGF23 is a bone-derived endocrine factor, which inhibits renal tubular phosphate reabsorption by activating receptor complexes composed of FGF receptor (FGFR) 1c and the coreceptor Klotho. As a major signalling pathway mitogen-activated protein kinase (MAPK) pathway is employed. For the investigation of FGF23 in an *in vitro* model we established FGF23-inducible HEK293 cells that stably express Klotho (HEK293-KL). The induction of HEK293-KL cells by FGF23 was shown by detecting the activation of MAPK pathway, which could be reduced by the use of two known small-molecule inhibitors of MAPK pathway: SU5402 and U0126. To identify novel small-molecule compounds that modulate FGF23/FGFR1c/Klotho signalling, we have developed a cell-based reporter assay that is suited for high-throughput screening (HTS). The assay is based on the AlphaScreen SureFire platform of Perkin Elmer to monitor the phosphorylation of endogenous extracellular signal-regulated kinase 1 and 2 (ERK1/2) in cellular lysates of HEK293-KL after the induction with FGF23 in the presence of small-molecule compounds. Since increased plasma concentrations of FGF23 are the main cause of many phosphatemic disorders, a modulation of its effect could be a potential strategy for drug discovery and new therapeutic approaches in disorders affecting phosphate homeostasis.

2175S

Phenylbutyrate increases pyruvate dehydrogenase complex activity in cells harboring a variety of defects. *R. Ferriero[1], A. Boutron[2], M. Brivet[2], D. Kerr[3], E. Morava[4], R.J. Rodenburg[5], L. Bonafé[6], M.R. Baumgartner[7], Y. Anikster[8], N.E. Braverman[9,10], N. Brunetti-Pierri[1,11].* 1) Telethon Institute of Genetics and Medicine, Naples, Italy; 2) Laboratoire de Biochimie, AP-HP Hôpital de Bicêtre, Le Kremlin Bicêtre, France; 3) Center for Inherited Disorders of Energy Metabolism, Case Western Reserve University, Cleveland, USA; 4) Hayward Genetics Center, Tulane University Medical School, New Orleans, USA; 5) Nijmegen Center for Mitochondrial Disorders, Department of Pediatrics, Radboud University Medical Center, Nijmegen, The Netherlands; 6) Center for Molecular Diseases, Lausanne University Hospital, Lausanne, Switzerland; 7) Division for Metabolic Diseases and Children's Research Center, University Children's Hospital, Zürich, Switzerland; 8) Metabolic Disease Unit, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel-Hashomer and Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel; 9) Department of Human Genetics, McGill University, Quebec, Canada; 10) Department of Pediatrics, Montreal Children's Hospital Montreal, Quebec, Canada; 11) Department of Translational Medicine, Federico II University of Naples, Italy.

Deficiency of pyruvate dehydrogenase complex (PDHC) is the most common genetic disorder leading to lactic acidosis. PDHC deficiency is genetically heterogenous and most patients have defects in the X-linked E1-α gene but defects in the other components of the complex encoded by *PDHB*, *PDHX*, *DLAT*, *DLD* genes or in the regulatory enzyme encoded by *PDP1* have also been found. Phenylbutyrate enhances PDHC enzymatic activity in vitro and in vivo by increasing the proportion of unphosphorylated enzyme through inhibition of pyruvate dehydrogenase kinases and thus, has potential for therapy of patients with PDHC deficiency. In the present study, we investigated response to phenylbutyrate of multiple cell lines harboring all known gene defects resulting in PDHC deficiency. Drug responses were correlated with genotypes and protein levels by Western blotting. Large deletions affecting *PDHA1* that result in lack of detectable protein were unresponsive to phenylbutyrate whereas increased PDHC activity was detected in most fibroblasts harboring *PDHA1* missense mutations. Mutations affecting the R349-α residue were directed to proteasome degradation and were consistently unresponsive to short-time drug incubation but longer incubation resulted in increased levels of enzyme activity and protein that may be due to an additional effect of phenylbutyrate as a molecular chaperone. PDHC enzyme activity was enhanced by phenylbutyrate in cells harboring missense mutations in *PDHB*, *PDHX*, *DLAT*, *DLD*, and *PDP1* genes. In the prospect of a clinical trial, the results of this study may allow prediction of in vivo response in patients with PDHC deficiency harboring a wide spectrum of molecular defects.

2176M

Rapamycin and Chloroquine: the in vitro and in vivo effects of autophagy-modifying drugs show unexpected results in valosin containing protein multisystem proteinopathy. *A. Nalbandian, K. Llewellyn, N. Walker, C. Nguyen, V. Katheria, A. Gomez, A. Dunnigan, P. Yazdi, V. Kimonis.* University of California, Irvine, Irvine, CA, Department of Pediatrics.

Mutations in the valosin containing protein (VCP) gene cause hereditary Inclusion body myopathy (hIBM) associated with Paget disease of bone (PDB), frontotemporal dementia (FTD), more recently termed multisystem proteinopathy (MSP). Affected individuals exhibit scapular winging and die from progressive muscle weakness, and cardiac and respiratory failure, typically in their 40s to 50s. Histologically, patients show the presence of rimmed vacuoles and TAR DNA-binding protein 43 (TDP-43)-positive large ubiquitinated inclusion bodies in the muscles. We have generated a VCPR155H/+; knock-in mouse model which recapitulates the disease phenotype and impaired autophagy typically observed in patients with VCP disease. Autophagy-modifying agents such as rapamycin and chloroquine at pharmacological doses have previously shown to alter the autophagic flux and autophagic signaling intermediates. Herein, we report results of administration of rapamycin, a specific inhibitor of the mechanistic target of rapamycin (mTOR) signaling pathway, in 19-month old VCPR155H/+ mice. Rapamycin-treated mice demonstrated significant improvement in muscle performance, quadriceps histological analysis, and rescue of defective autophagy by decreasing the protein expression levels of LC3-I/II, p62/SQSTM1, optineurin, ubiquitin, and TDP-43, and inhibiting the mTORC1 pathway substrates. Conversely, chloroquine-treated 19-month old VCPR155H/+ mice revealed progressive muscle weakness, cytoplasmic accumulation of TDP-43, ubiquitin-positive inclusion bodies and increased LC3-I/II, p62/SQSTM1, and optineurin expression levels. These findings suggest that targeting the mTOR pathway and restoring autophagic flux ameliorate an increasing list of disorders and can now include in VCP disease and related neurodegenerative multisystem proteinopathies.

2177S

Unfolded protein response induced by mutant alpha1-antitrypsin (AAT) activates JNK-MAPK pathway that modulates AAT levels and toxicity in AAT deficiency. *N. Pastore[1], B. Granese[1,2], C. Mueller[3], J. Teckman[4], N. Brunetti-Pierri[1,2].* 1) Telethon Institute of Genetics and Medicine, Italy; 2) Department of Translational Medicine, Naples, Italy; 3) Department of Pediatrics and Gene Therapy Center, UMass Medical School, Worcester, MA, USA; 4) Department of Pediatrics, Saint Louis University School of Medicine, Cardinal Glennon Children's Medical Center, Saint Louis, MO, USA.

Alpha-1-antitrypsin (AAT) deficiency is the most common genetic cause of liver disease in children and liver transplantation is the only available treatment in severe patients. The majority of patients carry a missense mutation (lysine for glutamate at amino acid position 342) in the AAT gene that alters protein folding. Mutant AAT is prone to polymerize and aggregate in the endoplasmic reticulum (ER) of hepatocytes, causing liver injury by a gain-of-toxic mechanism. Accumulation of these polymers results in release of ER calcium, NFκB activation, and apoptosis. The involvement of unfolded protein response (UPR) in AAT deficiency has been controversial. We observed that the PiZ mouse, a transgenic mouse model that expresses the mutant human AAT gene and recapitulate the features of human liver disease, has an age-dependent accumulation of mutant AAT that decreases with aging. UPR markers ATF6, PERK, IRE1, BiP, and GRP94 were all found to be overexpressed in younger PiZ mice with more abundant load of mutant AAT. Next, we injected PiZ mice with a recombinant serotype 8 adeno-associated virus (rAAV) vector that incorporate microRNA (miRNA) sequences targeting the AAT gene and resulting in mutant AAT knockdown in the liver. These mice displayed reduced activation of UPR as shown by decreased ATF6 and BiP protein levels thus confirming that UPR activation is dependent upon mutant AAT protein dosage at a given age. Finally, we observed that UPR induced by mutant AAT results in the activation of JNK-MAPK pathway in both PiZ mice and human liver samples from affected patients. Interestingly, inhibitors of this pathway decrease mutant AAT levels in a cell model of the disease, suggesting that these components might be new targets for mitigating AAT deficiency. Collectively, these data reveal new therapeutic entry points for treatment of AAT deficiency.

2178M

Case report: Efficacy of L-Citrulline supplementation in a patient with Mitochondrial Encephalopathy, Lactic Acidosis and Stroke-like episodes (MELAS) with behavioral disturbances and failure to thrive. *E. Alkhunaizi[1], A. Shuen[1], M. Lefrancois[2], M. Larose[3], C. Saint-Martin[4], W. Al-Hertani[1].* 1) Department of Medical Genetics, McGill University Health Centre, Montreal, QC, Canada; 2) Department of Nutrition, McGill University Health Centre, Montreal, QC, Canada; 3) Department of Psychiatry, McGill University Health Centre, Montreal, QC, Canada; 4) Department of Diagnostic Imaging, McGill University Health Centre, Montreal, QC, Canada.

MELAS (MELAS [MIM 540000]) is a mitochondrial multisystem condition that often manifests in childhood with seizures associated with stroke-like episodes and transient hemiparesis or cortical blindness and lactic acidosis. Therapeutic options for MELAS are limited at this time, and under trial. We report on an 8 year-old female of Sri Lankan origin, who presented with status epilepticus and found to have severe lactic acidosis (5-8.2 mmol/L). She subsequently developed mild to moderate bilateral sensorineural hearing loss. Her past medical history was significant for a seizure disorder, Wolff-Parkinson-White syndrome, failure to thrive and behavioral disturbances. Magnetic resonance imaging with spectroscopy (MRI/MRS) showed progressive increased T2 and FLAIR signal in multiple cortical and subcortical regions and presence of lactate peaks. She was later confirmed to have the most common MELAS mutation (m.3243A>G) in the *MT-TL1* (MTTL1 [MIM 590050]) gene. After her delayed diagnosis at 9 years of age, we started her on arginine supplementation (0.3mg/kg/day) for approximately one year. Her weight remained unchanged on the 0.1-3rd percentile and height on the 3rd percentile in that time period, and she continued to deteriorate in her behavior and mood as noted on neuropsychiatric assessment. One year later, as a trial we switched her to citrulline supplementation (0.3mg/kg/day) and a remarkable improvement in her neuropsychiatric function, especially mood and behavior, was noted by her parents. Furthermore, admissions for seizures and stroke-like episodes were less frequent. Finally her growth parameters increased up to the 15th-50th percentile for weight and 15th percentile for height, over a period of seven months, after initiation of citrulline. It is known that citrulline increases nitric oxide (NO) production more than arginine, and as published by El Hattab et. al., citrulline may potentially have a better therapeutic effect in MELAS, compared to arginine, with larger trials currently underway. This case report illustrates impressive improvement in clinical outcome and quality of life with citrulline supplementation in our MELAS patient.

2179S

Combination Therapy To Enhance Antisense Mediated Exon Skipping for Duchenne Muscular Dystrophy. *D.W. Wang[1], E.I. Mokhonova[1, 2, 3], L. Martinez[2, 3], D. Becerra[2, 3], M.J. Spencer[2, 3], S.F. Nelson[3, 4, 5, 6], M.C. Micheli[1, 3, 5].* 1) Microbiology Immunology and Molecular Genetics, University of California, Los Angeles, Los Angeles, CA., USA; 2) Neurology, David Geffen School of Medicine, University of California, Los Angeles; 3) Center for Duchenne Muscular Dystrophy, University of California Los Angeles, California, USA; 4) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, California, USA; 5) Molecular Biology Institute, University of California Los Angeles, California, USA; 6) Department of Pathology and Laboratory Medicine, University of California Los Angeles, California, USA.

Duchenne muscular dystrophy (DMD) is a lethal X-linked recessive disorder primarily caused by frameshifting deletions in the DMD gene that result in a loss of dystrophin protein expression. Antisense oligonucleotide-mediated exon skipping shows promise as a therapeutic approach by restoring the DMD reading frame and correcting dystrophin protein localization to the DGC. Using a small-molecule screen, our lab has previously found that dantrolene, an FDA approved compound used to treat malignant hyperthermia, increases antisense oligonucleotide-mediated exon skipping. Previous work has evaluated dantrolene synergy with antisense oligonucleotide (AO) in the context of twice daily intraperitonealy administered dantrolene. We tested if dantrolene administered via oral gavage or incorporated into mouse chow had a similar efficacy enhancement of Dmd exon 23 antisense oligonucleotide (PMOE23) skipping and dystrophin protein rescue. RNA was analyzed via qRT-PCR for amounts of skipped and full-length Dmd mRNA. Quantitative immunohistochemistry was performed to assess both rescued dystrophin levels as compared to wildtype control, as well as proper localization at the sarcolemma. We have thus assessed a series of alternate and more viable methods for dantrolene delivery with a large number of replicate 3 week experiments, and determined that dantrolene mixed in chow delivers optimal, stable and effective dosing. We further demonstrate that dantrolene administered with one half of the maximally effective dose of AO (greater than 100mg/kg) induces an equivalent amount of dystrophin as 100mg/kg per week AO or 300mg/kg per week AO. Thus, we have now determined route and method of dantrolene delivery that is compatible with long-term administration, and have established dosing of AO to assess both potential for enhancement of maximal effect and potential for dose saving effect in the upcoming 6 month treatment experiment. In addition to the animal studies, we aim to establish that dantrolene enhancement of exon skipping of human exon 51 occurs in multiple independent human induced myotubes. For this work, we have established protocols for skin punch fibroblast growth and transdifferentiation into myotubes. Seven skin punch biopsies have been obtained from patients with exon 51 skippable mutations, and 5 have been reprogrammed into iDRMs. We are assessing AVI4657 mediated skipping with and without dantrolene in myotubes generated from these iDRM cell lines.

2180M

Inhibition of AKT signaling in Proteus and PROS cells: A simple model for cancer therapeutics targeting the AKT/PI3K pathway. *L.G. Biesecker[1], M.J. Lindhurst[1], M.R. Yourick[1], D.T. Dransfield[2].* 1) NHGRI, NIH, Bethesda, MD; 2) ArQule, Inc., Woburn, MA.

The AKT/PI3K (phosphatidylinositol-3-kinase) signaling pathway is critical for cellular growth, survival, protein synthesis, and glucose metabolism. Mutations in several key genes that result in activation of this pathway have been implicated in disorders characterized by overgrowth and/or hypoglycemia and have been identified in numerous tumor samples. Identification of inhibitors that ameliorate the affects of these mutations is an important first step in the development of protocols to treat not only cancer, but also disorders such as Proteus syndrome (PS) and PROS (PIK3CA-related overgrowth spectrum). We tested a small molecule imidzopyridine pan-AKT inhibitor in skin fibroblasts harboring the *AKT1* p.Glu17Lys mutation isolated from several patients with PS. Cells treated for 24 hours with 31-500 nM with this compound grown in both serum-free and serum-containing medium had reduced levels of AKT phosphorylation, with levels at the higher three doses approaching those of quiescent wild-type cells. The inactivation of AKT, as measured by AKT phosphorylation, occurred rapidly as the levels were reduced 20-50 fold within two hours of addition of 125 nM of this inhibitor to the media. Similar results were found using fibroblasts containing *PIK3CA* p.His147Arg or p.His1047Leu mutations. *AKT1* mutation-positive cells treated with 625 nM of the inhibitor for 72 hours had a 25-40% reduction in cell viability compared to untreated cells, whereas mutation-negative cells only had a 0-20% decrease. These results indicate that basal AKT signaling in cells containing activating mutations in *AKT1* and *PIK3CA* can be inhibited in PS and PROS cells using an imidzopyridine inhibitor. Furthermore, our results suggest that reduced signaling can be achieved without complete disruption of cellular processes. This work has implications beyond PS and PROS as these mutations have been shown to be driver mutations in tumors. Because these cells harbor only a single mutation, as compared to the hundreds or thousands of mutations in tumors, they can be considered as single gene models of tumor therapy. Thus, the evaluation of such cells from patients with only a single mutation will simplify studies and allow investigators to assess the efficacy of a compound on a specific pathway without confounding effects resulting from additional driver and passenger mutations typically found in tumor cells.

2181S

Cycloheximide enhances skipping of mutated DMD exons synergistically with TG003. *A. Nishida[1], Y. Takeshima[2], T. Lee[3], T. Takarada[4], M. Matsuo[1].* 1) Medical Rehabilitation, Kobegakuin University, Kobe, Japan; 2) Department of Pediatrics, Hyogo Medical College, Nishinomiya, Japan; 3) Department of Pediatrics, Kobe University Graduate School of Medicine, Kobe, Japan; 4) Department of Clinical Pharmacology, Kobe Pharmaceutical University, Kobe, Japan.

Duchenne muscular dystrophy (DMD) is a progressive and fatal muscle wasting disease. DMD is characterized by muscle dystrophin deficiency caused by mutations in the DMD gene. Exon skipping that produces in-frame functional DMD mRNA is the most plausible way to express dystrophin in DMD therapy. In our previous report, we showed that TG003, a chemical Clk1 inhibitor, enhances skipping of mutated DMD exons and the subsequent expression of dystrophin (Nishida et al. Nat Commun 2011). It has been questioned whether TG003-mediated exon skipping is potentiated. Here, we searched for chemicals that enhance skipping of 14 mutated DMD exons. Hybrid-minigenes inserting one of mutated exons were transfected into HeLa cells and resultant mini-gene product was analyzed RT-PCR amplification. We found that cycloheximide (CHX), a protein synthesis inhibitor, enhanced skipping of DMD exon 31 encoding a nonsense mutation (c.4303G>T). Other protein synthesis inhibitors, anisomycin and puromycin did not enhance skipping, but emetine did, indicating inhibitor-specific enhancement. As observed in TG003 treatment, CHX enhanced the skipping of mutated exons 27 and 39. CHX showed no effect on splicing of further 11 mutated or 14 wild-type DMD exons. CHX-mediated exon skipping was observed in the limited mutated exons that have been shown responsive to TG003 for exon skipping. This suggested that both TG003 and CHX enhance skipping of the mutated exons through common mechanism. Unexpectedly, the level of exon 31-skipping with TG003 was increased higher by combined treatment with CHX than the sum of increased levels with single chemical treatment. CHX was found to enhance the skipping of mutated exon 31 synergistically with TG003. This suggested that TG003 and CHX enhanced mutated exon 31 skipping in their own pathway. TG003 has been shown to increase full-length mRNA of the Clk1 gene which product phosphorylates SR protein. According TG003 treatment increased full-length Clk1 mRNA in HeLa cells. Remarkably, combination treatment of TG003 and CHX shifted the splicing completely to produce full-length mRNA only. This indicated that CHX modulates splicing of the Clk1 gene. CHX was shown to enhance skipping of mutated DMD exon 31 synergistically with TG003, and nominated as a leading compound to potentiate TG003-mediated exon skipping.

2182M

Treatment of *LMNA* Laminopathies with the Rapamycin Analog RAD001. *A.J. DuBose, N.M. Petrash, M.R. Erdos, F.S. Collins.* NHGRI, NIH, Bethesda, MD.

The lamina lies beneath the inner nuclear membrane (INM) and is composed of a meshwork of intermediate-sized filament proteins called lamins. The lamina is crucial for the structural support of the nucleus and interacts with INM proteins and chromatin. Two types of lamins make up the lamina - type A lamins and type B lamins. Through alternative splicing, *LMNA* encodes the four type A lamins, the major forms of which are lamin A and lamin C. Mutations in *LMNA* cause a wide array of disorders (laminopathies) including lipodystrophies, skeletal and cardiac myopathies, peripheral neuropathies, and accelerated aging syndromes.

Hutchinson-Gilford Progeria Syndrome (HGPS) is a rare premature aging disorder usually caused by a single nucleotide mutation in *LMNA*. This mutation activates a cryptic splice donor site causing abnormal mRNA splicing. As a result, an aberrant lamin A isoform called progerin is created. Progerin causes deleterious downstream effects including increased DNA damage, misregulation of gene expression, chromatin disorganization, altered histone modification, disruption of nuclear envelope proteins, replicative senescence, and blebbing (bulging of nucleus).

Rapamycin inhibits the mTOR pathway and thus triggers autophagy - a process that clears unnecessary proteins and organelles from cells. Treatment of HGPS fibroblasts with Rapamycin or RAD001 (Everolimus) has been previously shown to increase progerin clearance, delay cellular senescence, and reduce nuclear blebbing.

The purpose of this study is to expand investigation of RAD001 to other types of laminopathies. In addition to progerin, other mutant forms of Lamin A are known to cause defects in lamina function and could be removed by autophagy. We are examining the effects of RAD001 treatment on cells from patients with Hutchinson-Gilford Progeria Syndrome (HGPS), Atypical HGPS, Atypical Werner's Syndrome, and Emery-Dreifuss Muscular Dystrophy. We found that RAD001 increases autophagy, and some of the laminopathy cell lines show a decrease in nuclear blebbing. We hypothesize that RAD001 treatment will increase cell proliferation and rescue premature senescence.

2183S

Novel Cystine Derivatives as Potential Inhibitors of Cystine Stone Formation in Cystinuria. *A. Sahota[1], M. Yang[1], Y. Yanghui[2], H. Aloysius[2], H.M. Albanyan[2], L. Hu[2], J.A. Tischfield[1].* 1) Dept Gen, Rutgers Univ, Piscataway, NJ; 2) Dept Medicinal Chem, Ernest Mario Sch Pharmacy, Rutgers Univ, Piscataway, NJ.

Background: Cystinuria, caused by mutations in SLC3A1 or SLC7A9, is a common cause of urolithiasis in children. Slc3a1 knockout mice develop cystinuria and cystine stones in the bladder. Here we report the effects of two novel cystine derivatives (LH707 and LH708) on cystine solubility in vitro and in Slc3a1 knockout mice. The results were compared with cystine dimethyl ester (CDME), which is known to inhibit cystine crystal growth in vitro and cystine stone formation in knockout mice. **Methods:** The novel derivatives were designed to be absorbed through the intestine and are more resistant to esterases compared with CDME. The concentration of each derivative to double the solubility of cystine (EC2x) in aqueous solution was determined. The inhibitors were then administered by stomach tube to 2-3-month-old Slc3a1 knockout male mice daily for four weeks (200 μl of 3 mM solution in water). The bladders were dissected and the stones counted and weighed. In a second study, mice were prescreened for bladder stones by micro computed tomography and those with stones treated with LH708 and urine samples collected before and after treatment. Mutant and wild-type mice were treated with LH707 and LH708 for one week and the concentration of these compounds in the urine determined. **Results:** The EC2X for LH707, LH708 and CDME was 0.261, 0.856, and 6.37 μM, respectively. Thus, the new derivatives are 7.4 and 24.4 times more effective than CDME in the in vitro assay. The majority of stones in the LH708-treated group were less than 1 mm in the longest dimension, whereas stones in the LH707-treated group were distributed over a wider range. Only the concentration of cystine was increased in urine following LH708 treatment ($p = 0.033$). The concentration of LH708 in urine was higher than that of LH707 in wild-type or mutant mice and urine output was also greater in the LH708-treated mice (wild-type or mutant). **Conclusions:** Both LH707 and LH708 were detected in urine following treatment, indicating that these compounds were absorbed. LH708 showed better efficacy in inhibiting stone formation than LH707. The increase in cystine concentration in urine after treatment with LH708 supports our in vitro findings. The increase in concentration of LH708 in urine from mutant mice and the increase in urine output from these mice suggest that there may be enhanced transport of LH708 in mutant mice and that LH708 may have a diuretic effect in addition to inhibiting stone formation.

2184M

Towards treatment of Cantu syndrome. *G. van Haaften.* Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Utrecht, Netherlands.

The CantuTreat project aims to develop a novel therapeutic approach in treatment of Cantú syndrome. Cantú syndrome is a rare genetic disorder, affecting a small number of patients (currently 100 cases are known worldwide) who suffer from multiple symptoms including hypertrichosis, lymphedema, distinctive facial features and cardiac abnormalities. Cantú syndrome is caused by dominant gain-of-function mutations in the ATP-dependent potassium channel subunits ABCC9, and KCNJ8. This recent discovery offers a promising opportunity for therapeutic intervention. The IKATP potassium channel is a known pharmaceutical target: pharmaceutical correction of these channels by sulfonylurea drugs results in a nearly complete cure of patients with neonatal diabetes. In analogy, we intend to develop a sulfonylurea treatment for Cantú syndrome, combining in silico, in vitro and in vivo approaches. This novel treatment on the basis of existing pharmacology and approved drugs might prevent invasive clinical procedures such as heart surgery and frequent laser treatment, thus significantly improving the quality of life and life expectancy of the patients. CantuTreat brings together an international, multidisciplinary consortium for translational research into the development of a cost-effective treatment. The project unites the leading researchers and clinical experts in the field. Modelling and experimental testing of specific Cantú related mutations will further our understanding of the biology of the various symptoms. In addition we will build a global patient registry to generate a critical mass of patients in order to ensure a rapid progression towards clinical application when this preclinical study has finished. The translational approach towards a novel therapeutic strategy in this study could also serve as a blueprint for tackling related channelopathies.

2185S

Early intravenous enzyme replacement therapy improves white matter myelination in canine mucopolysaccharidosis I. *P. Dickson[1], J.M. Provenzale[2,3], S. Chen[2], I. Nestrasil[4], J. Yee[1], S.h. Kan[1], S.Q. Le[1], J. Jens[5], E. Snella[5], M.A. Guzman[6], C. Vite[7], E. Shapiro[4], N.M. Ellinwood[5].* 1) LA BioMed/Harbor-UCLA, Torrance, CA; 2) Duke University, Durham, NC; 3) Emory University, Atlanta, GA; 4) University of Minnesota, Minneapolis, MN; 5) Iowa State University, Ames, IA; 6) St. Louis University, St. Louis, MO; 7) University of Pennsylvania, Philadelphia, PA.

Diffusion tensor imaging studies in people with mucopolysaccharidosis I (MPS I) show reduced volume and reduced fractional anisotropy in the corpus callosum, which correlates with inattention on neurobehavioral tests. We have found decreased myelination in MPS I dogs as early as 6 weeks of age, and the fractional anisotropy of the corpus callosum varies with the amount of myelin basic protein in myelin extracts. We therefore studied the impact of enzyme replacement therapy (ERT) with recombinant human alpha-L-iduronidase on myelination. MPS I dogs received intrathecal (IT) ERT 0.05 mg/kg every 3 months and/or intravenous (IV) ERT 0.58-2.0 mg/kg weekly. Untreated MPS I and normal carriers were used as controls. The expression of several myelin-related genes was reduced in the corpus callosum of MPS I dogs compared to normal carriers. Expression of myelin-related genes improved in MPS I dogs treated with IT and/or IV ERT beginning at ≤30 days of age: the expression of myelin-related genes was normal in dogs treated with 0.58 mg/kg weekly IV ERT + IT ERT, and was greatly improved in dogs receiving 1.57 or 2.0 mg/kg weekly IV ERT without IT ERT. There was no improvement in myelin gene expression in dogs treated with IT ERT beginning at age 4m, although there was some improvement in fractional anisotropy, myelin basic protein levels, and myelin lipids. We have previously published improvement in brain lysosomal storage in MPS I dogs receiving early IV ERT. We suggest that small quantities of IV ERT cross the blood-brain barrier and that is sufficient to prevent some of the neuropathology of MPS I.

2186M

The impact of enzyme replacement therapy on immunity in Gaucher disease. *R. Limgala[1,2], C. Loanou[1], E. Komlodi-Pasztor[1], M. Plassmeyer[2], M. Ryherd[2], L. Austin[2], S. Lipinski[1], A. Hebert[1], M. Brown[2], O. Alpan[2], O. Goker-Alpan[1].* 1) Lysosomal Disorders Treatment Unit, O&O Alpan LLC, Fairfax, VA 22030; 2) Amerimmune, O&O Alpan, LLC, Fairfax, VA 22030.

Introduction: Gaucher disease (GD) is an autosomal recessive inherited disorder of metabolism caused by mutations in the GBA gene. Subsequent deficiency of the lysosomal enzyme glyucocerebrosidase results in accumulation of glycosphingolipids in macrophages. Patients with GD often present with an abnormal immune response that may be the result of cellular and/or humoral immune dysregulations. One of the most accepted therapies to treat GD is Enzyme replacement therapy (ERT). Here, we are examining the impact of ERT on immunity of GD patients. Cases and Methods: In an ongoing study (NCT01358188), the impact of ERT on the immunity was assessed in twenty-seven GD patients (19F/8M, mean age 40.2 yrs), five of whom were treatment-naïve at the start of the study. Flow cytometry based immunophenotyping was performed from peripheral blood samples obtained before and after ERT administration. Lymphocyte subsets, memory, NK, B and dendritic cell populations were assayed along with chemokine receptors and activation markers. Results: In a treatment naïve patient at baseline, prior to commencing ERT, CD4/CD8 ratio was less than 1 indicating immunosenescence, and transitional B-cells (characterized by CD21low expression) were markedly elevated suggesting a B-cell maturation defect. After establishment of stable-dose therapy, not only was the CD4/CD8 ratio reversed, but CD21 low cells were normalized, indicating improved B cell maturation. An increase in dendritic cells was also observed. Although the majority of subjects on long-term ERT exhibited normal ranges of lymphocyte subsets, in less than 20%, all of whom were diagnosed before the advent of ERT, the CD8 T-cell fraction was either elevated or expressed increased chemokine receptors like CXCR3, CCR6. Conclusions: Patients with GD can present with multiple immune abnormalities beyond the scope of activated macrophages resulting from accumulated sphingolipids. Our data suggest ERT or its effects may improve immunological parameters indicating the role of ERT on long-term outcome may extend beyond the known effects of decreasing organ size and improvement of hematological parameters.

2187S

An open-label, multicenter, ascending dose study of the tolerability and safety of recombinant human acid sphingomyelinase (rhASM) in patients with ASM deficiency (ASMD). *M.P. Wasserstein[1], S.A. Jones[2], H. Soran[2], G. Diaz[1], B. Thurberg[3], K. Culm-Merdek[3], A. Cunningham[3], T. Singh[3], A.C. Puga[3].* 1) Icahn School of Medicine at Mount Sinai, New York, NY, USA; 2) Manchester Centre for Genomic Medicine, St. Mary's Hospital, CMFT, University of Manchester, Manchester, UK; 3) Genzyme, a Sanofi company, Cambridge, MA, USA.

Background: Enzyme replacement therapy (ERT) with recombinant human acid sphingomyelinase (rhASM) is in clinical development for the treatment of the non-neurological manifestations of acid sphingomyelinase deficiency (ASMD). In a phase 1a single-ascending-dose study in adult patients, the maximum tolerated starting dose of rhASM was determined to be 0.6 mg/kg. Objectives: The primary objective of this phase 1b study was to determine the safety and tolerability of within-patient dose escalation of rhASM in five adult patients with Niemann-Pick Disease type B (NPD B), the non-neuronopathic form of ASMD. Each patient was to receive a starting dose of intravenous rhASM at 0.1 mg/kg and advance every two weeks according to a predetermined schedule to 3.0 mg/kg, or their maximum tolerated dose. The secondary objective was to study the pharmacokinetics, pharmacodynamics, and exploratory efficacy of rhASM administered intravenously every two weeks for 26 weeks. Methods: Study assessments included continuous adverse event (AE) reporting and periodic evaluations of safety, pharmacokinetics, pharmacodynamics, and exploratory efficacy. Safety biomarkers (e.g., C-reactive protein, bilirubin, IL-6, IL-8) were evaluated; plasma ceramide was used as a biomarker for the breakdown of sphingomyelin. Sphingomyelin content in liver was used as a pharmacodynamic endpoint. Results: The dose escalation regimen was well tolerated, with all patients reaching the maximum dose of 3.0 mg/kg. No serious or severe adverse events or deaths were reported. Related AEs consisted predominantly of infusion-associated reactions, the majority of which were mild and resolved without sequalae. A positive response to treatment with rhASM was observed in liver sphingomyelin content and several exploratory efficacy parameters, including spleen and liver volumes, pulmonary function testing, lung imaging, lipid profile, and quality of life assessments. Results will be presented. Conclusions: Within-patient dose escalation of rhASM was well tolerated. Repeat-dose safety and exploratory efficacy of rhASM support its continued development for the treatment of the non-neurological manifestations of ASMD. This study was funded by Genzyme, a Sanofi company.

2188M

A proprietary human acid α-glucosidase with high mannose 6-phosphate and its application in chaperone-advanced replacement therapy (CHART ™) as a potential next-generation treatment for Pompe disease. *S. Xu, R. Gotschall, M. Frascella, A. Garcia, R. Soska, Y. Lun, J. Feng, K. Chang, R. Khanna, H. Do, K. Valenzano.* Amicus Therapeutics, Cranbury, NJ.

Pompe disease is an inherited lysosomal storage disease that results from deficiency in acid α-glucosidase (GAA) activity. It is characterized by progressive accumulation of lysosomal glycogen in cardiac and skeletal muscles. The only treatment currently available is enzyme replacement therapy (ERT) using recombinant human GAA (rhGAA). The targeting and uptake of the exogenous enzyme into lysosomes requires the specialized carbohydrate, mannose-6-phosphate (M6P), which binds to cation-independent M6P receptors (CIMPR) at the cell surface for internalization. However, the quality and quantity of M6P present on existing rhGAA is not optimal, and may limit lysosomal uptake and clinical efficacy. While existing rhGAA products at 20 mg/kg or higher doses do achieve some clinical efficacy, unmet needs due to inadequate uptake and substrate reduction in key tissues still exist. Also the risk for developing immune response to rhGAA is high. Other drawbacks include long infusion times for patients and high demand for the manufacture of large quantities of rhGAA. We have developed a proprietary mammalian cell line and purification process that yields an rhGAA enzyme with significantly higher M6P content compared to existing products. This new rhGAA (designated as AT-B200) binds the intended CIMPR with high affinity in vitro, and is more efficiently internalized by skeletal muscle myoblasts. In the present study, AT-B200 was intravenously administered to *Gaa* KO mice using bi-weekly regimens, and achieved up to 75% reduction of accumulated glycogen in multiple skeletal muscles. Importantly, lower doses of AT-B200 (5 and 10 mg/kg) compared to Lumizyme® (20 mg/kg) showed similar or greater glycogen reduction in key skeletal muscles. At 20 mg/kg, AT-B200 showed significant superiority to Lumizyme, indicating that the higher M6P content in AT-B200 enables better drug targeting and substrate reduction in vivo. Moreover, co-administration of a pharmacological chaperone which was previously shown to improve the pharmacological properties of Pompe ERT via binding and stabilization resulted in a trend of further glycogen reduction in key skeletal muscles. Taken together, these data demonstrate that a well-targeted rhGAA mono-therapy leads to greater substrate reduction in KO mice, with additive benefits from the application of the CHART platform, thus warranting further investigation of this next-generation treatment for Pompe disease.

2189S

A Phase II Multicenter, Open-label Trial to Evaluate the Safety and Efficacy of Fabagal® (Agalsidase beta) in Patients with Fabry Disease. *H. Yoo[1], W.S. Kim[2], C.H. Lee[3], D.S. Kim[4], J.K. Jeon[5], H. Kook[6].* 1) Dept Pediatrics, Asan Medical Ctr, Seoul, KoreaKorea; 2) Dept Internal Med, Kyunghee Univ Hosp, Seoul, Korea; 3) Dept Internal Med, Hanyang Univ Hosp, Seoul, Korea; 4) Dept Neurology, Pusan Natl Univ Hosp, Pusan, Korea; 5) Dept Pediatrics, Pusan Natl Univ Hosp, Pusan, Korea; 6) Dept Pediatrics, Cheonnam Natl Univ Hosp, Hwasoon, Korea.

Background: Fabry disease is caused by a deficiency of α-galactosidase A (α-gal A), resulting in the accumulation of glycosphingolipids, mainly globotriaosylceramide (GL-3) in the vascular endothelial cells or tissues throughout the body. Fabagal® is a recombinant human α-gal A (agalsidase beta) produced from CHO cells through serum-free suspension culture. This study was planned to prove whether Fabagal® could be used as an enzyme replacement therapy for Fabry disease. We evaluated the efficacy and safety of enzyme replacement therapy (ERT) with Fabagal® in patients with Fabry disease. Methods: The study was a multicenter, open-label, and phase II study in Fabry disease patients for 24 weeks. A total of ten patients (7 male and 3 symptomatic female patients including 1 naiive patient) were enrolled (n = 10). The objective was to evaluate safety and efficacy during Fabagal® administration (1.0 mg/kg every 2 weeks) for 24 weeks. The primary endpoint of efficacy was to maintain the level of plasma GL-3 within normal range. The secondary endpoint variables were changes of urine GL-3, serum creatinine (Cr), plasma lysoGL-3 and urine lysoGL-3, echocardiographic findings, the score of Short Form McGill Pain Questionnaire & SF-36 Health Status Survey, and Subject symptom diary at 24 weeks of treatment compared to the baseline. Results: Nine of 10 enrolled patients completed the study (n = 9). Treatment with Fabagal® proved efficacious since plasma GL-3 level maintained within the normal range. Also urine GL-3 value was normalized after ERT with Fabagal® compared to the baseline. There were no significant changes of plasma lysoGL-3, urine lysoGL-3 (remained within normal range), echocardiographic findings and serum Cr level during the study period. As a result of the adverse event (AE) analysis, there was no serious adverse event or any drop out or death due to AEs. All AEs were resolved without any complications and none were related to Fabagal®. There was no significant infusion-associated reaction. Conclusions:This study proved the beneficial effect and safety of Fabagal® (agalsidase beta) that generates clinically significant decrease of GL-3 with no significant adverse drug reaction in Fabry disease patients. However, further study is needed in larger number of naiive patients for long-term period.

2190M

Choroid plexus-directed viral gene therapy for alpha-mannosidosis, a prototypical lysosomal storage disease. *E. Choi, S.G. Kaler.* Section on Translational Neuroscience; Molecular Medicine Program, NICHD/NIH, Bethesda, MD.

The choroid plexuses are highly vascularized structures that project into the cerebrospinal fluid (CSF) of the four cerebral ventricles. The specialized polarized epithelia of choroid plexuses produce CSF by transporting water and ions into the ventricles, and turn over at an extremely slow rate. We hypothesized that remodeling these epithelia to secrete a missing lysosomal enzyme by one-time administration of a recombinant AAV (rAAV) gene therapy vector into the cerebrospinal fluid could be an attractive and efficacious approach for long-term treatment of lysosomal storage diseases (LSD). Lysosomes function as the primary digestive units within cells and specific enzymes within lysosomes normally break down nutrients. Patients with LSDs cannot metabolize certain nutrients, resulting in diminished lifespans and reduced quality of life. There are no ideal therapeutic options presently available for the neurological manifestations of LSDs. Brain-directed recombinant enzyme replacement has shown promise for several LSDs but requires repeated instillations due to short enzyme half-lives. In contrast, rAAV-mediated gene transfer to the choroid plexus would enable continuous synthesis and secretion of missing lysosomal enzymes into the CSF and steady penetration to the cerebral cortex and cerebellum. To evaluate this hypothesis, we obtained a mouse model of alpha-mannosidosis in which a targeted disruption of the lysosomal acid α-mannosidase (LAMAN) gene is present. Using mice from our developing colony, we devised a genotyping assay that distinguishes wild type, heterozygous and homozygous animals. We cloned the human (hu) LAMAN cDNA into an adeno-associated virus serotype 5 (AAV5) shuttle plasmid and documented robust LAMAN expression in transfected HK293T cells. We next generated high titer rAAV5-huLAMAN expressing huLAMAN and administered 1×10e10 or 1×10e11 viral particles to homozygous affected mice by lateral ventricle injection on day 2. We documented efficient, dose-dependent viral transduction in the initial group of treated animals. Biochemical analyses and neurobehavioral testing in treated animals is in progress. We also established a NIH clinical protocol (14-CH-0106) to evaluate biomarkers in CSF of human subjects with alpha-mannosidosis for potential use in a future clinical trial. If the choroid plexus viral gene therapy approach were successful, the largest current barriers to health for patients with certain LSDs would be circumvented.

2191S

Genotoxicity after adeno-associated virus (AAV) gene therapy is dependent upon dose, treatment age and enhancer-promoter selection. *R.J. Chandler, M.C. LaFave, G.K. Varshney, A.G. Elkahloun, S.M. Burgess, C.P. Venditti.* Nat Human Genome Res Inst, National Institutes Hlth, Bethesda, MD.

AAV gene therapy has recently been approved for clinical use and shown to be efficacious and safe in a growing number of clinical trials. However, the safety of AAV as a gene therapy has been challenged by studies that documented hepatocellular carcinoma (HCC) after AAV gene delivery in mice. The association between AAV and HCC has been difficult to reconcile and is the subject of intense debate because numerous AAV studies have not reported toxicity. Here, we report a comprehensive study of HCC in a large number of mice following therapeutic AAV gene delivery. Using a sensitive high-throughput integration site-capture technique and global expressional analysis, we found that AAV integration into the Rian locus and the over-expression of a proximal gene, *Rtl1*, were associated with HCC. In addition, we identify a number of genes with differential expression that may be useful in the study, diagnosis and treatment of HCC. We demonstrate that AAV vector dose, enhancer-promoter selection, and the timing of gene delivery are critical factors in AAV-mediated insertional mutagenesis. Our results help define a mechanism underlying AAV-mediated genotoxicity and have important implications for the design of both safer AAV vectors and gene therapy studies.

2192M

Mutational and functional analysis of Glucose transporter 1 deficiency syndrome. *S. Nakamura[1], H. Osaka[1], S. Muramatsu[2], S. Aoki[1], E. F. Jimbo[1], T. Yamagata[1].* 1) Department of Pediatrics, Jichi Medical University, Shimotsuke-shi, Tochigi, Japan; 2) Department of Neurology, Jichi Medical University, Shimotsuke-shi, Tochigi, Japan.

Glucose transporter 1 deficiency syndrome (Glut-1 DS) is an autosomal dominant disorder caused by mutations in the SLC2A1 gene. The Glut-1 is expressed in the blood-brain barrier and responsible for hexose transport into the brain. The classic phenotypes with Glut-1 DS are infantile drug resistant seizures, intellectual disability (ID) and cerebellar ataxia. Seizures of Glut-1 DS can be treated with a ketogenic diet, but other symptoms persist. We detected a new mutation on SLC2A1 in a classical type patient on whom ketogenic diet was not effective. To aim more effective treatments of Glut-1 DS, we have been investigating the potentiality of gene therapy using AAV vector. As the first step, we established a functional system that monitor the ability of introduced SLC2A1 to transport glucose into cells. (Materials and methods) The patient was a 16-year old boy. He had frequent convulsion from early infancy and had severe ID. Ketogenic diet started from 6 years was not effective. After obtaining informed consent, we isolate DNA from a patient and sequenced the SLC2A1. Total RNA was also extracted and reverse transcribed to cDNA. For transfection study, we used SLC2A1 expression vector by Origene technologies (RC222696). We created plasmids with SLC2A1 mutation by ligation of oligonucleotide corresponding to R333W, A405D and c.906_907insG. To assess glucose transport, we transfected wild type and mutated SLC2A1 vectors into HEK293 cells, and 2-deoxyglucose (2DG) uptake was analyzed 48 hours after gene transfection. (Results) We identified a novel frame shift mutation, c.906_907insG in our patient. 2DG uptake increased in SLC2A1 gene expressing HEK293 cells than control vector transfected cells. 2DG uptake by mutated SLC2A1 vectors transfection to HEK293 cells decreased compared with wild type SLC2A1 vector or control vector transfection. (Discussion) Glut-1 DS is caused by haploinsufficiency of the SLC2A1 and expected to benefit from gene therapy without highly gene expression level. We could achieve the increase of 2DG uptake by wild type SLC2A1 gene transfection to HEK293 cells, which was not observed by the mutated SLC2A1. This system allows us to assess the 2DG uptake by vectors SLC2A1.

2193S

Online DuchenneConnect self-report data indicates Exon 44 skippable DMD patients have prolonged wheelchair-free survival. *J.W. Ulm[1,2], R.T. Wang[1,2], A. Eskin[1,2], S.F. Nelson[1,2,3].* 1) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles; 2) Center for Duchenne Muscular Dystrophy, University of California, Los Angeles; 3) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles.

Duchenne muscular dystrophy (DMD [MIM 310200]) is a debilitating and fatal human genetic disorder, resulting from frameshift mutations at the large *Dystrophin* gene locus on the X chromosome that disrupt the dystroglycan complex (DGC) in muscle tissue. Loss of dystrophin function causes a stereotypical clinical pattern of progressive lower-extremity proximal muscle weakness and depleted muscle mass, commencing usually prior to age 6, with muscle damage provoking inflammation and fibrosis. This leads eventually to wheelchair dependence (generally before adolescence) and eventually respiratory failure before the third decade. One treatment modality of significant interest is exon-skipping, which employs an antisense-mediated approach to bypass exons near disease-causing deletions while yielding a mostly intact transcript, thus restoring partial dystrophin function. DMD patients show significant heterogeneity in the age at which they lose the capacity for ambulation and become wheelchair-dependent (WCage), suggesting wide variation in the phenotypes ensuing from specific deletions and, therefore, in the impacts from skipping at different exons. At present, however, there is limited data available on which "skippable" exons would most likely give rise to an ameliorated phenotype, knowledge that would be of significant value in testing possible exon-skip protocols and in formulating eventual treatment regimens. In the work described here, we used online self-report data from the DuchenneConnect registry alongside gene characterization data from patients, grouping them according to which exon skip (based on their respective deletion mutations) would restore the *Dystrophin* transcript to being in-frame. We then considered the patients' particular skippable exons in the context of the reported severity of their phenotypes, focusing especially on their ages at loss of ambulation. We found that Exon 44 skippable patients demonstrated significant mitigations in severity compared to Exon 45, 50, 51, and 53 skippable patients and those with nonsense mutations, with statistically significant increases in WCage for the Exon 44 skippable group relative to other groupings (median 2 years at $p = 0.009$). These results support other studies of exon skippability and, given the comprehensive data and high statistical power of the DuchenneConnect registry, suggest that online self-report data can be of tangible value in guiding further research and treatment protocols.

2194M

TALEN-mediated Genome Modification Leads to Ablation of Intranuclear Foci in Neural Stem Cells Derived from Human Myotonic Dystrophy Type 1 iPS Cells. *G. Xia[1], Y. GAo[1], S. Jin[2], SH. Subramony[1], N. Terada[3], L. Ranum[2], M.S. Swanson[2], T. Ashizawa[1].* 1) Department of Neurology, University of Florida, College of Medicine, Gainesville, FL, USA; 2) Department of Molecular Genetics and Microbiology, College of Medicine, Gainesville, FL, USA; 3) Department of Pathology, Immunology & Laboratory Medicine, College of Medicine, Gainesville, FL, USA.

Objective: Myotonic dystrophy type 1 (DM1) is caused by expanded CTG repeats in 3'UTR of the DMPK gene. The advancement of iPS cell technology has introduced new possibilities for developing cell-based therapy. Correcting the mutation in DM1 iPS cells would be an important step towards autologous stem cell therapy. In this study, we developed an approach of in vitro gene editing to prevent production of mutant expanded CUG transcripts. Methods: Gene editing was performed in DM1 neural stem cells derived from human DM1 iPS cells. Integration of a cassette containing a selectable marker, SV40 polyA and bGH polyA sequences upstream of the DMPK 3'UTR CTG repeats was mediated by TALEN-induced double-strand break (DSB) and homologous recombination (HR). Cells that contained the insertion cassette were enriched by daGFP expression or puromycin resistance. The expression of mutant CUG repeats was monitored by foci formation using RNA fluorescence in situ hybridization (RNA-FISH). Results: The insertion cassette containing the selectable marker and polyA signals was successfully inserted into DMPK intron 9. The insertion resulted in the complete loss of intranuclear RNA foci. After removal of puromycin selectable marker by Cre recombinase, which leaves only poly A signals in the genome, cells continue to be foci-free. Conclusion: This study demonstrates that genome modification by integration of polyA signals upstream of CTG repeats prevents the production of toxic RNA mutant transcripts. This genome modification technique should overcome one of the major hurdles in the development of autologous cell replacement therapy for DM1.

2195S

Choroid plexus-directed gene therapy using an α-*N*-acetyl-glucosaminidase-IGF2 fusion protein in MPS IIIB mice. *S.H. Kan[1], S.Q. Le[1], K.N. Vondrak[1], M.R. Haddad[2], E.Y. Choi[2], A. Donsante[2], S.G. Kaler[2], P.I. Dickson[1].* 1) Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; 2) Eunice Kennedy Shriver National Institute of Child Health and Development, National Institutes of Health, Bethesda, MD, USA.

Mucopolysaccharidosis type IIIB (MPS IIIB; Sanfilippo B) is an inherited neurodegenerative disorder for which no reliably treatment is currently available. The cause of MPS IIIB is deficiency of the lysosomal enzyme, α-*N*-acetyl-glucosaminidase (NAGLU) and accumulation of heparan sulfate glycosaminoglycans (GAG). Impediments to enzyme replacement therapy (ERT) include the short half-life, absence of mannose 6-phosphate (M6P), and poor blood-brain barrier penetration associated with recombinant human NAGLU. A modified human NAGLU fused to the receptor binding motif of insulin-like growth factor 2 (hNAGLU-IGF2) has been shown to enhance entry to MPS IIIB fibroblasts via M6P/IGF2 receptor-mediated endocytosis. In this study, we administered a recombinant adeno-associated virus, serotype 5 (AAV5) vector expressing rhNAGLU-IGF2 that targets the choroid plexus epithelia via lateral ventricle injection to deliver the deficient enzyme to the cerebrospinal fluid (CSF) of MPS IIIB mice.

We cloned the hNAGLU-IGF2 cDNA into an AAV vector plasmid and generated rAAV5 by the triple transfection method. NAGLU activity assay and western blots confirmed robust expression of the recombinant protein in HEK 293T cells. An *in vivo* pilot study was then performed in MPS IIIB mice by injecting 5×10^{10} vector genomes (v.g.) of rAAV5-rhNAGLU-IGF2 to the lateral ventricles of adult mice, or 5×10^{9} v.g. in neonatal mice. Mouse brains were analyzed 2 weeks after viral injection and compared with brains of untreated MPS IIIB and control (heterozygous) mice. NAGLU activity reached twice normal levels in the AAV5-treated brains. β-hexosaminidase activity, which is abnormally elevated in MPS IIIB, was significantly reduced in the AAV5-NAGLU-IGF2 treated brains. Histochemical evaluations confirmed NAGLU-IGF2 expression in the choroid plexus epithelia. NAGLU-IGF2 was also detected in brain parenchyma, even contralateral to the administered site, indicating delivery from the CSF. Evaluation of the efficacy of AAV5 gene therapy in preventing brain lysosomal storage and neurobehavioral abnormalities is in progress.

These preliminary results suggest that the combination of M6P/IGF2 receptor-mediated endocytosis and choroid plexus-targeted viral gene therapy may overcome the major obstacles associated with ERT for MPS IIIB and enable, sustained, and efficient distribution of NAGLU throughout the brain.

2196M

Glycogen storage disease type Ia mice receiving gene therapy are protected against age-induced obesity and insulin resistance. *GY. Kim[1], JH. Cho[1], YM. Lee[1], CJ. Pan[1], HS. Jun[1], BC. Mansfield[1,2], JY. Chou[1].* 1) Section on Cellular Differentiation, Program on Developmental Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD; 2) Foundation Fighting Blindness, Columbia, MD.

Glycogen storage disease type Ia (GSD-Ia), characterized by impaired blood glucose homeostasis, fasting hypoglycemia, and hepatocellular adenoma (HCA), is caused by a deficiency in glucose-6-phosphatase-α (G6Pase-α or G6PC) activity. In a long-term, dose-ranging study, we showed that systemic administration of rAAV-G6PC, a recombinant AAV2/8 vector expressing G6Pase-α directed by the G6PC promoter/enhancer, delivered the G6Pase-α transgene to the liver of GSD-Ia mice. The old (70-90 week-old) rAAV-G6PC-treated GSD-Ia mice expressing 3-63% of normal hepatic G6Pase-α activity (AAV-mice) maintained glucose homeostasis, tolerated prolonged fasting, and showed no evidence of HCA. Interestingly, the AAV-mice exhibit better metabolic controls than their control littermates with fasting blood insulin levels closer to the normal values and a leaner phenotype. Mice overexpressing the transcription factor carbohydrate response element binding protein (ChREBP) in the liver exhibit improved glucose tolerance and are protected against insulin resistance, and ChREBP signaling is activated by G6P. The elevated levels of hepatic G6P in the AAV-mice suggests that ChREBP signaling would be activated. The AAV-mice also produced reduced levels of hepatic glucose, averaging 61-68% of the control littermates, suggesting that these AAV-mice lived under a chronic calorie restriction (CR). A hallmark of aging is mitochondrial dysfunction, and several longevity factors in the CR pathway have been identified that modulate mitochondrial function. These include: 1) the NADH shuttle systems that reoxidize cytosolic NADH produced by glycolysis, interconnecting glycolysis, lipogenesis, and mitochondrial oxidative phosphorylation; 2) intracellular levels of NAD^+ that are decreased in obesity and the aged; and 3) the AMP-activated protein kinase (AMPK)/peroxisome proliferator-activated receptor-γ coactivator 1α (PGC-1α) pathway, the major regulator of mitochondrial biogenesis in response to energy depletion. We now show that during aging, AAV-mice are protected against age-induced insulin resistance and obesity and that this correlates with: activation of ChREBP signaling; increases in glycerol-3-phosphate shuttle system expression; increases in NAD^+ concentrations; and activation of the AMPK/PGC-1α pathway in the liver of the AAV-mice, suggesting that reduced hepatic G6Pase-α activity is beneficial to longevity and a lean phenotype.

2197S

Developing Treatment for Neurofibromatosis. *A. Cosco, W. Salloom, A. Metzenberg, A. Metzenberg.* CSUN, 18111 Nordhoff Street, Northridge, CA 91330.

Neurofibromatosis type 1 (NF1) is a relatively frequent autosomal dominant disorder. The symptoms include neural tumors, which can affect vision if they occur in the area of the optic nerve. NF1 shows genetic heterogeneity, and different mutations in the NF1-GRD domain result in the NF1 phenotype. Furthermore, this disorder shows variable expressivity, in that different members of a single family may have a different set of symptoms. Tumor development occurs unexpectedly, making prevention difficult. The NF1 gene is a tumor suppressor, and encodes a protein called Neurofibromin. This protein has an arginine finger motif, and is able to aid in the hydrolysis of RAS-GTP to RAS-GDP. The faulty Neurofibromin protein is no longer able to enhance the intrinsic ability of RAS to turn itself off from a GTP bound state to a GDP bound state. The result may be that a permanently active RAS, leading to the development of tumors in the central and peripheral nerve systems. One possible preventive treatment for NF1 would involve gene therapy. We hypothesize that it is possible that a plasmid vector for this gene therapy would be preferable to a viral vector, because it would avoid the problems of inactivation of other genes by random insertion, and activation of the immune system. A plasmid was constructed using the One- Step Gibson cloning ligation technique, inserting the NF1-GRD domain into the vector. The construct was transfected into a cultured NF1-/- cell line derived from an individual with NF1, using Lipofectamine 2000. The effectiveness of the introduction of pEPito vector containing NF1-GRD was evaluated by estimating the level of active RAS in cells transfected with the construct, using an ELISA kit. The captured GTP-Ras was detected by an Anti-pan-Ras Antibody and HRP conjugated secondary antibody. A significant difference in RAS-GTP levels ($p<0.05$)between cells that were transfected vs non transfected cells suggests that the GAP domain of Neurofibromin was successfully transcribed and compliments the inability of these cells to inactivate RAS. These results support the idea that it would be possible to develop a gene therapy treatment for NF1 that would not cause certain side effects, and would cross the blood-brain barrier , allowing treatment for a disorder that affects the central nervous system, and can cause blindness.

2198M

Cell Reprogramming Technologies for the Treatment of Genetic Disorders of Myelin. *A. Lager[1], Z. Nevin[1], J. Heaney[2], P. Tesar[1,3].* 1) Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) New York Stem Cell Foundation, New York, NY.

Cellular therapies are a promising treatment modality for a variety of human myelin diseases. In previous work we describe reprogramming methods for generating functional oligodendrocyte progenitor cells (OPCs) from autologous sources (Najm & Lager et al Nat Biotech 2013). However, in cases of dysmyelinating diseases with genetic etiology (e.g. Pelizaeus-Merzbacher and Spastic Paraplegia Type 2), autologous cells harbor genetic abnormalities that preclude oligodendrocyte function. Here we present technology for generating functional OPCs from fibroblasts isolated from shiverer mice. Shiverer mice harbor a 20kb deletion in a locus containing myelin basic protein (MBP) and consequently lack compact myelin and exhibit severe motor deficits. We outline methods for correction of the MBP-shiverer allele in induced pluripotent stem cells (iPSCs). These methods utilize the CRISPR/Cas9 genome editing system for induction of homology-directed repair based on an exogenously supplied bacterial artificial chromosome that contains the MBP-wt allele. Gene corrected iPSCs will be differentiated into OPCs, and we hypothesize that introduction of gene-corrected shiverer OPCs back into the central nervous system of shiverer mice will allow myelination of hypomyelinated axons and behavioral improvements. This work will provide proof-of-concept for gene correction of large deletions by homologous recombination, with important clinical implications in treatment of microdeletion syndromes.

2199S

Rescue of lethal hypophosphatasia model mice by adeno-associated virus mediated muscle specific expression of bone targeted alkaline phosphatase. *A. Nakamura[1], O. Iijima[1], K. Miyake[1], A. Watanabe[1,2], Y. Hirai[1], H. Kinoshita[3], T. Noguchi[3], S. Abe[3], T. Okada[1], T. Shimada[1,2].* 1) Department of Biochemistry and Molecular Biology, Division of Gene Therapy Research Center for Advanced Medical Technology, Nippon Medical School, Tokyo, Japan; 2) Division of Clinical Genetics, Nippon Medical School Hospital, Tokyo, Japan; 3) Department of Anatomy, Tokyo Dental College, Tokyo, Japan.

[BACKGROUND] Hypophosphatasia (HPP) is a systemic skeletal disease, caused by the deficiency of tissue-nonspecific alkaline phosphatase (TNALP). The TNALP knockout (*AKP*2-/-) mice phenotypically mimic the severe infantile form of HPP and usually die by 3 weeks of age suffering epileptic seizures and apnea. Enzyme replacement therapy (ERT) is a potential approach, but repeated injection of large amounts of recombinant TNALP is required for clinical benefit. As an alternative approach, we have recently shown that *AKP*2-/- mice can be treated by a single intravenous injection of adeno-associated viral (AAV) vector expressing bone-targeted TNALP (TNALP-D10) in the neonatal period. However, the safety of a systemic injection of AAV into the neonates has not yet been established. In this study, to exploit a safe and clinically applicable protocol, we designed a new self-complementary type 8 AAV vector containing the muscle-specific muscle creatine kinase (MCK) promoter and examined the feasibility of muscle directed gene therapy for HPP. **[METHODS]** Self-complementary AAV8 expressing enhanced green fluorescence protein (scAAV8-MCK-EGFP) and TNALP-D10 (scAAV8-MCK-TNALP-D10) driven by the MCK promoter were generated and injected into the quadriceps femoris muscle of newborn C57BL6 (WT) or *AKP*2-/- mice. **[RESULTS and DISCUSSION]** Stringent muscle specificity of the MCK promoter was demonstrated after intramuscular (IM) injection of scAAV8-MCK-EGFP (5×10^{11} vector genome/mouse) into WT mice (n=3). Neonatal *AKP*2-/- mice (n=10) were treated with a single IM injection of scAAV8-MCK-TNALP-D10 (2.5×10^{12} vector genome/mouse). The plasma ALP activity was rapidly elevated and maintained at a therapeutic level (>1.0 unit/ml) for at least 3 months. The treated *AKP*2-/- mice grew well and survived over 3 months with healthy appearance (9/10), while untreated *AKP*2-/- mice died within 3 weeks ($P<0.001$). Improved mineralization of the knee joints was demonstrated on X-ray images and micro CT analysis. Ectopic calcification was not detected in treated mice. These results suggest that muscle directed gene therapy using scAAV8-TNALP-D10 would be safe and effective to cure the severe infantile form of HPP.

2200M

Functional genomic analysis of the interplay between the retrograde transport machinery and HIV-1 replication. *D. Dykxhoorn[1,2], S. Liu[1].* 1) John P. Hussman Institute for Human Genomics, Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 2) Department of Microbiology and Immunology University of Miami Miller School of Medicine.

Human immunodeficiency virus (HIV) remains a global health concern affecting millions of individuals worldwide. By attacking the host's immune system, HIV leaves the infected individual susceptible to infection by a wide range of pathogens. Like all viruses, HIV-1 is dependent on host cell factors for successful infection, replication and release of progeny virus. Previously, a large-scale functional genomic screen was performed which identified over 230 novel factors, termed HIV-1 dependency factors (HDFs) whose silencing inhibited viral replication. Detailed analysis of these HDFs showed enrichment for factors involved in the trans-Golgi network (TGN), a pathway which had not been previously implicated in HIV replication. This included components of large multiprotein complexes that facilitate the recognition, docking and fusion of different vesicles as they are trafficked through the cell. The targeted silencing of these HDFs, including components of the conserved oligomeric golgi (COG) complex and the Golgi-associated retrograde protein/vesicular protein sorting fifty three (GARP/VFP) complexOur results show that the targeted silencing of components of COG complex each impaired HIV-1 replication in HIV-1-susceptible cell lines and primary monocyte-derived macrophages and inhibited replication against multiple strains of the virus. More detailed analysis of the different forms of the HIV-1 genome showed that the defect in HIV-1 replication occurred prior to late RT product formation, suggesting that these HDFs are required for very early steps in the HIV-1 life cycle. These results highlight a novel pathway required for HIV-1 replication and infection which may hold promise for the development of novel approaches to prevent HIV-1 transmission.

2201S

Toward Treatments for Genetically-Based Protein Misfolding Diseases Using Antisense Technology. *S. Guo[1], S. Booten[1], S. Murray[1], E. Ackermann[2], S. Hughes[2], M. McCaleb[1], B. Monia[1].* 1) Antisense Drug Discovery, Isis Pharmaceuticals, Carlsbad, CA; 2) Clinical Development, Isis Pharmaceuticals, Carlsbad, CA.

Protein misfolding diseases, also known as proteopathies, are a class of diseases that are caused by changes in protein conformation which often lead to gain of toxic function. Antisense technology utilizes RNase H mechanism to degrade RNA and subsequently remove unwanted protein products in the cell, potentially providing an optimal platform for the treatment of these diseases. We have evaluated potent antisense oligonucleotides (ASOs) for three genetically-based protein misfolding dsieases: 1) Transthyretin amyloidosis (ATTR), 2) Alpha-1 antitrypsin liver disease (AAT serpinopathy), and 3) autosomal dominant retinitis pigmentosa due to a mutation (P23H) in the rhodopsin gene. ISIS-TTR_{Rx} targets all mutant forms as well as wild type TTR mRNA for the treatment of ATTR. In a Phase 1 normal volunteer study, ISIS-TTR_{Rx} was well tolerated and achieved mean reductions in plasma TTR of ~75% at the 300mg dose level. This compound is currently being evaluated in a multicenter, double-blind, placebo-controlled Phase 3 trial in ATTR familial amyloid polyneuropathy patients. In a mouse model of AAT liver disease (PiZ mice), treatment with an antisense oligonucleotide targeted to the alpha-1 antitrypsin mRNA (AAT-ASO) prevented liver disease progression after short-term treatment; reversed liver disease after long-term treatment, and prevented liver disease in young animals. Furthermore, ASO treatment markedly decreased liver fibrosis in these mice. In addition to the reduction of peripheral targets like TTR and AAT which are mainly produced by liver, ASOs can be delivered directly into the eye via intravitreal injection and distribute to all the cell layers in the eye. P23H rhodopsin transgenic rats express both mutant mouse and normal rat rhodopsin, representing an excellent model of retinitis pigmentosa. We developed ASOs targeting the diseased allele (mouse mutant Rho), but not the normal endogenous allele in this model. Using allele-specific ASO, we demonstrated a slower progression of photoreceptor degeneration and improved electroretinography (ERG) measurements 30 days after a single intravitreal injection. Eyes injected with P23H ASO had a 128 ± 22% improved amplitude response and maintained more outer nuclear layers compared to PBS injected contralateral eyes. Our data demonstrate that antisense technology provides promising treatments for genetically-based protein misfolding diseases.

2202M

AUDITORY VERBAL THERAPY IN PEDIATRIC PATIENTS WITH CONGENITAL DISEASES. *M. Conde-Pacheco.* Audiology and Rehabilitation, Hospital para el Nino Poblano, Puebla, Puebla, Mexico.

The hearing, speech and language are essential parts in the development of a child, to speak and master the language the child needs to hear. If the hearing is the starting point of the main distinguishing characteristic of human beings that is language, their absence or reduction imply important consequences of the type; personal , family, social, educational and cultural, is closely linked to intelligence, thought and imagination. A child who does not hear, not only cannot learn to speak, but cannot appropriate the reading and writing that are the basis of cultural development. Language, in conclusion, is a fundamental requirement for a useful life. If the child's hearing loss is detected early, (considering that most children with this problem have residual hearing) and if this remnant increases with technological resources and appropriate auditory verbal therapy, these children can lead a normal life. The Ministry of Health of the State of Puebla has implemented in the Hospital para el Niño Poblano and other hospitals, health programs which include a program for the habilitation of hearing impaired children, this program includes; hearing screening in neonates, implantation of hearing aids or cochlear implants and conducting habilitation therapies. It is very important to mention that technological resources, (hearing aids or cochlear implant), are not solutions by themselves, are just only an opportunity for the child to live a normal life always and when the auditory is stimulated early, appropriately and continuously. These points are the objectives of the area of Auditory Verbal therapy inside the program to hearing impaired children in the HPNP. The Auditory Verbal Therapy provided in the HPNP, has as main objective the development of listening and speaking skills in children with hearing aids or cochlear implants. In conclusion, under all circumstances and mainly when the cochlear implant or the hearing aid were implanted late, auditory verbal therapy is essential to obtain the best development of the child. To obtain this is strictly required that the cochlear implant or the hearing aids are rightly adjusted or mapped, the child must hear at least a minimum range of frequencies. If due to different circumstances, the last point is not obtained, the development of the child will be unpredictable.

2203S

Combination therapy with artificial dermis, growth factors, and cell culture techniques for massive hematoma-induced skin necrosis in a patient with dermatan 4-O-sulfotransferase 1 (D4ST1)-deficient Ehlers-Danlos Syndrome (DDEDS). *F. Nagai[1], S. Yano[2], K. Wakui[3], Y. Higuchi[4], S. Shimodaira[4], K. Fujita[1], M. Noguchi[1], K. Matsuo[2], T. Kosho[3], S. Yuzuriha[2].* 1) Department of Plastic Surgery, Nagano Children's Hospital, Azumino, Nagano, Japan; 2) Department of Plastic and Reconstructive Surgery, Shinshu University School of Medicine, Matsumoto, Japan; 3) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 4) Cell Processing Center, Shinshu University Hospital.

Dermatan 4-O-sulfotransferase 1 (D4ST1)-deficient Ehlers-Danlos syndrome (DDEDS), caused by recessive loss-of-function mutations in CHST14, is a recently delineated form of EDS, characterized progressive multisystem fragility-related manifestations (skin hyperextensibilty and fragility, progressive spinal and foot deformities, massive subcutaneous hematoma) and various malformations (facial features, congenital multiple contractures) [Miyake et al., 2010; Kosho et al., 2011]. Massive hematoma is the most serious complication for patients with DDEDS. We present successful combination therapy with artificial dermis, growth factors, and cell culture techniques for massive hematoma-induced skin necrosis in an 18-year-old girl with DDEDS. She developed progressive necrosis of the skin and soft tissue throughout the left leg, resulting from massive subcutaneous hematoma occurring associated with manual reposition of the left hip dislocation caused by a traffic accident. She was referred to our hospital on day 33 after the injury to treat the necrosis. Autologous cultured dermis was generated by cultivating the patient's own fibroblasts and seeding them in artificial dermis before the first operation. The first operation, on day 43, included excision of all necrotic tissue and application of autologous cultured dermis prepared from her skin fibroblasts and artificial dermis. The second operation, on day 69, included application of split-thickness chip skin grafts taken from the right thigh and autologous cultured epithelia. The third operation, on day 109, included grafting small pieces of split-thickness skin from the right thigh again to cover skin ulcer. Intravenous hyperalimentation improved wound heeling and the tissue defect was completely epithelialized on day 230. She was able to walk without contracture in the knee joint in spite of immobility over 10 months. This is the first report of massive hematoma-induced skin necrosis as a potentially fatal complication of DDEDS and the successful treatment with application of autologous cell therapy and grafting techniques.

2204M

Feasibility of Exon-Skipping Therapy for Juvenile Neuronal Ceroid Lipofuscinosis. *M. Velinov[1,2], N. Dolzhanskaya[1].* 1) Human Genetics, NYS Institute for Basic Research in Developmental Disabilities, Staten island, NY; 2) Albert Einstein College of Medicine, Bronx, NY.

Rationale: Juvenile Neuronal Ceroid Lipofuscinosis (JNCL) is a progressive lethal, neurodegenerative disorder and is the most common type in the group of Neuronal Ceroid Lipofuscinoses. 95% of the JNCL patients carry a common 1 kb deletion in one or both alleles of gene CLN3. The deletion includes exons 7 and 8, and leads to reading frame shift and formation of truncated protein. No effective therapy for JNCL is available. Splicing out (skipping) of exon 9 may be induced in cells with the common CLN3 deletion using antisense oligonucleotides (AO). Such exon skipping would restore the correct reading frame and is expected to produce a modified protein with correct aminoacid sequence downstream of exon 9. Induction of exon skipping (ES) may lead to the increase of the CLN3 transcript in mutant cells by decreasing the nonsense-mediated decay. ES may also help restore some of the CLN3 protein function since the correct aminoacid sequence downstream of exon 9 would be restored. Such approach is currently used in a clinical trial for patients with Duchenne Muscular Dystrophy. **Exon skipping:** Exon splicing enhancer sequences within exon 9 were identified using the program ESEFinder. Four modified oligonucleotides with 2'-O-methyl modified bases on a phosphothiolate backbone, complementary to splicing enhancer sequences were synthesized. Fibroblast cell lines from patient, homozygous for the 1 kb deletion in CLN3 and phenotype of JNCL were plated 24 h prior transfection and were transfected with all four AO using Lipofectamine 2000 reagent. The cells were harvested 18 h after transfection. Total RNA was isolated and used for RT-PCR analysis. The AO transfection resulted in complete exon 9 skipping in the detectable mutant CLN3 transcript. These results were confirmed with sequencing of the RT-PCR product. . **Increase of the transcript level:** Total RNA extracted after AO treatment was used for real time quantitative RT-PCR. The total mutant transcript level increased with 30% after AO treatment compared to the non-treated mutant transcript. **Intranasal administration:** Two mouse-specific AO were labeled with immunofluorescent marker, and were applied intranasaly in a mouse model of JNCL. The intranasaly applied AO penetrated the blood brain barrier in detectable amounts, and were observed in the mouse neuronal cells. **Conclusions:** This prove of concept study showed that AO based ES therapy for JNCL may be feasible using intranasal application.

2205S

Improving the communication skills of Down Syndrome patients through speech therapy. *A. Umrigar[1], A. Musso[2], K. Foley[1], M. Banajee[2], F. Tsien[1].* 1) Department of Genetics, LSU Health Sciences Center, New Orleans, LA; 2) Department of Communication Disorders, LSU Health Sciences Center, New Orleans, LA.

Two major obstacles that Down Syndrome patients encounter in the process of communicating effectively are the craniofacial abnormalities and cognitive deficits associated with the syndrome. Physiological features, such as a small oral cavity, large tongue, high arched palate, and facial hypotonia impede the clear articulation of speech. In addition to the physical abnormalities, cognitive deficits further interfere with speech and comprehension. Early intervention in the forms of speech therapy and hearing loss treatments such as hearing aids are essential in overcoming these impediments and improving the communication skills of Down Syndrome patients. This study focuses on detailed communication evaluations of Down Syndrome patients enrolled in the Louisiana State University Health Sciences Center (LSUHSC) Speech Language Therapy program from 6 to more than 15 years. The patients' treatment plans and progress were examined weekly in order to determine the effect of speech therapy on the patient's comprehension and articulation of both verbal and written forms of communication. Patient improvement was monitored using the Goldman-Fristoe test of articulation, repetitive one word vocabulary test, expressive one word vocabulary test, and oral-motor/range of motion evaluation. In addition, these patients received audiological assessments. Concrete progress goals and objectives were constructed specific to each patient. All assessment methods demonstrated a significant improvement in effective communication skills, including progression from one-word to full sentences, vastly improved articulation, and communication in a group setting compared to those who did not receive any form of communication or speech therapy. It is therefore important for geneticists to recognize the improved quality of life benefits of communication intervention programs and refer Down Syndrome patients to speech pathologists.

2206M

The outcome of N-carbamylglutamate (Carbaglu®) therapy in a Korean patient with N-acetylglutamate synthase deficiency. *K. Woo[1], J. Choi[2], B. Lee[1,2], H. Yoo[1,2].* 1) Genome, Asan Medical Center, Seoul, South Korea; 2) Dept. Pediatrics, Asan Medical Center, Univ. Ulsan College of Medicine,Seroul, South Korea.

N-acetylglutamate synthase (NAGS) deficiency is a rare autosomal recessive disorder of the urea cycle. Affected individuals present with poor oral intake, vomiting, seizure, lethargy, and coma within a first few days of life. N-carbamylglutamate (Carbaglu®), a synthetic derivative of glutamic acid as well as a structural analogue of human N-acetylglutamate, is a treatment choice for NAGS deficiency. We report an 8-day-old boy with NAGS deficiency who has been treated with N-carbamylglutamate (Carbaglu®). The patient was born from non-consanguineous Korean parents with a birth weight of 3.4 kg after 40 weeks of gestation. At age 8 days, he showed poor oral intake and lethargy with hyperammonemia (800 μM), which was managed by peritoneal dialysis before referral to our institution. Plasma amino acid profiles revealed elevated glutamate (1117 μmol/L), low-normal arginine (51 μmol/L), ornitine (49 μmol/L), and low citrulline (7 μmol/L) levels. Urine orotic acid was normal (1.77 μmol/molCr; range 0.2-6 μmol/molCr). He has been placed on sodium benzoate, carnitine and citrulline therapy with protein restriction diet with presumptive diagnosis of carbamoyl phosphate synthase (CPS) deficiency. He was referred to our center at age 2 years. Genetic testing for CPS1 was negative. At age 7 years, the NAGS genes analysis using cDNA extracted from liver tissues identified compound heterozygote of p.V310A and p.H488Qfs*2, which were inherited from both parents. He showed mild mental retardation (IQ 51), epilepsy, and attention-deficit/hyperactivity disorder. Despite protein restriction and administration of nitrogen scavenger medicines, he presented recurrent hyperammonemia, stunted growth velocity, and poor weight gain. N-carbamylglutamate (Carbaglu®) replaced since age of 8.1 years, and he discontinued nitrogen scavenger drugs. After 4 months, he discontinued special formula and has been on liberal diet without protein restriction. During past 9 month, he has maintained plasma ammonia within normal range and showed normal prepubertal growth velocity. In conclusion, N-carbamylglutamate (Carbaglu®) has been effective for NAGS deficiency to maintain stable metabolic state and growth without any complications.

2207S

Modeling Autosomal Dominant Optic Atrophy in Induced Pluripotent Stem Cells (iPSCs) and Identifying the Potential Therapeutic Targets. *J. Chen[1], H. Riazifar[2], M. Guan[1], T. Huang[1].* 1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 2) Department of Pediatrics, Division of Human Genetics, University of California, Irvine, CA, USA.

Purpose: Many retinal degenerative diseases are caused by loss of retinal ganglion cells (RGCs). Autosomal dominant optic atrophy (DOA) is the most common hereditary optic atrophy characterized by central vision loss and degeneration of RGCs. DOA is largely caused by mutations of a nuclear gene *OPA1*, which encodes an inner mitochondrial membrane protein. Currently, there is no effective treatment for these types of diseases. However, stem cell therapy holds great potential by replacing patient lost RGCs. Compared with embryonic stem cells, induced pluripotent stem cells (iPSCs) are derived from adult somatic cells, associated with fewer ethical concerns, and are less prone to immune rejection. In addition, these patient-derived iPSCs may serve as a cellular model, allowing us to closely study pathogenesis and therefore to identify potential therapeutic agents. **Methods:** In this study, we were able to generate specific iPSCs from the patient carrying an *OPA1* mutation and diagnosed with optic atrophy. We differentiated iPSCs into RGCs by using our established protocol. **Results:** We found that the mutation of *OPA1* in patients with optic atrophy led to significantly increased apoptosis and *OPA1*-iPSCs were not able to differentiate into RGCs. However, adding neuron induction medium or beta-estrogen into differentiation medium efficiently promoted *OPA1*-iPSC differentiation into RGCs. **Conclusions:** Our results suggest that apoptosis mediated by OPA1 mutations plays a very important role in DOA pathogenesis and both noggin and beta-estrogen could be potential therapeutic agents for *OPA1* mutations related optic atrophy.

2208M

Complete Huntingtin Haplotypes for Allele-Specific Silencing. *C. Kay[1], J.A. Collins[1], N.H. Skotte[1], A.L. Southwell[1], A. DiPardo[2], C. Ross[1], F. Squitieri[2], M.R. Hayden[1].* 1) Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada; 2) Neurogenetics Unit and Rare Disease Centre, IRCCS Neuromed Pozzilli (IS), Italy.

BACKGROUND: The dominant CAG repeat expansion causing Huntington disease (HD [MIM 143100]) occurs on specific haplotypes of the *HTT* gene. In humans, at least 2300 polymorphisms occur across *HTT* and offer potential targets for allele-specific silencing of the mutant transcript encoding the CAG expansion. The organization of all *HTT* polymorphisms into gene-spanning haplotypes, and the portion of the HD patient population treatable by selective silencing of disease-associated haplotypes, is unclear and needed to guide allele-specific *HTT* silencing strategies.

AIMS: Our study sought to identify all genetic variants occurring on the most common HD-associated haplotypes, A1 and A2, and determine the frequency of these haplotypes among distinct HD patient populations of European ancestry. We additionally sought to demonstrate selective *HTT* silencing using novel HD-associated polymorphisms identified by this work.

RESULTS: Using whole-genome sequencing from the 1000 Genomes Project, and direct genotyping of diverse HD patient cohorts from Canada, Sweden, France, and Italy, we assemble complete haplotypes for all known common variants (>5%) in *HTT*. The most common HD-associated haplotype, A1, is uniquely defined by three transcribed polymorphisms; the second most common HD haplotype, A2, is uniquely defined by five intragenic polymorphisms. A1 is most heterozygous in Northern European patient populations, where HD is most common, whereas A2 is most heterozygous in Southern European patients. Using antisense oligonucleotides (ASOs) targeting a novel A1-defining polymorphism, we silence mutant A1 *HTT* mRNA and protein in patient-derived cells, maintaining wild-type *HTT* expression at untreated levels.

CONCLUSIONS: Across populations of European ancestry, silencing the A1 haplotype may offer allele-specific therapy in approximately half of HD patients. Our results suggest that targeting the A2 haplotype may allow selective *HTT* silencing in an additional quarter of patients across all four populations, and over half of HD patients in Italy. In combination, antisense reagents silencing A1 and A2 may offer selective silencing of mutant *HTT* in >70% of HD patients.

2209S

iPS-cell derived basal keratinocytes and melanocytes to study severe monogenic genodermatoses in patient-specific 3D tissue systems. *K.M. Eckl[1], D.M. de Lima Cunha[1,2], R. Plank[1,2], R. Casper[2], H. Traupe[3], M. Rauch[1], M.K. Gupta[4], T. Saric[4], H.C. Hennies[1,2].* 1) Ctr. for Dermatogenetics, Div. of Human Genetics, Innsbruck Medical University, Innsbruck, Austria; 2) Ctr. for Dermatogenetics, Cologne Ctr. for Genomics, Univ. of Cologne; 3) Dept. of Dermatology, University Hospital Münster, Germany; 4) Inst. for Neurophysiology, Medical Center, Univ. of Cologne, Cologne, Germany.

Induced pluripotent stem (iPS) cells from a somatic source like dermal fibroblasts from a patient with a known severe, often fatal, but always rare genodermatosis are a unique and excellent tool to study the pathomechanism during neuro-ectodermal differentiation, melanocyte development and epidermal commitment and finally epidermal stratification. In contrast to primary keratinocytes and melanocytes, genetic modification of iPS cells can be easily performed and allows stable gene correction and fast in-vitro gene-therapy approaches. Up to date, for most genodermatoses, like neurofibromatosis 1 (NF-1), congenital ichthyosis (autosomal recessive congenital ichthyosis, ARCI), and Cole disease, therapy is only symptomatic and rarely sufficient. Although genetically and functionally well characterised, patient-specific, individualized therapies have so far not been generated for all of these disorders. Lately we published 3D skin models for different monogenic and severe genodermatoses. These models can be used for the development of protein replacement therapeutics (PRT). To overcome the shortage of cells for large-scale patient-specific skin modelling, keratinocytes and melanocytes can be differentiated from patient iPS cells and incorporated into the 3D skin model. We established a protocol to differentiate diseased and healthy iPS cells into the neuro-ectodermal fate (p63 positive cells) using a feeder-free culture system. Our protocol is based on a defined medium, which is supplemented with BMP4, retinoic acid, and ascorbic acid. At specific time points during the 30-40 days of differentiation enhancers of the Wnt/β-catenin pathway are additionally supplemented. Neuroectodermal cells are then specifically differentiated in either keratinocytes or melanocytes. For melanocyte differentiation cells are passaged daily in the presence of a melanocyte-specific, defined medium which is further supplemented. During keratinocyte differentiation we observe first keratin 18 expression (day 12), then around day 30 a strong keratin 14 expression indicating basal keratinocytes; at day 30 at least 50% of all keratinocytes are positive for CD104. iPS cell-derived melanocytes are positive for Mel5, Tryp1 and Mart1. Only iPS-cell derived keratinocytes and melanocytes allow us to generate patient-specific 3D tissue models in larger quantities, which will be essential to assess drugs for these orphan diseases and pave the way for individualized therapies.

2210M

A Fine Balance of Dietary Lipids Improves a Murine Model of VCP-associated Disease. *K.J Llewellyn, A. Nalbandian, N. Walker, J. Tang, A. Gomez, V. Kimonis.* Pediatrics, 2501C Hewitt Hall, Irvine, CA.

Despite intense investigations, the discovery of effective novel advancements/therapies and the disease mechanisms underlying Valosin containing protein (VCP)-associated myopathies and neurodegenerative disorders remain elusive. VCP diseases, caused by mutations in the VCP gene, are a clinically and genetically heterogeneous group of disorders with manifestations varying from hereditary inclusion body myopathy, Paget's disease of bone, frontotemporal dementia (IBMPFD), amyotrophic lateral sclerosis (ALS), and other neurodegenerative changes. Affected individuals exhibit scapular winging and progressive muscle weakness and die from cardiac and respiratory failure. Histologically, patients display rimmed vacuoles and TAR DNA Binding Protein-43 (TDP-43)-positive ubiquitinated inclusion bodies in muscles. Currently, there are no effective treatments for patients with VCP-related myopathies. VCP mouse models carrying the common R155H mutation include several of the clinical features typical of human diseases. Here, we examined the effects of varying dietary lipid percentages on VCPR155H/+and wild type (WT) mice from birth till 9 months of age. Disease progression was monitored and analyzed using survival curves, pathological and immunohistochemical methods. Mice on the 9% lipid-enriched diet demonstrated an improvement in muscle strength measurements, histology, and autophagy signaling pathway when compared to mice on normal diet. In this report, we demonstrate VCP mice fed increasing lipid diets of 12%, 15%, 30%, and 48% showed no improvement in muscle pathology or the autophagy cascade, suggesting that significantly increased lipid diets only depict a detrimental effect in mice. Thus, a balanced lipid supplementation suggests a promising translational therapeutic strategy for patients with VCP-associated neurodegenerative diseases.

2211S

Development of autologous myogenic stem cell therapy for carriers of a heteroplasmic mtDNA mutation: a proof of principle study in m.3243A>G mutation carriers. *F. van Tienen[1,2], E. Timmer[1,2], M. Quattrocelli[3], M. Sampaolesi[3], I. De Coo[4], H. Smeets[1,2,5].* 1) Clinical Genetics, MaastrichtUMC, Maastricht, Netherlands; 2) Research School GROW, MaastrichtUMC, Maastricht, Netherlands; 3) Interdepartmental Stem Cell institute, KU Leuven, Leuven, Belgium; 4) Dept Neurology, Erasmus UMC, Rotterdam, Netherlands; 5) Research School CARIM, MaastrichtUMC, Maastricht, Netherlands.

Mitochondrial diseases due to heteroplasmic mitochondrial DNA (mtDNA) mutations are severe multisystem disorders (frequency 1/5,000), caused by defects in oxidative phosphorylation (OXPHOS). Severe progressive myopathy and exercise intolerance occur in >50% of the mtDNA disorders with a variety of other clinical manifestations and seriously affect the quality of life and well-being. Effective therapy is currently unavailable. We hypothesize that inducing muscle regeneration by transplantation of mtDNA-mutation free autologous myogenic stem cells (pericytes/mesoangioblasts -MABs) can compensate for the OXPHOS defect and may reverse myopathy and exercise intolerance, resulting in increased mobility and quality of life for heteroplasmic mtDNA mutation carriers. MABs are currently the only myogenic precursors that fulfill all requirements for being used in this therapeutic strategy and a phase I/II clinical trial using allogeneic MABs is currently ongoing in DMD patients. The key characteristic of our project is transplantation of the patients' own MABs that have been freed of the mtDNA mutation. This approach allows targeting all proximal muscles safely without adverse reactions to the transplant and the use of immunosuppressive agents. To this end, we have succeeded in generating mutation-free MABs from m.3243A>G carriers (with adult onset myopathy and a mutation load in skeletal muscle ranging from 15% to 93%) in sufficient amounts for initial local treatment. The in vitro growth rate and myogenic differentiation capacity (ranging between 20%-57%) of these mutation-free MABs were comparable with MABs derived from healthy controls and their in vivo myogenic capacity was subsequently verified in immunodeficient mice. The next step is to perform a proof of principle clinical study to demonstrate the local effect of transplantation of mutation-free autologous MABs, injected in the leg of m.3243A>G carriers. Muscle regeneration, mutation load and OXPHOS capacity will be determined. From this study, we expect to get a clear insight in the potential of this exciting new therapeutic approach, warranting further steps into a systemic clinical trial.

2212M

Online Self-Report Data for Duchenne Muscular Dystrophy: observations of current natural history and explorations of therapeutic responses. *A. Eskin[1], R. Wang[1], J. Ulm[1], C. Silverstein[2], I. Jankovic[1], A. Lu[1], V. Miller[3], R. Cantor[1], N. Li[4], R. Elashoff[4], A. Martin[5], H. Peay[5], S. Nelson[1,6].* 1) Human Genetics, UCLA, Los Angeles, CA; 2) Department of Internal Medicine, College of Osteopathic Medicine of the Pacific, Western University of Health Sciences, Pomona California and the Western Diabetes Institute, Western University of Health Sciences, Pomona, CA; 3) PatientCrossroads, San Mateo, California; 4) Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles; 5) Parent Project Muscular Dystrophy, Hackensack, NJ; 6) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles.

To assess the utility of online patient self-report outcomes in a rare disease, we attempted to observe the well established therapeutic benefit of corticosteroids for Duchenne muscular dystrophy. First, we assessed how comparable an online self report registry for Duchenne, called DuchenneConnect, compared to prior natural history data in regard to age at diagnosis, mutation spectrum, and age at loss of ambulation. Second, we assessed the correlation of reported corticosteroids usage with reported age at fulltime wheelchair use in Duchenne muscular dystrophy (DMD) only using data from 1,057 males from DuchenneConnect. Because registrants reported differences in steroid and other medication usage, as well as age and ambulation status, we could explore these data for correlations with age at loss of ambulation. Using multivariate analysis, current steroid usage was the most significant and largest independent predictor of improved wheelchair-free survival. Thus, these online data were sufficient for the retrospective observation that current steroid use by patients with DMD is associated with a delay in loss of ambulation (a key functional endpoint difficult to assess in traditional study designs). Interestingly, reported use of Vitamin D, CoEnzyme Q10, insurance status, and age at diagnosis after 4 years were also significant, but smaller, independent predictors of longer wheelchair-free survival. This study demonstrates the utility of DuchenneConnect data to observe therapeutic differences, and highlights needs for improvement in quality and quantity of patient-report data, which may allow exploration of drug/therapeutic practice combinations impractical to study in clinical trial settings. Further, with the low barrier to participation, we anticipate substantial growth in the dataset in the coming years.

2213S

Prevention and reversal of transcriptional changes in an induced model of spinal muscular atrophy by administration of an antisense oligonucleotide promoting inclusion of *SMN2* exon 7. *J.F. Staropoli[1], H. Li[1], S.J. Chun[2], N. Allaire[1], P. Cullen[1], A. Thai[1], C. Fleet[1], F. Bennett[2], A. Krainer[3], D. Kerr[1], A. McCampbell[1], F. Rigo[2], J. Carulli[1].* 1) Biogen Idec, Boston, MA; 2) ISIS Pharmaceuticals, Carlsbad, CA; 3) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Spinal muscular atrophy, the leading genetic cause of infant mortality, is a neuromuscular disease characterized by progressive loss of alpha-motor neurons in the anterior horn of the spinal cord. It is caused by disruption of the *SMN1* gene and insufficient compensation of its function by the neighboring, nearly identical paralogous gene *SMN2*, which harbors a splicing silencer element in intron 7 that suppresses inclusion of exon 7 critical for production of full-length SMN protein. Antisense oligonucleotide (ASO)-mediated blockade of the splicing silencer was previously shown to promote inclusion of *SMN2* exon 7 in multiple mouse models of SMA and to mediate phenotypic rescue. To date, however, the molecular signature of this rescue has not been defined. Here we characterize the transcriptional changes that occur in an induced model of type III SMA and the substantial prevention or reversal of those changes in CNS tissue upon intracerebroventricular administration of an ASO promoting inclusion of exon 7. We further show that the timing of ASO administration is correlated with prevention or reversal of specific subsets of transcriptional changes. Collectively, these changes represent potential biomarkers of SMN depletion and therapeutic response to SMN repletion.

2214M

A hemizygous *GYG2* mutation in Japanese siblings showing Leigh syndrome without marked elevation of lactate and pyruvate. *E. Imagawa[1], H. Osaka[2], A. Yamashita[3], M. Shiina[4], E. Takahashi[5], H. Sugie[6], M. Nakashima[1], Y. Tsurusaki[1], H. Saitsu[1], K. Ogata[4], N. Matsumoto[1], N. Miyake[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Division of Neurology, Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Department of Molecular Biology, Yokohama City University School of Medicine, Yokohama, Japan; 4) Department of Biochemistry, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 5) Division of Infection and Immunology, Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 6) Department of Pediatrics, Jichi Medical University, Tochigi, Japan.

Leigh syndrome (LS: MIM#256000) is a progressive neurodegenerative disorder characterized by unique, bilateral neuropathological findings in brainstem, basal ganglia, cerebellum and spinal cord. The estimated incidence of LS is 1:40,000 live births. LS is genetically heterogeneous and the majority of the causative genes affecting mitochondrial malfunction. However, the genetic causes in the majority patients still remain unsolved. Here, we present male sibling affected with LS showing developmental delay, convulsion, athetoid movements, nystagmus, hypotonia, spasticity, increased deep tendon reflex, ketonemia, and bilateral lesion of basal ganglia (caudate nuclei, globus pallidus, and putamen), but no marked elevation of lactate and pyruvate. To identify their genetic cause, we performed genetic analysis using whole exome sequencing. Based on the hypothesis of autosomal recessive and X-linked recessive models, we identified only one hemizygous missense mutation (c.665G>C, p.Trp222Ser) in glycogenin-2 (*GYG2*, isoform a: NM_001079855) in both affected siblings. The heterozygous change was identified in their mother, which is consistent with the X-linked recessive trait. *GYG2* is located at Xp22.33 and encodes glycogenin-2 (GYG2) which plays an important role in the initiation of glycogen synthesis. Based on the structural modeling based on its homologue, glycogenin-1 (GYG1), the alteration from hydrophobic Trp222 to hydrophilic Ser222 may lead to structural instability and create a cavity around the mutated Ser222 residue possibly resulting in protein malfunctioning. Furthermore, *in vitro* experiments showed mutant GYG2 expressed in COS-1 cells was unable to undergo the self-glucosylation observed in wild-type GYG2. We proposed a pathomechanism of the GYG2 impairment in this family based on the canonical pathway of glycogen metabolism. LS is known as one of metabolic disorders that impaired mitochondrial function resulting an energy depletion. In present patients, the central nervous system was dominantly affected. This predominance could be explained by high glucose consumption as the primary energy source in brain. In conclusion, these data highlight the possible link between mutation in *GYG2* and LS. Further studies are needed to conclude whether GYG2 abnormality leads to atypical LS.

2215T

Role of ACBD3 protein in the mitochondrial energy metabolism. *M. Tesarova[1], H. Kratochvilova[1], M. Rodinova[1], V. Stranecky[2], H. Hartmannova[2], A. Vondrackova[1], J. Spacilova[1], M. Hulkova[1], H. Hansikova[1], T. Honzik[1], J. Zeman[1].* 1) Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; 2) Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic.

A homozygous mutation c.460A>G in ACBD3 gene was found by whole-exome sequencing in patient with mitochondrial encephalomyopathy and combined deficiency of respiratory-chain complexes. ACBD3 encodes a 60 kDa acyl-coenzyme A binding domain containing 3 protein (also known as PAP7 or GCP60) localized in the Golgi apparatus, endoplasmic reticulum, and mitochondria. ACBD3 protein plays an important role in many cellular processes (lipid metabolism, membrane transport, neuronal division, embryogenesis, steroidogenesis, structure of the Golgi apparatus, apoptosis, iron homeostasis). In mitochondria, ACBD3 protein is a part of the complex (140-200kDa) mediating cholesterol transport into the organelle. An integral part of this complex is VDAC protein (Porin) which forms freely permeable pores in the outer mitochondrial membrane and further associates with ANT protein (adenine nucleotide transporter) in the inner mitochondrial membrane. ANT is an ADP/ATP carrier that controls a pool of mitochondrial adenine nucleotides required for mtDNA maintenance. The aim of the study is characterize the consequences of stably down-regulated expression of ACBD3 protein on the biogenesis and functions of OXPHOS complexes in HEK293 to elucidate its role in the pathogenesis of the patient's phenotype. The first results show that deficit of ACBD3 protein leads to the decrease of amount and activity of complex IV and decrease of selected subunits of complex IV. Supported by research projects IGA NT/13114-4, IGA NT/11186-5, RVO-VFN64165/2012, and GAUK 1308214.

2216M

Encephalopathy and urea cycle disorder after gastric bypass in a heterozygous G209S short chain acyl-CoA dehydrogenase gene carrier. *H. Wang[1,2], A. Singh[3], N. Shimizu[1], V. Patel[1].* 1) Department of Surgery, Metabolic Support Service, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115; 2) Department of Medical Genetics, Mayo Clinic, 200 First Street SW, Rochester, MN 55905; 3) Department of Surgery, Thomas Jefferson University Hospital, Philadelphia, PA 19107.

INTRODUTION: There is growing evidence of urea cycle disorder (UCD) recognized in adult. Encephalopathy after gastric bypass has been reported increasingly and fetal hyperammonia related encephalopathy has been reported in several cases. Among of them only two cases were reported to have investigated and identified UCD as a culprit of encephalopathy. We describe a UCD whose phenotypic expression became manifest with the stress of malnutrition in an adult bariatric patient. CASE : A 47 year old morbidly obese female with no history of chronic liver disease status post gastric bypass surgery eight years ago presents with altered mental status requiring intubation and ventilatory support. Her post surgical course is notable for severe malnutrition over the preceding 4 months. Admission labs revealed a markedly elevated NH3 139ug/dL peaking at 276 [normal range 40-80] and multiple micronutrient deficiencies. Head MRI is normal. Therapy was initiated with an ammonia scavenger, ammunol and TPN nutrition support. Metabolic panel showed urine orotic acid 2.8 mmol/molCr (0.4-1.2), alanine 184umol/L (240-600), arginine 30umol/L (40-160), citrulline 5umol/L (10-60), glutamine 366umol/L (410-700), isoleucine 0umol/L (30-130), methionine 11 umol/L (17-53), serine 36umol/L (60-200), tyrosine 9umol/L (30-120), valine 44umol/L (140-350). Ornithine, carnitine and acylcarnitine levels were normal. With the exception of low glutamine and alanine, the biochemical findings were consistent with ornithine transcarbamylase (OTC) deficiency. The patient recovered after aggressive treatment. Further molecular genetic analysis did not reveal a detectable mutation in the OTC gene. However, a heterozygous G209S (GGC>AGC) in exon 6 of the acyl-CoA dehydrogenase (ACAD) gene (NM_000017.2) disease-associated variant was identified. DISCUSSION: Mutations in the ACAD gene are associated with short-chain acyl-CoA dehydrogenase (SCAD) deficiency, an autosomal recessive disorder. Only those patients homozygous or compound heterozygous for the mutation are considered to be at risk with most carriers remaining asymptomatic. SCAD-deficient mice show depressed mRNA expression and enzyme activity for urea cycle enzymes carbamyl phosphate synthetase I (CPS) and OTC. The risk of this mutation confers to this patient has not yet been clearly defined but exemplifies the need to suspect UCD in severely malnourished bariatric patients with acute onset of encephalopathy and hyperammonia.

2217T

Persistent elevations in C3 and C5OH, heralding signs for an underlying mitochondiral defect: Neurogenic weakness with ataxia and retinitis pigmentosa (NARP). *N. Hauser.* Medical Genetics Dept, Childrens Hospital Central California, Madera, CA.

INTRODUCTION Mitochondrial diseases are a heterogeneous group of disorders, particularly those caused by defects in mitochondrial DNA. These disorders can affect single or multiple organs and can present with varied symptoms. Here, a patient was eventually diagnosed with NARP who presented with mild biochemical abnormalities, poor coordination and speech delay. PATIENT CASE Patient was a 3.5yrs old male referred for speech delay. Additional concerns were clumsiness and decreased endurance. Metabolic testing revealed mild elevations in C3 2.42uM (0.11-1.25), C5-OH 0.23uM (0-0.06), CK 368U/L (<160),] lactic acid 2.8mmol/L (0-2.2), alanine 796 umol/L (157-481umol/L) and urine 3OH-butyric, 3OH-isovaleric and octanoic acids. Basic chemistries showed persistently low CO2 17mmol/L (21-33mmol/L). Lymphocyte carboxylases performed were normal. Family history revealed an older full sibling who passed away in Mexico at 7 days of age from a cardiomyopathy. Subsequent metabolic studies continued to reveal the same mild abnormalities. Six weeks prior to his death at age 5 years, the patient showed signs of decreasing endurance, worsening balance, strabismus and increasing sleepiness and fatigue. These symptoms increased in severity until the patient presented to the Emergency Room with encephalopathy and inability to walk. Muscle biopsy was performed; the mitochondrial respiratory chain was evaluated (Buffalo, NY) and showed no abnormalities of any of the mitochondrial complexes. The pathology on the muscle biopsy showed a variety of mild abnormalities including pale cytochrome C oxidase staining, mild mitochondrial cytopathy, Z-band streaming, cap-like filamentous inclusions, and mild type I fiber predominance. The patient rapidly progressed to coma, respiratory failure, and passed away. Mitochondrial genome sequencing (GeneDx) revealed a homoplasmic mutation m.8993T>G associated with Neurogenic weakness with ataxia and retinitis pigmentosa (NARP). CONCLUSION Mitochondrial disorders can present with varying abnormalities. Our patient had signs of acidosis, elevations in C3 and C5-OH, not typically associated with NARP. Investigations into conditions causing C3 and C5-OH elevations were negative, leaving us to conclude the C3 and C5-OH elevations are connected to the m.8993T>G mutation. Research to determine the precise biochemical interactions between the mitochondrial effects of the m.8993T>G mutation and the two acylcarnitine abnormalities are needed.

2218M

Evidence of secondary mitochondrial dysfunction in patients with organic acidemias. *G.M. Enns[1], T. Moore[2], A. Le[2], M.K. Shah[1], K. Cusmano-Ozog[1,3], A.K. Niemi[1], T.M. Cowan[1].* 1) Pediatrics, Stanford University, Stanford, CA; 2) Pathology, Stanford University, Stanford, CA; 3) Genetics and Metabolism, CNMC, Washington, DC.

Growing evidence suggests that organic acidemias are accompanied by secondary mitochondrial dysfunction, decreased energy production, and oxidative stress. We previously reported that reduced and oxidized glutathione (GSH and GSSG), as measured in whole blood by LC-MS/MS, are useful biomarkers of mitochondrial dysfunction, and that patients with primary mitochondrial disorders show significant redox imbalance which further worsens during periods of metabolic crisis. Here we extend these findings to the organic acidemias, and provide direct evidence that these disorders are associated with secondary mitochondrial dysfunction. Whole blood GSH and GSSG levels were evaluated in 105 samples from 37 organic acidemia patients (29 methylmalonic acidemia, 5 propionic acidemia, 3 isovaleric acidemia). All patients were evaluated at least once during outpatient clinic visits (i.e., clinically stable), and 7 were also tested while hospitalized for metabolic decompensation. Redox potential was calculated using the Nernst equation, and compared to the mean redox potential from 59 healthy controls (-260 mV ± 6.4). Statistical comparisons were conducted using Student's t test, with significance set at $p<0.05$. As a group, organic acidemia patients showed significant redox imbalance (redox potential -250 mV ± 6.8, $p<0.0001$) with an increased level of oxidation of ~10 mV compared to controls. This was seen in all three organic acidemias studied, with no significant differences noted between the groups. Patients experiencing a metabolic crisis had significantly more oxidized redox potential (-245 mV ± 3.5) than those who were clinically stable (-250 mV ± 6.9; $p=0.036$). These results support the idea that measurements of blood GSH and GSSG may give insights into the contribution of oxidative stress and mitochondrial dysfunction to the pathophysiology of organic acidemias.

2219T

A newborn with persistent mild elevations of succinylacetone in bloodspot, plasma and urine, with an identified homozygous variant in the *GSTZ1* gene. *W. Al-Hertani[1], L. You[1], K. Phommarinh[1], S. Yak[1], S. El Tarazi[3], M-T. Berthier[4], Y. Giguère[4], T. Gagnon[2], D. Cyr[2], P.J. Waters[2].* 1) Department of Medical Genetics, McGill University Health Centre, Montreal, QC, Canada; 2) Biochemical Genetics Laboratory, Medical Genetics Service, Centre hospitalier universitaire de Sherbrooke (CHUS), Sherbrooke, QC, Canada; 3) Department of Pediatrics, McGill University Health Centre, Montreal, QC, Canada; 4) Quebec Newborn Bloodspot Screening Program, Service de Biochimie, CHU de Québec, Quebec City, QC, Canada.

Case report: The proband is a Native Canadian female, born to non-consanguineous parents. At age 18 days, she was referred as screen-positive for tyrosinemia, following slightly elevated succinylacetone (SUAC) in bloodspot (age 40 hours): hydrazine method, 1.19 micromol/l, cutoff 0.67; enzymatic method 3.53, cutoff 2.50. Meanwhile, a urine organic acid profile at age 6 days (test requested independently, for unrelated reasons) revealed mild SUAC elevation. SUAC results obtained using stable isotope dilution GC-MS on samples collected between age 4 days and 4 weeks were: in plasma, 1282, 689, 617 nanomol/L (reference < 24); in urine, 776, 766, 604 micromol/mol creatinine (reference < 34). While unambiguously elevated, and persistent, these values were approximately 1000-fold lower than typical of patients with tyrosinemia type 1. Tyrosine and phenylalanine, normal on the newborn bloodspot, remained normal in plasma during follow-up, as did the remainder of the plasma amino acid profile. At age 2 months, she is doing well clinically. Initial minimal elevations of transaminases have been resolving spontaneously. Abdominal ultrasound and ophthalmological examination were within normal limits, as were coagulation factors. Sequencing of the *FAH* gene did not show any deleterious mutations or variants. Considering the clinical course together with biochemical and molecular findings, a diagnosis of tyrosinemia type 1 is highly unlikely. Despite no reported human phenotype in the literature, we considered the possibility that reduced activity of maleylacetoacetate isomerase, due to hypomorphic variants in the *GSTZ1/MAAI* gene (J Clin Pharmacol 2012;52:837-849), as the most likely potential etiology for the persistent elevation of SUAC in this infant. Interestingly, molecular analysis in our patient, did reveal a homozygous variant of uncertain significance in exon 7 of the *GSTZ1* gene; c.449C>T (p.Ala150Val), predicted to be pathogenic by SIFT and Mutation Taster. Parental analysis is currently under way.

2220M

Novel mutations in human lipase A gene. *J. Huang, Z. Wu, C.R. Scott.* Dept Pediatrics, Univ Washington, Seattle, WA.

Lysosomal acid lipase (LAL) is a lysosomal enzyme that is involved in intracellular lipid metabolism. Deficiency of the LAL enzyme results in accumulation of cholesterol esters and triglycerides in most tissues of the body, and causes Wolman disease and cholesteryl ester storage disease (CESD). Wolman disease is fatal within the first year of life due to severe hepatomegaly, persistent diarrhea and failure to thrive, while CESD is a milder disease that is characterized by hyperlipidemia and hepatomegaly that can be observed in childhood or develop in adulthood. LAL deficiency is an inherited genetic condition, caused by mutation in the lipase A (LIPA) gene that encodes the LAL enzyme. Spanning about 40 kb with 10 exons and coding 400 amino acids, the LIPA gene is located on the long arm of chromosome 10 between positions 23.2 and 23.3. So far more than 40 mutations in the LIPA gene have been reported, including missense, nonsense, splice site, small deletion/insertions and large deletions. We have examined twenty specimens worldwide with LAL deficiency for mutation analysis of the LIPA gene, performing bi-directional sequence analysis of coding exons and corresponding intron/exon boundaries on genomic DNA. Among mutations we identified, the common mutation c.894 G>A is about 20% of alleles; and four novel mutations have never been reported previously. These new mutations include: (1) missense mutation c.67G→A (p.G23R), (2) missense mutation c.266T→C (p.L89P), (3) small deletion TGCCAACAGCAGCCT in c.297-311, and (4) missense mutation c.421G→C (p.A141P). Interestingly, the mutation c.421G→C has only been detected in specimens from persons of Asian ethnicity.

2221T

Inherited GPI anchor deficiency: biochemical, molecular, and clinical presentation of a patient with *PIGW* mutations. *T. Chiyonobu[1], N. Inoue[2], M. Morimoto[1], T. Kinoshita[3], Y. Murakami[3].* 1) Pediatrics, Kyoto Prefectural University of Medicine, Kyoto, Japan; 2) Molecular Genetics, Osaka Medical Center for Cancer, Osaka, Japan; 3) Immunoregulation, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan.

Background: Glycosylphosphatidylinositol (GPI) is a glycolipid that anchors 150 or more kinds of proteins to the human cell surface. Recently, inherited GPI anchor deficiencies (IGDs) were reported which cause developmental delay often accompanied by epilepsy, facial dysmorphism and multiple anomalies. There are 26 known genes involved in the biosynthesis and remodeling of GPI anchored proteins (GPI-APs). IGDs due to germline mutations in nine of these genes (*PIGA*, *PIGM*, *PIGN*, *PIGV*, *PIGL*, *PIGO*, *PIGT*, *PGAP2* and *PGAP3*) have been reported so far. Additionally, elevated levels of serum alkaline phosphatase (ALP), one of the GPI-APs, is commonly seen in IGDs caused by mutations in *PIGV*, *PIGO*, *PGAP2*, and *PGAP3*, and these deficiencies are termed hyperphosphatasia with mental retardation syndrome (HPMRS). Here we report the first patient with mutations in *PIGW*, which is involved in the addition of the acyl-chain to inositol in an early step of GPI biosynthesis. Patient and Results: The patient was the second child from healthy non-consanguineous Japanese parents. Severe developmental delay was evident from early infancy. At 7 months old, he developed clusters of tonic spasms with hypsarrhythmic pattern in EEG and was diagnosed as having West syndrome, a type of infantile epileptic encephalopathy. Brain MRI disclosed symmetric lesions in the basal ganglia and brainstem. Laboratory investigations including metabolic disorder screening were normal but the serum ALP was constantly elevated at around 2000 U/L. Because he showed severe developmental delay with dysmorphic facial features (broad nasal bridge and tented upper lip) and hyperphosphatasia, characteristics often seen in IGDs, we tested for GPI deficiency. Flow cytometry using blood granulocytes from the patient revealed reduced surface expression of GPI-APs. He was identified to be compound heterozygous for Thr71Pro and Met167Val mutations in *PIGW* by targeted sequencing. Functional analysis using the PIGW deficient CHO cells revealed that both mutations affect the PIGW activity. Conclusions: Hypomorphic mutations in *PIGW* cause infantile epileptic encephalopathy and HPMRS. The characteristic findings in brain MRI may be a diagnostic clue, and flow cytometric analysis of blood cells is effective in screening IGD.

2222M

Structural modeling and bioinformatics analysis of Acid Alpha-Glucosidase in Colombian patients with p.[G576S; E689K] polymorphism associated with p.W746C mutation. *J. Gonzalez Santos[1], LP. Barragán-Osorio[2], S. Bustos-Hémer[2], P. Ayala[2], R. García-Robles[3,4], F. Suarez-Obando[2].* 1) Nutrición y bioquímica, Pontificia Universidad Javeriana, Bogota, Colombia; 2) Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogotá D.C., Colombia; 3) Instituto de Investigación en nutrición, genética y metabolismo. Universidad del Bosque, Bogotá, Colombia; 4) Genzyme de Colombia S.A.

Introduction: Pompe disease is caused by pathogenic sequence variations in the Acid α-Glucosidase gene (GAA) causing lysosome glycogen accumulation. SNPs c.1726G>A (p.G576S) and c.2065G>A (p.E689K) were predicted to cause pseudodeficiency of GAA. Substitution p.E689K reduces GAA activity by 50%; at most. Substitution p.G576S reduces GAA activity to clinical spectrum ranges. Objective: to assess the effects of p.[G576S; E689K] and their association with mutation p.W746C found in 2 Colombian patients. Methods: Three-dimensional model of human GAA based on the N-terminal Subunit of Human Maltase-Glucoamylase complex with Acarbose (MGA PDB 2QMJ) template. Energy minimization was performed. Physicochemical properties were validated with ProtParam. Quality/Reliability of stereo chemical parameters of the homology model was confirmed with Qmean6, GROMOS and Ramachandran plot. Alignments were calculated with ClustalW and Muscle. Conserved amino acids were obtained as a consensus of MultiDisp, Consurf and ConSeq. Solvent accessible surface area value (ASA) was determined with PYMOL. The mutant model was superimposed on the wild type structure to determine the influence of amino acid replacement in the 3D structure on the basis of the total root-mean-square distance (RMSD). Results: Amino acid identity between GAA and the template was 44%. Both proteins belong to the glycosyl hydrolases family 31 in which the structure is conserved and shares similar binding site, overcoming the low sequence identity. GAA and GMA are hydrophilic with similar composition of negative and positive residues. Stereo-chemical quality of the predicted model showed 85.7%; of residues in favored regions, 8.4%; in allowed regions and 5.9%; in outlier regions. The conservation analysis recognized a highly conserved amino acid position G576, in a lesser extent to E689. The preliminary ASA measurement indicated that both polymorphisms and W7466C mutations are buried. Wild type predicted model value, significantly differs from the mutant model (27125.766 $Å^2$ and 30854.953 $Å^2$, respectively). The Superposition of the two models reveals that G576 is located inside the barrel structure and G576S substitution change the side chains of the Aα4-helix. E689 is located at N-terminus of Aα8-helix where the side chain is fully exposed. The superposition RMSD value was 0.84 Å.

2223T

Familial hypermangenesemia among Egyptian families: further delineation of the phenotype and management. *M.S. Zaki[1], MS. Abdel-hamid[2], H. Hossni[3], MY. Issa[1], A. El-Safty[4], A. Oraby[5].* 1) Clinical Genetics Department, National Research Centre , Cairo, Egypt, Dokki, Cairo, Egypt; 2) Human Molecular Genetics Department, National Research Centre, Cairo, Egypt; 3) National Institute of Neuromotor System, Cairo, Egypt; 4) Toxicology Department, Cairo University, Cairo, Egypt; 5) Pediatrics Department, Cairo University, Cairo, Egypt.

Familial hypermanganesemia or inherited manganism is a rare metabolic error caused by recessive mutations in SLC30A10 gene. This gene is functioning to maintain homeostatic control of intestinal absorption and biliary excretion of manganese through responsible for manganese transporter. Manganese accumulates in the liver, muscle, bloodstream, and brain, specifically the basal ganglia resulting in extrapyramidal manifestations, polycythemia and liver cirrhosis. We describe 7 patients from 3 unrelated consanguineous families with cardinal manifestations of manganism. They were five males and 2 females and their aged ranged from 1 4/12y to 18 years old. Six patients (Family 1(4 sibs), 2 (2 sibs)) presented with slowly regression of walking around the age of 2 9/12, manifested as recurrent falling due to limb dystonia. Interestingly, one patient (Family 3) had sudden acute dystonia of lower limbs with inability to walk at the age of 1 year and 5 months with suspicious of inflammatory etiology but the weak response to intravenous immunoglobulin together with the striking neuroimaging pointed to this metabolic error. All affecteds had polycythemia with hemoglobin concentration ranged from 17 -21 g/dl and blood manganese level was above 2000 nmol/L. Elevated liver enzymes with no sign of cirrhosis were in the older 2 sibs in family 1 (17 and 15 ys old), however normal enzymes were in the older affected in family 2 in spite of reaching 18 years old and was still ambulant. Brain MRI showed the typical picture of manganese deposition in basal ganglia classically on T1-weighted images, with affection of midbrain and the tegmentum of the pons as well as the middle cerebellar peduncle with no corresponding abnormality on T2-weighted scans. Molecular analysis for these 3 families revealed novel mutations in SLC30A10 gene in 2 of them. We designated oral treatment with 2,3 dimercaptosuccinic acid (DMSA), iron supplementation and levodopa that showed satisfactory preliminary results in chelating manganese with remarkable improvement of biochemical markers in all and alleviation of clinical manifestations in the younger probands. This report represents further delineation of familial hypermanganesemia. It emphasizes the importance of considering this treatable metabolic error in any patient presenting with acute or insidious dystonia and polycythemia.

2224M

Abnormal phospholipid metabolism due to variants in *PCYT1A* causes spondylometaphyseal dysplasia with cone-rod dystrophy. *J. Jurgens[1], N. Sobreira[1], J. Hoover-Fong[1], P. Modaff[2], G. Yamamoto[3], W. Baratela[3], D. Bertola[3], F. Collins[4], J. Christodoulou[4], M.B. Bober[5], R. Pauli[2], D. Valle[1].* 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Pediatrics, University of Wisconsin-Madison, Madison, WI; 3) Unidade de Genética, Instituto da Criança, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil; 4) Western Sydney Genetics Program, Children's Hospital at Westmead, Sydney and Disciplines of Pediatrics & Child Health and Genetic Medicine, Sydney Medical School, University of Sydney, New South Wales, Australia; 5) Division of Medical Genetics, Department of Pediatrics, A.I. duPont Hospital for Children, Wilmington, DE.

Spondylometaphyseal dysplasia with cone-rod dystrophy (SMD-CRD) is a rare, autosomal recessive condition characterized by short stature, bowing and metaphyseal flaring of the long bones, rhizomelic shortening, platyspondyly, scoliosis, and early-onset progressive degeneration of cone and rod photoreceptors. By whole exome sequencing and Sanger sequencing, Hoover-Fong et al. (2014) identified novel homozygous or compound heterozygous variants in *PCYT1A* in 6 unrelated probands. Independently, Yamamoto et al. (2014) described two novel homozygous *PCYT1A* variants, p.Glu129Lys and p.Ser323Argfs*38, in two unrelated SMD-CRD probands. Here, we used Sanger sequencing to identify homozygous or compound heterozygous variants in *PCYT1A* in four additional unrelated individuals with SMD-CRD. We detected a single homozygous variant (p.Ala99Val) in 2 unrelated probands and found a separate homozygous variant (p.Tyr240His) in a third unrelated proband. In the fourth proband, we detected the compound heterozygous variants p.Ser323Argfs*38/p.S114T. *PCYT1A* encodes CTP:phosphocholine cytidydyltransferase α (CCTα), an enzyme which catalyzes the rate-limiting step in de novo phosphatidylcholine biosynthesis by the Kennedy pathway. Phosphatidylcholine is the predominant membrane phospholipid in mammalian cells. Western blot analysis of CCTα expression in cultured skin fibroblasts of 3 unrelated individuals with SMD-CRD shows that the protein levels vary among SMD-CRD patients. When compared to controls, p.Ala99Val homozygotes show a 30% reduction in protein levels; p.Glu129Lys homozygotes show a 70% reduction; and patients homozygous for a 1-bp frameshifting insertion (p.Ser323Argfs*38) show no protein levels. Preliminary northern blot analysis of RNA isolated from the same fibroblast cells suggests that phosphatidylethanolamine N-methyltransferase (*PEMT*) mRNA levels are increased in patients as compared to controls. Based on these results, we hypothesize that patient cells may be compensating for their deficiency in phosphatidylcholine synthesis by upregulating the PEMT pathway, which converts phosphatidylethanolamine into phosphatidylcholine. We suggest that abnormal phospholipid metabolism is a largely unexplored cause of progressive retinal degeneration with or without accompanying skeletal abnormalities.

2225T

Mutation spectrum of six genes in Chinese phenylketonuria patients obtained through next-generation sequencing. *Y. Gu[1], K. Lu[2], G. Yang[2], Z. Cen[2], Y. Li[2], L. Lin[2], J. Hao[2], Z. Yang[2], J. Peng[2], S. Cui[2], J. Huang[2].* 1) Lianyungang Maternal and Child Health Hospital, Lianyungang, China; 2) Chinese National Human Genome Center at Shanghai Shanghai, CHINA.

Background: The identification of gene variants plays an important role in the diagnosis of genetic diseases. Methodology/Principal Findings: To develop a rapid method for the diagnosis of phenylketonuria (PKU) and tetrahydrobiopterin (BH4) deficiency, we designed a multiplex, PCR-based primer panel to amplify all the exons and flanking regions (50 bp average) of six PKU-associated genes (PAH, PTS, GCH1, QDPR, PCBD1 and GFRP). The Ion Torrent Personal Genome Machine (PGM) System was used to detect mutations in all the exons of these six genes. We tested 93 DNA samples from blood specimens from 35 patients and their parents (32 families) and 26 normal adult individuals. Using strict bioinformatic criteria, this sequencing data provided, on average, 99.14% coverage of the 108 exons at more than 70-fold depth. We found 23 well-documented variants in the PAH gene and six novel mutations in the PAH and PTS genes. A detailed analysis of the mutation spectrum of these patients is described in this study. Conclusions/Significance: These results were confirmed by Sanger sequencing. In conclusion, benchtop next-generation sequencing technology can be used to detect mutations in monogenic diseases and can detect both point mutations and indels with high sensitivity, fidelity and throughput at a lower cost than conventional methods in clinical applications.

2226M

Functional Characterization of RYR1 Sequence Variants Associated with Malignant Hyperthermia Susceptibility Identified through Exome Sequencing. *S.G. Gonsalves[1, 2], C.E. Kasper[2], S. Perry[2], S.M. Muldoon[3], L.G. Biesecker[1].* 1) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; 2) Uniformed Services University of the Health Sciences, Graduate School of Nursing, Bethesda, Maryland, USA; 3) Uniformed Services University of the Health Sciences, School of Medicine, Department of Anesthesiology, Bethesda, Maryland, USA.

Malignant hyperthermia susceptibility (MHS) is a life-threatening, inherited disorder of muscle calcium metabolism, triggered by anesthetics and depolarizing muscle relaxants. Exome sequence (ES) data from ClinSeq® were screened to identify putative pathogenic MHS variants in the RYR1 and CACNA1S genes. To study the effects of RYR1 variants on channel function, we selected benign and pathogenic variants for comparison—using an in-vitro assay measuring calcium release from participant-derived Epstein-Barr virus immortalized B-lymphocytes—to determine if incidental MHS-associated variants identified from ES showed abnormal calcium release from lymphoblasts. Annotation of 870 exomes for RYR1 and CACNA1S variants used an algorithm that filtered results based on genotype quality, allele frequency, mutation type, and information from mutation databases. Sixty-three RYR1 and 41 CACNA1S variants passed the quality and frequency metrics and were scored on a four-point pathogenicity scale (2-5). Three RYR1 class 5 (p.RArg614C, p.S1728F, p.D3986E) and one class 3 (variant of uncertain significance) p.R1667C, were selected for functional testing. Pharmacological sensitivity of the RyR receptor type-1 in EBV-immortalized B-lymphocytes derived from four ClinSeq® participants were tested and compared to MH-negative controls. RYR1 p.R614C (with published functional data) was selected as a positive control. B-lymphoblastoid cell lines were used to study the transient peak calcium release induced by the ryanodine receptor agonist 4-chloro-m-cresol (4-CMC). Calcium release (area under curve) was averaged from 12 replicates. Differences in calcium release between cell lines were analyzed by t-test. EBV-lymphocytes from class 5 variants showed increased sensitivity of calcium release to 4-CmC compared to normal controls. Elevated calcium release was observed in RYR1 p.R614C and p.D3986E, (*$P<0.05$) at progressive doses (0.4 mM, 0.6 mM and 1.0 mM) of 4-CMC. RYR1 S1728F showed increased levels compared to wild type (WT), but did not achieve statistical significance. RYR1 p.R1667C was similar to WT. Two class 5 MHS variants showed significantly altered calcium release. Although this assay is time consuming and impractical for large-scale screening of putative MH-susceptibility variants, variant RYR1 function can be successfully assayed using a biochemical assay on participant-derived EBV-lymphocytes, without resorting to muscle biopsy for ex-vivo contraction testing.

2227T

Screening for Lysosomal Storage Disorders using an integrated enzymatic and molecular approach. *O. Bodamer[1], G. Ghaffari[1], P. Noriega[2], C. Hung[1], B. Illagan[1], A. Hosseini[1], B. Johnson[1].* 1) University of Miami, Miami, FL; 2) Genzyme-Sanofi Inc, Mexico City, Mexico.

Background: Lysosomal storage disorders (LSDs) comprise a heterogeneous group of genetic conditions that affect lysosomal metabolism. The clinical phenotype typically follows a clinical continuum of progressive multi-organ involvement. Effective therapies including enzyme replacement therapy (ERT) are available for a subset of LSDs including Fabry, Pompe, Gaucher diseases, MPS I and MPS II. Additional therapies are under development. Timely diagnosis and initiation of ERT is vital to reduce and/or prevent disease morbidity. Methods: We have validated an integrated diagnostic approach utilizing enzyme testing and second tier molecular testing on dried blood spots (DBS) for 6 LSDs including Fabry, Pompe, Gaucher, Niemann Pick Type A/B, Krabbe diseases and MPS I. Samples were collected from at-risk patients based on clinical phenotype suspicious for LSD from larger tertiary referral centers in Mexico. Enzyme analysis in DBS was done by tandem mass spectrometry followed by direct sequencing in samples with low enzyme activities. Results: 3206 samples were received for enzyme testing and 820 samples for direct sequencing of the GLA gene in females. 223 samples (7%) were reflexed to molecular testing based on low enzyme activities. 3/22 (14%) of samples were diagnostic for Pompe disease, 4/19 (21%) were diagnostic for Gaucher disease, 6/11 (55%) were positive for MPS I, 9/171 (5.2%) were diagnostic for Fabry disease and 12/820 (1.4%) of females tested for Fabry disease carried a heterozygous GLA mutation. Conclusion: The use of an integrated diagnostic approach for the above LSDs in DBS allows timely and inexpensive diagnosis of at-risk patients. The refinement of the clinic phenotype that prompts testing for LSDs will lead to further reduction of unnecessary testing.

2228M

Lysosomal acid lipase activity in dried blood spots from patients initially suspected of Gaucher disease. *V.G. Pereira[1], C.F. Chaves[1], J.U.S. Yamamoto[1], A.M. Martins[2], V. D'Almeida[1].* 1) Department of Psychobiology, Universidade Federal de São Paulo, São Paulo, Brazil; 2) Department of Pediatrics, Universidade Federal de São Paulo, São Paulo, Brazil.

Lysosomal acid lipase (LAL) deficiency is an autosomal recessive condition, which leads to the accumulation of cholesteryl esters and triglycerides in most tissues, causing two different lysosomal storage disorders: Wolman disease (WD) and Cholesterol ester storage disease (CESD). WD is the most severe form of the disease, leading to death before one year of age, while CESD presents a more attenuated clinical course, with highly variable phenotypes. Hepatic alterations are a common symptom between lysosomal storage disorders, mainly in CESD, Gaucher and Niemman-Pick diseases. Therefore, differential diagnosis is essential for proper treatment. In this sense, patients initially suspected of Gaucher disease, whose biochemical diagnosis had not been confirmed, may represent a high-risk population for LAL deficiency (LALD). The aim of this study was to determine LAL activity in dried blood spots (DBS) referred to Laboratório de Erros Inatos do Metabolismo/UNIFESP - Brazil with an initial suspicion of Gaucher disease, which had beta-glicosidase activity within the normal range (Gaucher-negative), as a screening for LALD patients. LAL activity was determined by a fluorimetric assay. Until now, 110 Gaucher-negative samples were analyzed and mean LAL activity observed was 45.28 pmol/punch/h (standard error = 4.17), median = 33.99 pmol/punch/h, ranging from 0 to 299 pmol/punch/h. From these samples, 38 (~35%) of them presented LAL activity below the normal range, which had been previously determined in our laboratory (> 24 pmol/punch/h). These samples were then analyzed separately; in this group, mean LAL activity was 12.82 pmol/punch/h (standard error = 1.23), median = 12.97 pmol/punch/h, ranging from 0 to 23.91 pmol/punch/h. All the physicians responsible for patients in the group with low LAL activity were contacted and a new DBS was requested to confirm the diagnosis, to exclude the possible interference of reduced LAL activity due to long storage periods (all DBS samples used in this study were kept at 4°C through a maximum period of one year). Although we still have not received these new DBS to confirm LALD, the high rate of LAL activity below the normal range found in our study suggests that Gaucher-negative samples indeed represent a high-risk population to screen for LALD patients. In addition, Gaucher-negative patients should also have LAL activity determined as differential diagnosis. Funding: FAPESP, CAPES, IGEIM and AFIP.

2229T

Novel strategy for the diagnosis of late onset Pompe disease using next generation sequencing technologies. *S. Levesque[1], E. Gravel[1], S.L. Austin[2], S. Gravel[1], J. Keutzer[3], C. Auray-Blais[1], P.S. Kishnani[2].* 1) Department of Pediatrics, Division of Medical Genetics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada; 2) Departments of Pediatrics, Division of Medical Genetics, Duke University Medical Center, Durham, NC, USA; 3) Genzyme Corporation, a Sanofi company, Cambridge, MA, USA.

Background: Pompe disease is a rare lysosomal storage disorder caused by acid alpha-glucosidase enzyme deficiency, and is characterized by progressive lysosomal glycogen accumulation especially in muscles. Studies have shown that early diagnosis followed with prompt initiation of enzyme replacement therapy may maximize clinical benefits. However, early diagnosis remains challenging, especially in the actual adult cohort owing to the lack of universal screening. Despite published guidelines to educate clinicians, testing for Pompe is often overlooked owing to its low frequency. Next generation sequencing (NGS) might solve this issue given that it can analyse at once a large number of genes causing diseases with overlapping phenotypes. **Method:** We designed a NGS assay in order to analyze the coding sequences and splice site junctions of *GAA*, mutated in Pompe disease, along with 78 genes causing neuromuscular disorders with overlapping phenotypes. Analytical validity was then determined using DNA from 20 Pompe patients with known variations and the reference cell line NA12878. Enrichment of target sequences was performed using a custom in-solution oligonucleotide probes library (SeqCapEz®, Nimblegen), followed by NGS on the Illumina MiSeq® platform. Variations were detected with our bioinformatics pipeline based on BWA and GATK. **Result:** We obtained a median coverage of ~90X (sequences per base) per sample, with only 0.3% of all exons showing less than 20X. All *GAA* exons were successfully covered with >20X. The NGS assay showed a sensitivity of 100% (95% C.I. 98-100%) for the *GAA* gene, based on the analysis of 197 known single nucleotide variations and indels in Pompe patients (42 different variations). In addition, recurrent deletion of exon 18 was correctly detected by read depth analysis in 5/5 Pompe patients. Sensitivity of the assay across all selected genes was similar with 100% (95% C.I. 99%-100%), as determined by comparison between observed variations and 373 expected variations in NA12878. Specificity was determined to 98% (95% C.I. 95-99%), for 208 variations found in exons and splice site junctions of selected genes in the NA12878 cell line. **Conclusion:** We were successful in designing a highly sensitive and specific NGS assay for Pompe disease and related genetic disorders. NGS of patients presenting with a neuromuscular disorder of unknown etiology from neuromuscular clinics is in progress and results will be presented.

2230M

Molecular testing of 163 patients with Morquio A syndrome (mucopolysaccharidosis IVA; MPS IVA): 39 novel *GALNS* gene mutations. *M. Hegde[1], A. Morrone[2,3], K.L. Tylee[4], M. Al-Sayed[5], A.C. Brusius-Facchin[6], A. Caciotti[2], H.J. Church[4], M.J. Coll[7], K. Davidson[13], M.J. Fietz[8], L. Gort[7], F. Kubaski[6], L. Lacerda[9], F. Laranjeira[9], S. Leistner-Segal[6], S. Mooney[10], S. Pajares[7], L. Pollard[11], I. Ribeiro[9], R.Y. Wang[12], N. Miller[13].* 1) Emory Genetics Lab, Emory Univ Sch Med, Atlanta, GA, USA; 2) Molec and Cell Biol Lab, Pediatric Neurology Unit and Lab, Meyer Child Hosp, Florence, Italy; 3) Dept of Neurosciences, Psychology, Pharmacology and Child Health, Univ of Florence, Florence, Italy; 4) Willink Biochem Genetics, Central Manchester Univ Hosp NHS Foundation Trust, Manchester, UK; 5) Dept of Medical Genetics, King Faisal Specialist Hosp and Research Center, Riyadh, Saudi Arabia; 6) Lab de Genética Molecular, Serviço de Genética Médica, Hosp de Clínicas de Porto Alegre (HCPA), Porto Alegre, Brazil; 7) Sección de Errores Congénitos del Metabolismo-IBC, Servicio de Bioquímica y Genética Molecular, Hospital Clínic, CIBERER, IDIBAPS, Barcelona, Spain; 8) SA Pathology, Women's and Child Hosp, North Adelaide SA, Australia; 9) Unidade de Bioquímica Genética, Centro de Genética Médica Jacinto Magalhães (CGMJM) do Centro Hosp do Porto (CHP), Porto, Portugal; 10) The Buck Inst for Research on Aging, Novato, CA, USA; 11) Biochem Genetics Lab, Greenwood Genetic Center, Greenwood, SC, USA; 12) Child Hosp of Orange County, CA, USA; 13) BioMarin Pharmaceutical Inc., Novato, CA, USA.

Morquio A syndrome is an autosomal recessive lysosomal storage disorder caused by deficiency of N-acetylgalactosamine-6-sulfatase (GALNS) encoded by the *GALNS* gene. Mutations occur throughout *GALNS* and often are identified only in one patient/family, complicating mutation detection and interpretation. 7 laboratories reported mutations in *GALNS* from Morquio A patients and 2 clinicians directly reported molecular results. Individuals received a diagnosis of Morquio A prior to and independent of molecular results. *GALNS* mutations were checked against reported SNPs. Molecular testing of 163 Morquio A patients identified 99 separate changes, of which 39 are previously unpublished, novel *GALNS* mutations associated with Morquio A: p.Val16Glu, p.Leu36Arg, p.Asp40Asn, p.Val48Gly, p.Glu51Lys, p.Pro81Leu, p.Ala84Glu, p.Leu91Pro, p.Gly116Val, p.His145Tyr, p.Phe156-Leu, p.His166Arg, p.Gly201Glu, p.Leu214Pro, p.Phe216Ser, p.Thr235Lys, p.Ser264Thr, p.Asn289Asp, p.Arg380Gly, p.Gly415Val, p.Ile416Thr, p.Pro420Arg, p.Ala492Thr, p.Gly500Ser, p.Cys507Phe, p.Glu126Ter, p.Trp141Ter, p.Tyr209Ter, p.Arg251Ter, p.Pro357ArgfsTer21, p.Leu372SerfsTer6, p.Tyr385Ter, p.Gln414Ter, p.Val427SerfsTer13, p.Glu477_Gln485del, Complex del-dup (duplication from intron 5-intron 9, deletion in intron 10), c.120+1g>c, c.405_422+1del, c.758+4a>t. We also identified 26 SNPs. Reporting Morquio A-associated *GALNS* mutations improves genetic counseling and diagnosis capability. We recommend also to molecularly test both parents. While molecular testing provides useful diagnostic and genetic counseling information, enzyme activity testing of GALNS, along with other enzymes, remains the standard for diagnosis of Morquio A.

2231T

Menkes disease: the importance of precise diagnosis with molecular analysis in neonatal period. *C. Kim[1], L. Costa[1], S. Pegler[1], R. Lellis[2], V. Krebs[1], T. Morgan[3], R. Honjo[1], D. Bertola[1].* 1) Dept Pediatrics, Inst da Crianca, Hospital das Clínicas - Universidade de São Paulo - Sao Paulo, SP, Brazil; 2) Dept. Pathology - Hospital das Clínicas - Universidade de São Paulo - São Paulo - Brazil; 3) Department of Womens and Children's Health, Dunedin School of Medicine, University of Otago, Dunedin NEW ZEALAND.

Menkes disease (MD) is a congenital disorder in copper metabolism, caused by mutations in ATP7A. It is characterized by physical and neurological alterations. In neonatal period these alterations can be nonspecific, what makes the precocious diagnosis a challenge. Clinical includes: developmental delay, hypotonia, seizures, hair abnormalities and connective tissue disorders. Diagnosis can be suspected when there are low levels of ceruloplasmine and serum copper, but they can be in low levels even in normal newborns. Molecular analysis confirms the diagnosis. Precocious treatment is mandatory with copper histidine for clinical improvement. We report a familial case of MD with molecular confirmation. The proband was the second child of a healthy mother, with a history of eye twitching since 1mo2d.The physical exam of our proband showed: weight of 4080g (15%), length 52.8 cm (15%) and OFC of 40cm (> 97%), steely, kinky, sparse hair, divergent strabismus, pudgy cheeks. The mother had the first son, from different partner, who died at 3mo21d with a clinical diagnosis of MD. Laboratorial exams showed low serum ceruloplasmin <2mg/dL (reference range from 20-60 mg/dL) and copper <8 μg/dL (reference range from 70-160 μg/dL) confirming clinical diagnosis of MD. Treatment with copper histidine was indicated immediately but it was instituted only at 2mo27d. The reason for delay was the time demanded for the preparation. He did not presented clinical or neurology improvements and died at 6mo. Mother had another pregnancy (with the same partner), a male fetus was identified and copper histidine was manufactured. He was born healthy, biochemical markers were diminished and treatment was instituted since 3d and he presented no neurological symptoms. Molecular analysis was performed the proband was hemizygous for a mutation in ATP7A (c.1668-1680del, predicted to lead to the frameshift p.Ile556Metfs*11). Mother is heterozygous for the same mutation. The third son did not have the mutation and consequently treatment was discontinued. We support the clinical relevance of molecular confirmation for diagnosis and genetic counseling, once clinical findings in the neonatal period are nonspecific and early treatment must be indicated. Molecular confirmation in our case was indispensable for the correct diagnosis of MD, once biochemical markers are nonspecific of copper deficiency.

2232M

Mitochondrial Heteroplasmy and Clinical Variability in a MELAS Family. *K. McClelland[1, 2], K. Weisiger[2], T. Huang[3], J. Youngblom[1], C. Lee[2], N. Barasa[3], C.A. Valencia[3], S. Packman[2].* 1) California State University, Stanislaus; 2) University of California, San Francisco; 3) University of Cincinnati.

Genetic counseling and prognosis in patients with mitochondrial disorders is challenging due to highly variable expression and penetrance. One source of variability is the existence of heteroplasmy; the co-occurrence of wild-type and mutant mitochondrial DNA [mtDNA] within a cell. Heteroplasmy levels can differ both between individuals and between different tissue types in the same individual. Manifestation of pathology in a given tissue, is, in fact considered to depend on that tissue's threshold for mtDNA mutation load. While a relationship between mutation load and disease expression has been shown in in vitro studies, current literature is conflicting on the utilization of heteroplasmy for predictive and prognostic counseling. In the present work, we asked whether there was a correlation between mutation load, as measured in urine sediment cells, and clinical manifestations in an extended Hispanic family with the mA3243G MELAS [mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes] mutation. This study examined 23 maternal relatives of a patient diagnosed with MELAS for the presence of clinical symptoms, physical findings and degree of heteroplasmy in urine sediment cells. For the molecular analyses and assessment of percent heteroplasmy, mitochondrial DNA was isolated from cells in urine samples and the mA3243G mtDNA mutation was detected by real time PCR. Presence and number of clinical symptoms [patient reports] and physical findings [examination by SP and CL] were compared to percentage of mtDNA A3243G mutation detected in urine sediment cells for each individual. This study found that 65% of family members self-reported MELAS-associated symptoms, and 52% had physical findings, both neurologic and other systemic. A significant correlation was found between age and number of MELAS-associated symptoms, as well as age and number of physical findings. In contrast to a number of reported studies, no significant correlation was found between levels of mutation load measured in urine and the number of symptoms or physical findings. We conclude that urine sediment cell heteroplasmy analysis is not a uniformly effective test to predict clinical phenotype in individuals, or small sample sizes that carry the mA3243G mutation.

2233T

Morquio A syndrome (mucopolysaccharidosis IVA; MPS IVA): A review of 277 gene mutations curated in a new *GALNS* locus-specific database. *N. Miller[1], A. Morrone[2,3], A. Caciotti[2,3], R. Atwood[4], K. Davidson[1], C. Du[4], P. Francis-Lyon[5], P. Harmatz[6], M. Mealiffe[1], S. Mooney[4], T.R. Oron[4], A. Ryles[5], K.A. Zawadzki[7].* 1) BioMarin Pharmaceutical Inc., Novato, CA, USA; 2) Molec and Cell Biol Lab, Pediatric Neurology Unit and Laboratories, Meyer Child Hosp, Florence, Italy; 3) Dept of Neurosciences, Psychology, Pharmacology and Child Health, Univ of Florence, Florence, Italy; 4) The Buck Institute for Research on Aging, Novato, CA, USA; 5) Dept of Computer Science, Univ of San Francisco, San Francisco, CA, USA; 6) UCSF Benioff Children's Hospital Oakland, Oakland, CA, USA; 7) Health Interactions Inc., San Francisco, CA, USA.

Morquio A is caused by deficient activity of N-acetylgalactosamine-6-sulfatase (GALNS) resulting from mutations in *GALNS. GALNS* mutations are numerous and heterogeneous. To aid detection and interpretation of mutations, we summarize published mutations from 541 Morquio A patients, together with 81 published mutations not described as genotypes from individual Morquio A patients, and report a new public-access *GALNS* locus-specific database. 277 unique *GALNS* mutations were identified from 1091 alleles. Most alleles reported from patients (79%) are missense. Even the most frequent mutations are uncommon. The three most common alleles (R386C, I113F, G301C) together only represent 14% of alleles. Significant geographical and/or ethnic origin-based allele frequency variability exists. 11% of alleles are insertions/deletions, 6% intronic, 4% nonsense. Most Morquio A-associated mutations have only been reported 1-2 times. 48% of patients are homozygous for a *GALNS* mutation, 39% heterozygous, and 13% have only one mutation detected. Mutation detection and genotype-phenotype challenges are in part due to the heterogeneity of *GALNS* mutations and lack of multiple families with the same mutations. Parental testing is encouraged. Reporting new alleles facilitates distinguishing pathogenic from benign mutations. The standard for Morquio A diagnosis is still deficient GALNS enzyme activity measured in leukocytes or fibroblasts, together with normal control enzyme activities.

2234M

Expanding the toolbox for the diagnosis of mitochondrial disorders. *L. Kremer[1,2], G. Pichler[3], T. Schwarzmayr[1,2], T. Haack[1,2], T. Wieland[1,2], T. Strom[1,2], M. Mann[3], T. Meitinger[1,2], H. Prokisch[1,2].* 1) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 2) Institute of Human Genetics, Technische Universität München, Munich, Germany; 3) Max Planck Institute of Biochemistry, Martinsried, Germany.

Despite good progress in the field, many disease causing mutations in patients with mitochondrial disorders still have to be identified. We recently applied whole exome sequencing (WES) in 400 unrelated individuals with juvenile-onset mitochondrial disorder. In 50% of the patients we were unable to identify the disease causing mutation and hence additional tools are needed to promote the identification of the disease gene for the remaining 50% of cases. Sequencing of trios will expand the screening for de novo mutations, while whole genome sequencing (WGS) approaches will ease the discovery of additional DNA variation not seen by WES. However, the interpretation of the consequences of rare variants remains a challenge. To this end, we established a collection of more than 200 fibroblast cell lines from patients with mitochondrial disorders, WES and respiratory chain complex investigation for a quantitative proteome study. The combined determination of DNA variation and protein steady state levels in a cell line expressing impaired energy metabolism will advance the identification of DNA variants important for protein stability and turnover involved in the etiology of the disease. For quantitative proteomics, we are developing a protocol for a fast and comprehensive analysis based on a limited number of cells. We are using the coding variants from the exome data to predict a cell line specific proteome and we are also considering allele specific detection of peptides. Testing different HPLC gradient separation times, we were able to quantify up to 8 000 proteins (out of 10 000 expected to be expressed in fibroblast cells). Starting with controls and characterized fibroblast cell lines from patients with known diagnosis, we applied 3 different algorithms to verify the approach: A) A principle component analysis allowed a clear distinction between patient and control cell lines. B) Hierarchical cluster analysis of 18 samples followed by pathway enrichment analysis discovered a disease specific profile in FBXL4 patients. C) Case by case analysis revealed not only the absence of NDUFB3 in a cell line harboring mutations in NDUFB3 but also a significant reduction of the respiratory chain complex I subunits. Currently, we are analyzing the first set of unsolved patient cell lines as well as the impact of frequent variants on the corresponding protein level.

2235T

Novel mutation in BCKDHA gene in Iranian population. *m. abiri[1], M.R. Noori Daluie[1], M.R. Alaee[3], A. Sotoodeh[1], S. zeinali[2,4].* 1) Tehran university of medical sciences, Iran; 2) 2Pasteur Institute of Iran, Tehran, Iran; 3) 3Shahid Behesti University of Medical Sciences, Tehran, Iran; 4) Medical Genetics Laboratory, Kawsar Human Genetics Research Center, Tehran, Iran.

Novel mutation in BCKDHA gene in Iranian population M, 1 Noori Daluie M.R, 3 Alaee M.R, 1Sotoodeh A, 2, 4 zienali S 1,2 Abiri 1Tehran University of Medical Sciences, Tehran, Iran. 2Pasteur Institute of Iran, Tehran, Iran. 3Shahid Behesti University of Medical Sciences, Tehran, Iran. 4 Medical Genetics Laboratory, Kawsar Human Genetics Research Center, Tehran, Iran. Abstract: Maple syrup urine disease (MSUD) is a rare inherited metabolic disorder that inherited as an autosomal recessive disorder. The defect is in branched-chain alpha-keto acid dehydrogenase, the enzyme required for the catabolism of branched chain amino acids (valin, Leucin, Isoleucin). MSUD is clinically and genetically heterogeneous disorder. Affected neonates show a distinctive maple sugar odor in earwax, sweat, and urine, poor appetite, irritability and developmental delay. If untreated can lead to seizures, coma, and death. It is estimated MSUD affects 1 in 185,000 infants worldwide, but this estimation can be higher in countries with high rate of consanguineous marriage. Mutation in several gene can cause the disease, including BCKDHA, BCKDHB, DBT, DLD. In this study, we report evaluation of 20 MSUD families with 20 affected individuals from different parts of Iran. Homozygosity mapping with the help of STR markers showed the probable mutated gene. The Analysis followed by amplification of all exons and flanking intronic sequences. The next was bidirectional sequencing of the amplified sequences. Investigation of the mentioned genes in all families are ongoing and here we report a new mutation in a twin which is the outcome of a consanguineous marriage. The patient is an 8 years old boy with developmental delay and severe mental retardation. Homozygosity mapping showed the probable mutation is in BCKDHA gene. Direct sequencing of the exons and intron/exon boundaries showed homozygous deletion of a T in Cd 234 of the mentioned gene. Our results show that the parents are heterozygous carrier of the disease. The spectrum of the mutations underlying MSUD would expand by these previously unpublished mutations in the BCKDHA gene. The identified mutation forms the basis to confirm the clinical diagnosis by genetic testing, provided the availability of Prenatal Diagnosis and pre-genetic implantation Diagnosis for demanding families. Also it helps to identify heterozygous carriers to reduce the burden of MSUD in Iran.

2236M

Two children with Phenylketonuria with normal tetrahydrobiopterin biochemical testing and normal Phenylalanine Hydroxylase gene testing. *P.M. James[1], C. Candeleria[1], M. Colville[1], M. Kennedy[1], H.L. Levy[2].* 1) Divison of Genetics, Phoenix Children's Hospital, Phoenix, AZ; 2) Divison of Genetics and Metabolism, Boston Children's Hospital, Boston, MA.

We report two girls with biochemically confirmed hyperphenylalaninemia who do not have detectable mutations of the Phenylalanine Hydroxylase (PAH) gene. The first girl was seen at age 16 months for developmental delay, hyperactivity, microcephaly; her plasma phenylalanine (Phe) level was 1620umol/L (27mg/dL). Urine pterins were normal x2; the 1st biopterin 39.5%, and the 2nd 67.7% (nl 18-69%). RBC dihydropteridine reductase activity was 1.2 nmol/min/mg Hgb, the 2nd 1.3 nmol/min/mg Hb (nl 0.8-3.9 nmol/min/mg Hb). She was placed on a low Phe diet and responded rapidly. Within 36 hours her plasma Phe level decreased to 274umol/L (4.5mg/dL) , subsequently on a Phe intake of 400-600 mg/day she maintained plasma Phe levels ranging 18-150umol/L (0.3-2.5mg/dL). Thus, her Phe tolerance was that of mild PKU. PAH gene sequencing performed on two different dates by two laboratories was non-informative. Investigating the possibility she had BH4 deficiency despite the normal pterin results she was given a BH4 challenge. Fed a normal diet for 4 days, then given BH4 5mg PO, there was no reduction in plasma Phe level. CSF neurotransmitters were normal. Retrospective study revealed her Phe level on newborn screening at 24 hrs of age was <2mg/dL. The original newborn specimen was retrieved from storage and retested, again showing a normal Phe level of <2 mg/dL. The second girl of African American heritage was found on newborn screening to have a Phe level of 3.6mg/dL at dol1; 7.1mg/dL at dol22 with plasma Phe level of 471umol/L (7.8 mg/dL). She was started on a Phe restricted diet of 83mg/kg/d and maintained plasma Phe levels from 3.3-4.9mg/dL. Lost to follow-up, she presented at 4 and ½ years of age with severe Kwashiorkor, developmental delay, blond hair, short stature (<2sd), and cardiomyopathy. Her plasma Phe on admission was 0.4mg/dL, tyrosine 0.05mg/dL, and albumin 0.6gm/dL. She developed re-feeding syndrome, max plasma Phe was 12.5mg/dL. PAH gene sequencing with del/dup was non-informative. Urine biopterin was 48% (nl 49-74%), and RBC dihydropteridine reductase activity was 4.74 nmol/min/disc (nl 1.96-7.51). Five months post hospitalization she is regaining milestones on a high protein and moderate Phe restricted diet with plasma Phe 2.1-5.3mg/dL. These girls represents the only instances of PKU known to be true negatives in over 50 years of PKU screening in New England and Arizona. Consequently, we suspect they represent a new variant of hyperphenylalaninemia.

2237T

Molecular characterization of homocysteine metabolism in North Indian cohort with Hyperhomocysteinemia. *A. Lomash, S. kumar, S. pandey, A. singh, M. goyal, SK. polipalli, S. kapoor.* Division of Genetics, Department of Pediatrics, MAMC, New-Delhi, , India.

Objective: The current study was conducted to elucidate whether the elevated Hcy levels are associated with the mutations in the MTHFR(A1298C & C677T),TYMS(6bp I/D) & HLA-G(14bp I/D) gene in a North Indian Cohort in the patient with hyperhomocysteinemia Materials and Methods: Fasting plasma total homocysteine concentration were measured using RP-HPLC & Genetic polymorphisms in MTHFR (A1298C & C677T), TYMS (6bp Ins/Del) & HLA-G (14bp Ins/Del) gene were determined by using PCR-RFLP, SSP & ARMS-PCR method & p value <0.05 were considered significant. Result: Forty five subjects with elevated Homocyteine levels were detected with a mean homocysteine value (26.1+2.7 umol/L) out of 180 referral cases and characterize into Pediatric, Adolescent & Adult age group. Stroke & Ectopia Lentis were the commonest reason behind Pediatric (70% & 10%) & adolescent age group (80% & 20%) respectively. The frequencies of the AA, AC, CC genotype of MTHFR (A1298C) in cases v/s controls were 12.5%, 75%,12.5% and 75%, 25% 0%. The frequencies of the CC,CT,TT genotype of MTHFR (C677T) in cases v/s controls were 60%,40% and 75% & 25% respectively. The frequency of 6bp Ins/Ins,Del/Del &Ins/Del polymorphism in TS 3'UTR of TYMS gene in cases vs controls were 17.7%,51.1%,31.1% and 15.5%,28.8%,55.5%..The frequency of 14bp Ins/Ins,Del/Del &Ins/Del polymorphism of HLA-G in cases vs controls were 31.1%,22.2%,46% and 11.1%,31.1%,57.7%. Conclusion: Study explains that elevated homocysteine concentration may be a significant contributor & positively associated with the presence of stroke in patient with pediatrics & adolescent age group. Molecular evaluation show that homozygous Del/Del genotype(P value = 0.0332) of TS 3'UTR (6bp Ins/Del) of TYMS gene & homozygous Ins/Ins genotype(P value = 0.0251) of the HLA-G 14-bp Ins/Del polymorphism respectively is more common in patients with genotypes A1298C & C677T & it appears to confer significant susceptibility to the development of Hyperhomocysteinemia.

2238M

Identification of a Mutation in a Novel Gene Causing a Chédiak-Higashi Syndrome-Like Phenotype. *J.C. Roney[1], M. Gunay-Aygun[1], D. Gul[2], T. Vilboux[1], D.M. Maynard[1], C. Toro[1], W.J. Introne[1], J.C. Mullikin[3,4], M. Huizing[1], W.A. Gahl[1], A.R. Cullinane[1].* 1) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD, USA; 2) Gülhane Military Medical Academy, Department of Medical Genetics, Ankara, Turkey; 3) Comparative Genomics Analysis Unit, Cancer Genetics and Comparative Genomics Branch, NHGRI, NIH, Bethesda, MD, USA; 4) NIH Intramural Sequencing Center, NHGRI, NIH, Rockville, MD, USA.

Chédiak-Higashi syndrome (CHS) is a rare autosomal recessive disorder characterized by partial albinism, recurrent infections, a mild bleeding tendency, and a later onset neurological dysfunction with features that include learning difficulties, cerebellar dysfunction, polyneuropathies, and L-DOPA responsive Parkinsonism. Diagnosis of CHS is based on identification of enlarged granules on peripheral blood smear and enlarged lysosomes and lysosome related organelles in many cell types. Previously, CHS was only known to be caused by mutations in the lysosomal trafficking regulator gene (*LYST*) that encodes a cytosolic protein proposed to function in the regulation of lysosomal size and trafficking. However, there are patients with CHS who do not have mutations in *LYST*, suggesting that mutations in other genes could also result in the disorder. Here we describe a set of twins of consanguineous background who presented with hepatosplenomegaly and were diagnosed with CHS based on typical inclusions in peripheral leukocytes and in precursor cells in the bone marrow. Their subsequent clinical course was dominated by severe global developmental delay and refractory epilepsy; at age 6 years they were non-verbal and unable to sit unsupported. Cellular analysis showed enlarged lysosomes that clustered perinuclearly in both patients' fibroblasts. Whole exome sequencing of one of these patients revealed no mutations in *LYST*, but identified a homozygous 1bp deletion in *WDR91*, representing a novel gene causing a CHS-like phenotype. The same mutation was identified in the affected sibling by Sanger sequencing, and segregated with disease status within the family. *WDR91* encodes WD40 Repeat 91 protein that contains several WD40 domains similar to LYST. These domains fold into a beta-propeller structure known to form a platform that mediates protein-protein interactions. Consistent with the severity of the mutation, WDR91 protein expression is completely absent in the two affected individuals and likely accounts for their CHS-like phenotype. Identification of other CHS patients with *WDR91* mutations will help us understand if WDR91-related CHS causes the severe neurological phenotype observed in these siblings.

2239T

Molecular Characterisation of known & novel mutations in Congenital Adrenal Hyperplasia patients (CYP21A2 gene) : Genetic & diagnostic implications. *r. khajuria[1], r. prasad[1], r. walia[2], a. bhansali[2].* 1) biochemistry, PGIMER,Chandigarh, chandigarh, chandigarh, India; 2) endocrinology,-PGIMER,Chandigarh India.

Congenital Adrenal Hyperplasia (CAH) is autosomal recessive disease, with a wide range of clinical manifestations from severe classical form to late onset form. Most cases of CAH, the inherited inability to synthesize cortisol, aldosterone with subsequent overproduction of androgens, are caused by mutations in steroid 21 Hydroxylase genetic analysis. As a complement to hormonal measurements, mutation analysis of CYP21A2 gene is potentially important tool in diagnosis of steroid 21 Hydroxylase defiency as well as genetic counselling. Our aim was to determine the frequency of common CYP21A2 gene mutations and identify novel mutations. Clinical and hormonal evaluations were used to categorize the patients in Salt Wasting (SW), Simple Virilizing (SV) and Non Classical (NC) forms. About 95 % of mutant alleles have apparently been transferred from linked pseudogene (CYP21P) to active gene (CYP21A2).Molecular analysis of CYP21A2 was performed in 50 patients for detection of common mutations viz gene deletion,Q318X,R356W,V281L,E6 cluster,i2g, ins T in exon 7 or L307 frameshift mutation, P267L by PCR, ACRS, restriction method and finally confirmed by sequencing . Polymorphisms viz D183E & S268T were identified in 50 patients and 50 control by DNA RFLP and sequencing. Novel mutations were identified by SSCP technique and subsequently sequencing of amplified product. Disease causing mutations were identified in patients comprising SW(n=14), SV(n=22) and NC(n=14). Single gene deletion was found with frequency of 18.4 % in SW, SV patients whereas homozygosity was found with 7.9% in these cases. Frequency of other known mutations were to be present at R356W mutation (20 %), Q318x (11 %) ,V281L (25 %), i2g (36 %), L307 (10 %) and P267L (4 %). D183E and S268T polymorphisms are also present in our population . The genotypic frequency of S268T polymorphism was present with 0.42% in normal and 0.86% in patients. H365N, F306V, P357P, D234D are novel mutations in CYP21A2 gene. Each novel mutation was present at frequency of 2 %. This is a comprehensive study showing compound heterozygosity in majority of our CAH population.

2240M

Molecular Analysis of two Indian Mucolipidosis-II (ML-II) patients and Identification of One Novel Mutation in GNPTAB gene. *M. Mistri[1], J. Sheth[1], K. Godbole[2], F. Sheth[1], N. Makadia[3].* 1) Biochemical and Molecular Genetics, FRIGE's Institute of Human Genetics, Ahmedabad, Satellite, India; 2) Department of Genetic Medicine, Deenanath Mangeshkar Hospital & Research Center, Pune, India; 3) N. M. Virani Wockhardt Hospital, Rajkot, Gujarat, India.

Mucolipidosis II (MLII or I-cell disease MIM# 252500) is an autosomal recessive genetic disorder resulting from defects in the membrane-bound enzyme UDPGlcNAc-1-phosphotransferase due to mutation in GNPTAB gene (MIM# 607840) which plays a key role in lysosomal enzyme trafficking. As per HGMD database, 135 mutations have been reported from different ethnic groups, although the disease causing mutations are not known in Indian population except one or two cases. In present study, mutation study for GNPTAB gene in 2 Indian patients with increased appropriate lysosomal enzyme level in plasma was address to ML -II. Mutation study was performed using direct sequence analysis. Both the patients were found to have 2 different mutations that includes novel homozygous missense mutation c.1144 A>C (p.T382P) in one patient and known small homozygous deletion c.3503_3504delTC (p.L1168QfsX5) in another patient, which is frequently observed in various populations with high carrier rate. This homozygous mutation was observed in two unrelated patients with non-consanguineous parentage. In silico analysis further confirmed the pathogenic effect of the novel mutations occurring at highly evolutionarily conserved domain residues in the protein, leading to conformational changes. This novel mutation also provides the new insights into the molecular basis of the disease that can be utilized for the molecular diagnosis, which will be helpful in offering precise genetic counseling, carrier detection followed by prenatal diagnosis.

2241T

DUP 24BP IN CHIT1 IN SIX MEXICAN AMERINDIAN POPULATIONS. *T.D. Da Silva Jose[1, 2], J.A. Juarez[1, 2], A. Porras[1,2], K.J. Jaurez[2], M.T. Magaña[2], L. Sandoval[2], A. Valladares[3], M. Cruz[3], M. Gonzalez[3], A. Soto[3], J.E Garcia[2].* 1) Genetics, Universidad de Guadalajara , Guadalajara , °., jalisco, México; 2) Laboratorio de diagnóstico de enfermedades lisosomales. División de Genética, Centro de Investigación Biomédica de Occidente, CMNO-IMSS. Guadalajara, Jalisco, México; 3) Unidad de Investigación Médica en Bioquímica Centro Médico Nacional, "Siglo XXI" IMSS. Av Cuauhtémoc 330, Ciudad de México, México.

Introduction: In chitotriosidase (CHIT1), a biomarker used in Gaucher disease, 24bp duplication is the most frequent polymorphism at CHIT1 and results in deficient enzymatic activity compromising its use as biomarker. In this study we wanted to determine the allelic and genotypic frequency of dup 24bp on six Mexican Amerindian populations. Material and methods: 692 samples were analized: Purepechas (49), Tarahumaras (97), Huicholes (97), Mayan (139), Tenek (97) and Nahuas (213). DNA extraction and PCR was done as published elsewhere. Statistical analysis was done by direct counting of genotype and allele frequencies and HWE were using a chi-square and Markov chain by Arlequin 3.5. Results: We found the dup 24bp (CHIT1) in distributions observed in Table 1. All groups were in HWE. The allele frequency of dup 24bp was higher than expected in all analyzed groups compared with Mexican mestizo. Table 1. dup 24 bp in Mexican Amerindian populations PopulationsNumberGenotype Frequencies (%)Allele Frequencies (%)Hardy-Weinberg wt/wtwt/DupDup/DupwtDupP value Purépechas4924 (49)18 (38)7 (14)0.670.330.2487 Tarahumaras9750 (52)37 (38)10 (10)0.710.290.4261 Huicholes 9743 (45)42 (43)12 (12)0.660.340.7265 Mayas 13955 (58)66 (37) 18 (6)0.630.370.7948 Tének 9747 (48)40 (41) 10 (11)0.690.310.7316 Nahua 21389 (42)103 (48) 21 (10)0.660.340.2618 Mestizos1 306177 (58)112 (37) 17 (6)0.760.240.8961 (1) Juaréz-Rendón et al. 2011.

2242M

Molecular and Functional Characterization of Mucopolysaccharidosis Type VI in Indian population. *U. Anusha, S. Jamal Md Nurul Jain, B. Ashwin Dalal.* Centre for DNA Fingerprinting and Diagnostics, Hyderabad, Andhra Pradesh, India.

Mucopolysaccharidosis type VI (MPS VI) is a rare, autosomal recessive lysosomal storage disorder caused by deficiency of the enzyme N-Acetyl Galactosamine-4-sulfatase resulting from mutations in the Aryl Sulfatase B (ARSB) gene. In the present study, molecular and functional characterization of the ARSB gene was done for twenty three patients and their families who were enzymatically confirmed to have Mucopolysaccharidosis type VI. A total of fifteen mutations were identified, of which eleven were novel mutations: p.D53N, p.L98R, p.Y103SfsX9, p.F166fsX18, p.I220fsX5, p.I350F, p.W353X, p.H393R, p.S403Yfs, p.P445L, p.W450L, and p.W450C and three were known mutations (p.D54N, p.A237D and p.S320R). All the novel sequence variants were confirmed not to be polymorphic variants by performing sequencing in 50 normal individuals from the same ethnic population. Age of onset, clinical progression and enzyme activity levels in each patient were studied, to look for genotype-phenotype correlations. Haplotype analysis performed for five different families with the recurring mutation W450C, was suggestive of founder effect. Sequence and structural analysis of the ARSB protein using standard software was carried out for determining the impact of detected mutations on the function of the ARSB protein. Functional characterization of all the mutant cDNA constructs were done by transient expression in cultured COS-7 cells and Aryl Sulfatase B enzyme activity was measured in cell extracts. Summary: Many novel mutations which were unique to Indian population were identified in the present study. This is the first study describing mutations in MPS VI cases from India and also provides some insights into the genotype-phenotype correlation in MPS VI. Keywords: Mucopolysaccharidosis type VI, ARSB gene, mutations, functional characterization, genotype-phenotype correlation, ARSB protein structural analysis.

2243T

Antioxidant enzymes gene expression in Fabry Disease along the circadian rhythm. *A.C. Barris-Oliveira, V. D'Almeida.* Department of Psychobiology, Universidade Federal de São Paulo, São Paulo, São Paulo, Brazil.

Fabry disease (FD) is caused by a mutation in the *GLA* gene, located in the X chromosome, which leads to a deficiency in the α-galactosidase A and thus, globotriaosylceramide storage in lysosomes. The symptoms of this disease are not completely explained by the lipid storage and pathophysiological changes observed in this disorder have been correlated with the activation of secondary pathways, as oxidative stress. Our group has already investigated the erythrocytes levels of antioxidant enzymes in FD patients, and has observed higher catalase (*CAT*) activity levels and also, higher levels of total glutathione. The generation of reactive oxygen species and their removal have already been described as processes that vary along the day, following a circadian rhythm. The objective was to investigate gene expression of the antioxidant enzymes: superoxide dismutase (*SOD1*), glutathione peroxidase (*GPX1*) and *CAT* in FD and control fibroblast cultures along a circadian rhythm. Four fibroblast cultures were used, divided equally in FD and control cultures, from female and male individuals. To induce the circadian rhythm, it was used the serum shock protocol with DMEM 50% fetal bovine serum for 2 hours. The RNAs, extracted at the following times: T1, T4, T10, T16, T22 and T28, were further converted in cDNA and used for qPCR. The genes expressions were plotted in graphics in relation to *36B4*, used as endogenous control gene. Comparing the cell cultures from Fabry Disease Woman (FDW) with Control Woman (CW), *SOD1*, *GPX1* and *CAT* showed similar circadian patterns of expression, and there were statistical differences for *SOD1* at T28, for *GPX1* at T10, T16 and T22, and for *CAT* at T16; besides, the amplitude of expression from CW seemed higher. Comparing the cultures from Fabry Disease Man (FDM) with Control Man (CM), the circadian patterns of expression looked different for all genes, and there were statistical differences for *CAT* at T4, T16 and T28; moreover, FDM presented lower amplitude of expression. Comparing CW and CM for all genes, the circadian expressions followed similar patterns, but CW tended to have higher amplitude of expression, as expected. With the data obtained, we conclude that FD cells show a variable gene expression of antioxidant enzymes along the circadian rhythm, and this variation differs between FDW and FDM, adding information to the study of pathophysiological differences between FD male and female patients. Funding: CAPES, FAPESP, AFIP, IGEIM.

2244M

Characterization of NPC1 expression on mRNA and protein levels in a cohort of Niemann-Pick type C disease type 1 (NPC1; MIM #257220) patients. *M. Hrebicek[1], F. Majer[1], D. Musalkova[1], M. Reboun[1], I. Maresova[1], G. Storkanova[1], H. Vlaskova[2], O. Luksan[1], M. Jirsa[2], L. Dvorakova[1].* 1) Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic; 2) Institute for Clinical and Experimental Medicine Prague, Czech Republic.

NPC1 expression in skin fibroblast cell lines were characterized at mRNA and protein levels in a subset of a cohort of 45 patients carrying various combinations of *NPC1* pathogenic variations. Total NPC1 protein expression levels were semi-quantitated by protein immunoblot analysis of fibroblast lysates. Steady state amounts of immunoreactive NPC1 were lower when compared to controls (80% - 10%). In all tested cell lines carrying at least 1 missense *NPC1* variant treatment with proteasome inhibitor bortezomib increased NPC1 amounts, while chemical chaperones (glycerol, 4-phenylbutyrate) and low temperature lead to increases only in some of them.

NPC1 allelic RNA expression ratios were determined in 23 patients by next generation sequencing of RT-PCR products, or in 11 patients by RT-PCR/RFLP followed by fragmentation analysis. In 15 patients carrying two missense mutations the expression from both alleles was generally balanced, with the exception of p.V950G/p.P1007A, in which the ratio was 70/30. From 5 patients carrying a combination of a missense mutation and a small deletion or duplication the ratios were skewed in favor of the missense allele (> 89/11) in three, while in two others the ratios suggested balanced mRNA expression. We have also analyzed 2 kb of *NPC1* promoter region in 62 control alleles and found six common polymorphisms comprising five probable haplotypes. We did not identify any rare variants potentially altering *NPC1* transcription in the patients.

Thus far, the results suggest that the *NPC1* missense mutations generally allow for residual protein synthesis, which may be increased by treatment with chemical chaperones in some cases. Support: IGA MZ CR NT12239-5/201, PRVOUK-P24/LF1/3, MHCZ-DRO-VFN64165.

2245T

Enzyme activity values and Pseudodeficiency Findings of arylsulfatase A and α-glucosidase in a Mexican Population. *A. Juárez[1,2], S. Mendoza[2], A. Porras[1,2], T. Da Silva[1,2], E. García[2].* 1) Instituto de Genética Humana "Enrique Corona Rivera" , Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) Laboratorio de Bioquímica IB, División de Genética, Centro de Investigación Biomédica de Occidente, Guadalajara, Jalisco, México.

Introduction . Lysosomal storage diseases (LSD) are a group of disorders caused by the buildup of molecules due to genetic defects of lysosomal acid hydrolases or receptor, activator, membrane or conveyor proteins. Each condition is derived from the accumulation of one or more specific substrates . There are enzymes involved in LSD that have significantly decreased enzyme activity in general population; this condition is known as pseudodeficiency and is due to the presence of residual alleles. However, individuals with these alleles show no clinical manifestations. The present investigation was performed to determine the presence of reduced enzyme activity of arylsulfatase A (EC 3.1.6.8) and α -glucosidase (EC 3.2.1.20), deficient enzymes in Metachromatic Leukodystrophy and type II Glycogenosis (Pompe disease), in a Mexican population. Material and methods. A cross-sectional study was conducted in 50 individuals of Mexican population (without any clinical manifestation), of which, prior informed consent, leukocytes were extracted from peripheral blood. Subsequently, the proteins were determined by Lowry method, and determination of enzyme activity was performed by spectrophotometric method in arylsulfatase A and fluorometric in α-glucosidase. The following reference values were considered: for arylsulfatase A: 3.0 ± 2.3 nmol / mg of protein / min and for α-glucosidase: 0.1365 ± 0.56 nmol / mg protein / hr. Candidates to the presence of pseudodeficiency alleles were evaluated by sequencing in order to identify polymorphisms associated with this condition; in *ARSA*: c.1049A> G (rs2071421) and c * 96A> G and in *GAA*: c.1726G> A (rs1800307) and c.2056G> A (rs78653500). Results and discussion. Only two individuals had reduced enzymatic activity for α-glucosidase (less than 50% relative to the control range) and were confirmed by molecular study; both presented c.1726G> A and c.2056G> A polymorphisms. No cases were detected for arylsulfatase A. Conclusions. We show for the first time, the presence of pseudodeficiency in Pompe in our population (heterozygous state); the essay confirms that the biochemical study on leukocytes is reliable to detect potential cases of pseudodeficiency and alert to think of them when suspected cases are studied (cases with clinical picture of Pompe disease). This study creates the basis for further study as determine the frequency of pseudodeficiency in Mexican population.

2246M

Unexpected manifestations of Gaucher disease during enzyme replacement therapy. *YM. Kim[1], CK. Cheon[1], DH. Shin[2], SB. Park[3], GH. Kim[4], HW. Yoo[4].* 1) 1Department of Pediatrics, College of Medicine, Pusan National University Children's Hospital, Yangsan , South Korea; 2) Department of Pathology, College of Medicine, Pusan National University Yangsan Hospital, Yangsan, Korea; 3) Departement of Internal Medicine, College of Medicine, Pusan National University Yangsan Hospital, Yangsan, Korea; 4) Medical genetic center, Asan Medical Center Children's Hospital, University of Ulsan College of Medicine, Seoul, Korea.

Background: Gaucher disease (GD) is caused by the deficiency of glucocerebrosidase. This enzyme helps break down fatty substances in the body, and when the enzyme is deficient, fatty substances build up in parts of the body and cause damage. In many individuals with GD, especially Type I, enzyme replacement therapy (ERT) can reduce the enlargement of the liver and spleen and improve anemia and thrombocytopenia. Recently, we encountered a 35 year-old man with GD manifested gastrointestinal involvement during ERT. Case: The patient was monozygote twin and both were diagnosed with GD by bone marrow biopsy showing marked increase in foamy histiocytes and lipid laden macrophages at 9 years old. They took a splenectomy at 10 years old. ERT was begun in the subject at the age of 24. They visited tour hospital 3 years ago and have been treated with ERT as dosage of 60units/kg administered every other week. Their biomarkers including angiotensin converting enzyme and acid phosphatase were decreased but osteolytic bone lesions still remained and he was suffered from bone pain at that time. Whole body magnetic resonance imaging (MRI) showed severe bone necrosis in right hip and both tibia and liver was not enlarged. The subject complained of dyspepsia and difficulty of ingestion of any pills 3 years ago. He performed esophagogastroduodenoscopy (EGD) from another hospital and the result revealed normal. He continued complaining of dyspepsia despite medication with proton pump inhibitor. A follow-up EGD showed multiple nodular yellowish lesions on duodenum by gastrofibroscopy. Gastric biopsy was done and pathologic finding showed that nodular lesions consisted of Gaucher cell infiltrations. We increased the dose of cerezyme from 60IU/kg to 80IU/kg and performed genetic testing. Direct sequencing of GBAgene showed the compound heterozygous mutations, c.259C>T (p.R48W) and c. 5118G>A (p.R257Q). The mutations (R48W and R257Q) are known as rare mutations. The R48W regarded as mild mutation was reported in non-neuronopathic GD manifesting bone disease, whereas R257Q was previously reported in both acute neuronopathic and non-neuronopathic GD. We increased the dose of cerezyme from 60IU/kg to 80IU/kg and are planning to repeat gastroduodenoscopy 6 months later. Conclusion: As ERT might change the natural course of GD during ERT, close monitoring of adverse events and sharing the patient's course during ERT would be required.

2247T

Characterization of the Beta-Galactosidase Protein Isolated from Brain of Normal Sheep and from a Unique Ovine Model of GM1-Gangliosidosis. *D.A. Nevin, A.J. Ahern-Rindell.* Biology, University of Portland, Portland, OR.

Background: β-galactosidase (β-gal) is a lysosomal hydrolase that catalyzes the removal of terminal β-linked galactose moieties from glycoconjugates. Human β-gal is encoded by the *GLB*1 gene and synthesized as an 88 kDa precursor that is processed into a mature 64 kDa enzyme within the lysosome. Patients with mutations in the *GLB*1 gene develop GM1-Gangliosidosis (GM1 [MIM 230500]), a Lysosomal Storage Disorder (LSD) recognized by a decrease in β-gal activity and the buildup of GM1-ganglioside. Several animal models of GM1 are used to research this disorder. This study aims to characterize a unique ovine model of GM1 presenting with dual activity deficiencies of β-gal and alpha-neuraminidase (α-neur). These enzymes associate and are stabilized within the Lysosomal Multienzyme Complex (LMC), a structure responsible for facilitating step-wise substrate degradation in the lysosome. We hypothesize that a missense mutation in the *GLB*1 gene alters the structure of β-gal and compromises the formation of the LMC, producing a secondary deficiency in α-neur activity. **Methods**: Brain tissues were homogenized in lysis buffer containing protease inhibitors, centrifuged, and supernatant collected. Western blotting was used to determine the size of β-gal in ovine brain. Protein lysates were denatured and size-separated using SDS-PAGE. Ovine β-gal was detected using a feline-stimulated antibody that cross-reacted. A glyceraldehyde-3-phosphate dehydrogenase antibody was used as a protocol control (~38 kDa band) and followed by film capture of chemiluminescence. **Results**: Approximately 28, 72, and 79 kDa bands were consistently detected in normal and *GM1* sheep brain. **Conclusion**: Previously in our lab, a 64 kDa β-gal was isolated from normal and GM1-affected sheep fibroblasts. Subsequently, we found a G714T DNA sequence variation causing a C229F missense mutation in exon 6 of the *GLB*1 gene. These findings support our hypothesis that the disease-causing mutation alters the structure of β-gal, but not its size. Preliminary results suggest that tissue-specific protein expression plays a pronounced role in the size of β-gal. Post-translational modifications and/or the presence of β-gal isoforms may explain this size variation. Further analysis of this unique GM1 ovine model is imperative to characterize the specific interactions of the LMC components, to discover the underlying molecular mechanism of this disorder, and to create an effective therapeutic treatment for GM1.

2248M

Galactose Supplementation Improves Glycosylation in PGM1-CDG. *T.E. Gadomski, K.J. Scott, G. Preston, R.O. Crandall, L.A. Pro, T. Kozicz, E. Morava.* Hayward Genetics Center, Tulane University School of Medicine, New Orleans, LA.

Congenital disorders of glycosylation (CDGs) are phenotypically diverse genetic syndromes caused by impaired glycoprotein synthesis. Phosphoglucomutase 1 deficiency (PGM1-CDG; MIM 612941) is a newly discovered and possibly treatable CDG characterized by decreased protein glycosylation, hypoglycemia, abnormal liver function, muscle involvement and endocrine dysfunction. Intercellular cell adhesion molecule 1 (ICAM-1) serves as a useful biomarker for the measurement of cell surface glycosylation. Dietary galactose supplementation is a promising experimental therapy for PGM1-CDG. We present the cases of two female patients, Patients 1 and 2, ages 16 and 19 respectively, who presented with rhabdomyolysis, hypogonadotropic hypogonadism and hypoglycemic episodes. Each patient was diagnosed with PGM1-CDG. Fibroblasts from both patients were cultured and stained for ICAM-1 to measure glycosylation. In vitro galactose complementation assays with repeat ICAM-1 staining were then performed on fibroblasts to assess for improved glycosylation. ICAM-1 staining in fibroblasts from both patients revealed decreased glycosylation compared to a control cell line. Western blot analysis showed decreased PGM1 protein expression in both patients and each patient had decreased PGM1 enzyme activity. Galactose complementation and subsequent ICAM-1 staining revealed a restoration of glycosylation in fibroblasts from both patients. Non-ischemic exercise testing in Patient 2 demonstrated an abnormal lactic acid elevation curve, suggesting a predominantly muscular phenotype with abnormal glycogen release in response to exercise challenge. Following one year of oral galactose supplementation, Patient 1 demonstrated significant improvement in liver and endocrine function and her hypoglycemia resolved. Patient 2 recently started taking oral galactose and showed transferrin restoration after only five weeks. In summary, *PGM1* defect is an intriguing inborn error of metabolism showing a combined phenotype of insufficient glycogen breakdown and abnormal glycosylation. Both of our patients showed abnormal glycosylation and impaired glycogen mobilization but their different presentations demonstrate the phenotypic heterogeneity of PGM1-CDG. As galactose therapy appears to be a promising dietary intervention for normalizing protein glycosylation, future directions include targeted screening for PGM1-CDG and further clinical studies to evaluate the success of oral galactose supplementation.

2249T

MORQUIO SYNDROME: NEW HETEROZYGOUS MUTATION OF THE GALNS GENE IN TWO SIBLINGS FROM SOUTH-WEST COLOMBIA. *M.F. Hernandez-Amariz[1], F. Ruíz-Botero[1], S. Eichler[2], H. Pachajoa[3].* 1) Centro de Investigaciones en Anomalías Congénitas y Enfermedades Raras (CIACER), Universidad Icesi, Cali, Colombia; 2) Centogene. Rostock, Germany; 3) Fundación Clínica Valle del Lili, Cali, Colombia.

Introduction: Mucopolysaccharidosis type IV A, or Morquio syndrome, is an autosomal recessive lysosomal storage disorder that is caused by mutations on the GALNS gene, leading to a deficiency of N-acetylgalactosamine-6-sulfatase enzyme which is encoded by this gene, resulting in the accumulation of keratan sulfate and condroitin sulfate in certain tissues. It is expressed as generalized skeletal dysplasia with corneal clouding, resulting in short stature, pectus carinatum, platyspondylia, odontoid hypoplasia, kyphoscoliosis, and genu valgum. Cardiovascular and respiratory systems can also be affected. This phenotype has a broad variability associated to the more than 180 mutations that have been reported. Until recently there was no effective therapy for treating the disease aside from a multidisciplinary approach, and this year was approved the enzymatic replacement therapy. Objective: to report the case of two siblings identifying a new mutation. Methods: 9 and 6 six year old brothers with clinical characteristic of severe Morquio syndrome, deficit of enzymatic activity of GALNS enzyme, GALNS gene was analyzed by PCR and exome sequencing. Results: two mutations were found, a previously reported one in exon 3 (c.280C>T p.R94C) and a new heterozygous variant in exon 9 (c.998G>A p.G333D), with software analysis predicting it as probably damaging. Conclusion: These brothers have received an approach by a multidisciplinary team including pediatrician, orthopedist, geneticist, endocrinologist, neurosurgeon, physiatrist and neuropsychologist and the starting of enzymatic replacement therapy is being considered. This multidisciplinary approach is recommended for the best control of the disease.

2250M

Identification of genetic mutations in Malaysian patients with fructose-1,6-bisphosphatase deficiency. *L.H. Ngu[1], A.A. Nor Azimah[2], H.Y. Leong[1], B.C. Chen[1], H. Muzhirah[1], M.Y. Zabedah[3], H. Anasufiza[3], Y.K. Chor[4], Y. Yusnita[2].* 1) Genetics Department, Hospital Kuala Lumpur, Jalan Pahang, 50586 Kuala Lumpur, Malaysia; 2) Molecular Diagnostics and Protein Unit, Specialised Diagnostics Centre, Institute for Medical Research, Jalan Pahang, 50586 Kuala Lumpur, Malaysia; 3) Biochemistry unit, Specialised Diagnostics Centre, Institute for Medical Research, Jalan Pahang, 50586 Kuala Lumpur, Malaysia; 4) Paediatric Department, Sarawak General Hospital, Jalan Hospital, 93586 Kuching, Malaysia.

Fructose 1,6-bisphosphatase (FBPase) deficiency (MIM #229700) is an autosomal recessively inherited disorder of gluconeogenesis. Patients present notably with recurrent episodes of hypoglycaemia, ketosis and lactic acidosis, which can be fatal if not treated appropriately. Urinary analysis in acute samples may reveal increased glycerol and glycerol phosphate. However these biochemical diagnostic markers rapidly normalized following treatment and can be easily missed. FBPase (EC3.1.3.11) is encoded by *FBP1* (NM_000507.3) gene on chromosome 9q22.2-q22.3. Genetic analysis of *FBP1* is useful in diagnosing FBPase deficiency, avoiding the need of invasive liver biopsy. Although more than 20 *FBP1* gene mutations have been published, there has been limited investigation into the genetics of this disorder in Malaysian patients. We studied 5 patients (2 males, 3 females) from 4 unrelated families with clinical symptoms and laboratory findings consistent with FBPase deficiency. All of them have recurrent episodes of ketotic hypoglycaemia since infancy. Urinary glycerol and glycerol phosphate were only detected in 2 patients. Bidirectional sequencing of all seven *FBP1* coding exons plus approximately 50 base pairs of flanking noncoding intronic DNA on either side of each exon is performed using genomic DNA. Five mutations were identified in these families. In two siblings from consanguineous parents, we identified a previously reported homozygous missense mutation in exon 7, c.778G>A [p.(Gly260Arg)]. Three other unrelated patients from non-consanguineous parents have novel mutations discovered. One of them has a homozygous missense mutation in exon 5, c.472C>T [p.(Arg158Trp)]. One patient has a heterozygous deletion c.392delT [p.(Val131Glyfs*71)] in exon 4 in association with a heterozygous splice site mutation at c.568-2A>C in intron 5. Another patient has a homozygous deletion at c.603delG [p.(Val201Valfs*3)] in exon 6. This mutation causes a frame-shifting change after Valine-201 and creating a new reading frame that ends at a stop at position 3. All the newly identified mutations in the *FBP1* gene were not found in normal controls in our population and are predicted to result in loss of FBPase activities. Following their genetic diagnosis, adequate management aimed at avoiding fasting periods, particularly during febrile episodes have been initiated for these patients and yielded good outcome to date.

2251T

Molecular Spectrum of HFE Gene Mutations in Patients Referred for HFE Molecular Analysis: A Single Center Study. *F. Ozkinay[1], A. Aykut[2], T. Atik[1], M. Tombuloglu[3], H. Onay[2].* 1) Dept Pediatrics, Ege Univ, Izmir, Turkey; 2) Dept Medical Genetics Ege Univ, Izmir, Turkey; 3) Dept Internal Medicine , Ege Univ, Izmir, Turkey.

Hereditary hemochromatosis is a relatively common autosomal recessive disorder caused by the high absorption and deposition of iron in major organs. This leads to organ dysfunction, and eventually arthritis, cirrhosis, cardiomyopathy, diabetes mellitus, hypogonadism and hepatocellular cancers. Early diagnosis and treatment are of utmost importance in the prevention of end stage organ damage. In more than 90% of hereditary hemochromatosis patients, HFE gene mutations have been responsible for the clinical manifestations. The aim of this retrospective study is to investigate the frequencies of common mutations found in individuals referred for HFE gene molecular analysis to a single genetic center in the Aegean region of Turkey. A total of 230 individuals were tested for HFE mutations between the years 2005 and 2014-June 1. The HFE common polymorphisms (H63D, C282Y, S65C, Q283P, E168Q, E168X, W169X, P160delC, Q127H, H63H, V59M, and V53M) were studied using polymerase chain reaction. Of 230 patients studied for HFE mutations, 73% were found to have no mutation and 27% had mutations either homozygously or heterozygously. Three distinct mutation (C282Y, H63D, S65C) and five different genotypes, with the exception of mutation free genotype, were detected in the group studied. Although no one was homozygous for C282Y mutation, one patient was heterozygous and another compound heterozgyous for this mutation. Five patients (2.1%) were homozygous for H63D mutation, with another fifty three patients (23%) showing heterozygosity for it. One patient had the genotype H63D/C282Y and one patient was heterozygous for S65C mutation. Among the mutant alleles H63D was the most prevalent allele with a frequency of 94.1%. In conclusion; C282Y mutation of HFE gene, which causes severe form of hemachromatosis and is common in Northern Europe, was not found to be frequent in our hemochromatosis patients. The most frequent hemochromatosis causing HFE gene mutation in our population was found to be H63D mutation.

2252M

An infant with hyperhomocysteinemia, methylmalonic aciduria, and an atypical cellular distribution of protein-bound cobalamin. *M. Pupavac[1], F. Petrella[1], D. Watkins[1], S. Fahiminiya[1,2], J. Muenzer[3], J. Majewski[1,2], D. Rosenblatt[1].* 1) Human Genetics, McGill University, Montreal, Quebec, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 3) University of North Carolina, Chapel Hill, North Carolina, USA.

In mammalian cells, derivatives of vitamin B12 (cobalamin, Cbl) are required as cofactors for two enzymes. Adenosylcobalamin (AdoCbl) is required in the mitochondria as a cofactor to methylmalonyl CoA mutase (MCM), and methylcobalamin (MeCbl) is required in the cytoplasm as a cofactor to methionine synthase (MS). Inborn errors of Cbl metabolism can present with elevations of methylmalonic acid and/or homocysteine in the blood and urine. A sixteen-month-old boy of Hispanic ethnicity was investigated because of elevated serum methylmalonic acid and homocysteine levels. Additional clinical findings included cardiac defects (ventral septal defect, long QT syndrome), cleft palate, hypospadias, hyperbilirubinemia, intractable epilepsy, and global developmental delay. There was no family history of similar findings. Cultured patient fibroblasts showed decreased function of both MCM and MS, and decreased synthesis of AdoCbl and MeCbl from exogenous [^{57}Co]cyanocobalamin. Crude subcellular fractionation demonstrated that most internalized radiolabelled Cbl was found in a fraction containing mitochondria, lysosomes and possibly other membrane-bound cellular compartments. The radiolabel in control fibroblasts was predominantly in the cytoplasmic fraction. Size exclusion chromatography of radiolabelled patient fibroblast extracts showed that the majority of labelled Cbl eluted bound to transcobalamin (TC). The majority of label in control fibroblasts was present bound to MCM and MS. Somatic cell complementation analysis suggested that the disorder in this patient was not any known inborn error of cobalamin metabolism. Whole exome sequencing did not identify causal mutations in any gene known to play a role in Cbl metabolism. However, candidate causal variants have been identified in several genes not previously implicated in Cbl metabolism. Family and functional studies are being performed to validate the candidate variants.

2253T

Insightful investigation of mtDNA integrity in affected tissues of patients with mitochondrial disorders. *J. Wang, J. Lin, X. Tian, V.W. Zhang, E.S. Schmitt, LJ. Wong.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Introduction: Mitochondrial biogenesis and maintenance defects are due to mutations in a group of nuclear genes, including TK2, SUCLA2, RRM2B, SUCLG1, DGUOK, MPV17, TYMP, POLG, POLG2, C10orf2, and OPA1. Affected tissues usually showed mitochondrial DNA (mtDNA) depletion, multiple deletions, or point mutations. MtDNA multiple deletions are known to be associated with mutations in nuclear genes involved in mtDNA biosynthesis. Investigation of mtDNA alterations in affected tissues can reveal impacts of nuclear gene defects on mtDNA integrity and their correlation with clinical phenotype. Methods: The mitochondrial genome is amplified by long range PCR followed by massively parallel sequencing (MPS) to assess mtDNA point mutations, large deletion and multiple deletions. MtDNA content was examined by Real Time-qPCR. Nuclear gene was analyzed by target gene capture/MPS or Sanger sequencing. Results: In a total of 34 patients with nuclear gene mutations, mtDNA deletion or multiple deletions in muscle or liver were detected in 25 patients (73.5%). Ten patients with mutations in MPV17, DGUOK, TK2, SUCLG1 or POLG, had mtDNA depletion in the affected tissues, and presented with encephalohepatopathy or encephalomyopathy at young age (1 month to 4 years). Multiple deletions were detected in 11 patients with mutations in RRM2B, POLG, C10orf2 or OPA1. These patients presented with myopathy at adulthood (17 to 78 years old, mean 46). Four patients with encephalomyopathy (4 to 66 years old) and mutations in TK2, RRM2B or POLG had mtDNA depletion and multiple deletions. The number of low heteroplasmic variants is significantly increased compared with that of age and tissue matched individuals without mutations in the above mentioned genes. Conclusions: MtDNA multiple deletions and depletion are secondary defects to defects in nuclear genes responsible for mtDNA biosynthesis and maintenance. Mutations in the POLG, TK2 or RRM2B gene can cause both mtDNA depletion and multiple deletions. Mutations in the MPV17, DGUOK or SUCLG1 gene are mostly associated with mtDNA depletion in patients with early onset of encephalopathy or hepatopathy, whereas autosomal dominantly inherited mutations in C10orf2, POLG or OPA1 are associated with multiple deletions without mtDNA depletion. This study provides insights into the mtDNA integrity and its correlation with nuclear gene mutations and phenotypes. MtDNA depletion and/or deletions in affected tissues may guide molecular diagnosis.

2254M

X-chromosome inactivation in females heterozygotes for Fabry disease. *L. ECHEVARRIA[1], K. BENISTAN[1], A. TOUSSAINT[2], O. DUBOURG[3], A.A. HAGEGE[4], D. ELADARI[5], F. JABBOUR[6], I. DETRAIT[7], C. BELDJORD[2], P. DE MAZANCOURT[7], D.P. GERMAIN[1,7].* 1) Medical Genetics, University of Versailles, Garches, France; 2) Molecular biology, CHU Cochin, Paris, France; 3) Cardiology, CHU Ambroise Paré, Boulogne-Billancourt, France; 4) Cardiology, HEGP, Paris, France; 5) Physiology, HEGP, Paris, France; 6) Biochemistry, CHU Raymond Poincare, Garches, France; 7) EA 2493, UFR des Sciences de la Santé, Montigny, France.

Background: Fabry disease (FD) is an X-linked genetic disorder caused by the deficient activity of lysosomal α-galactosidase A. Inheritance is X-linked so that while males are usually severely affected, female patients range from asymptomatic to clinical disease equal to that of males. The underlying reasons for this phenotypic variability in female patients are still poorly understood. **Aims**: To evaluate the existence of X-inactivation in females with FD, its variability or concordance between tissues, and its potential contribution to clinical phenotype. **Patients and methods**: 56 female patients were enrolled in the study. Extensive clinical work-up was carried out and two validated scores assessed global clinical severity. Kidney function was evaluated using measured GFR, while magnetic resonance imaging (MRI) evaluated left ventricular mass index, left posterior wall thickness and interventricular septum thickness. Alpha-galactosidase A residual activity was measured in leukocytes. X-inactivation status was analysed using DNA methylation studies at the Hpa II sites in the highly polymorphic CAG repeat in the first exon of the androgen receptor gene in four different tissues (leukocytes, mouth smears, skin biopsies and urine sediment). **Results**: Disease severity increased with age. Skewed or highly skewed X-inactivation was found in 29 percent of the study population and a significant correlation between the direction and degree of skewing of X chromosome inactivation and residual α-galactosidase A activities, global severity scores of FD, progression of cardiomyopathy and deterioration of kidney function was evidenced. **Discussion**: This study shows that X chromosome inactivation is a key factor in determining the clinical phenotype of females affected with Fabry disease. Our findings support the inclusion of X chromosome inactivation studies as a tool to predict the natural history of the disease and as a decision making criteria for enzyme replacement therapy in female patients.

2255T

Multiple gene variants cause a phenocopy for Lysinuric Protein Intolerance: A case report. *S. Lipinski[1], A. Hebert[1], O. Alpan[2], S. Pattison[3], A. Tolun[3], S. Huguenin[3], E. Haverfield[4], J. Neidich[4], O. Goker-Alpan[1].* 1) Genetics Unit, O and O Alpan, LLC, Fairfax, VA; 2) Allergy and Immunology Unit, O and O Alpan, LLC, Fairfax, VA; 3) Laboratory Corporation of America® Holdings, Reseaerch Triangle Park, NC; 4) GeneDx, Gaithersburg, MD.

Absorption of amino acids from the renal tubular lumen and small intestine are mediated by solute carrier transport proteins. Cystinuria is a common selective renal aminoaciduria clinically associated with urolithiasis. The protein product of SLC3A1 in cystinuria is involved in cystine and dibasic amino acid exchange for neutral amino acids at the luminal membrane. Lysinuric protein intolerance (LPI), caused by mutations in SLC7A7, is the result of defective dibasic amino acid transport at the basolateral membrane which leads to poor intestinal absorption and urinary loss of dibasic amino acids. Biochemically, LPI is characterized by circulating deficiency of arginine and ornithine, which is thought to lead to classical symptoms of post-prandial hyperammonemia, protein intolerance and orotic aciduria. The metabolic defect in LPI can be partially corrected by supplementation with citrulline, while cystinuria is managed by preventing formation of kidney stones by hydration and urinary alkalization. An 11 year-old female presented with severe abdominal pain, diarrhea, rash, multiple fractures and osteoporosis (Z-score -2.7). There were pain and tingling sensations in the hands and muscle weakness after exercise. Fasting metabolic evaluation revealed elevations of dibasic amino acids in the urine: Cystine 2016.2 umol/g (0.0-141.45), Ornithine 354.8 umol/g (0.0-66.6), Lysine 4330.6 umol/g (0.0-562.2) and Arginine 430.0 umol/g (0.0-69.4), a pattern most consistent with cystinuria. Plasma amino acids were normal. A post-prandial (protein loaded) ammonia and urine orotic acid/creatinine ratio were elevated, 187 ug/dL(19-87) and 1.98 mmol/mol Crt (0.11 - 1.07), respectively, and were considered inconsistent with the classical biochemical and clinical picture of cystinuria. Although protein restriction and citrulline supplementation normalized the ammonia level, the patient continued with abdominal pain. Whole Exome Sequencing revealed homozygosity for the known pathogenic M467T variant in the SLC3A1 gene, consistent with cystinuria. Additional variants possibly related to the mixed phenotype included a S712X heterozygous mutation in the enterokinase (TMPRSS15) gene, homozygosity for the T196I variant in the lysine hydroxylase 2 (PLOD2) gene, and the H599R variant in the sodium channel (SCN4A) gene. The clinical presentation is suggestive of an alternate amino acid transport single gene disorder and could be a phenocopy resultant of multiple gene variants.

2256M

Clinical and molecular characterization of Korean patients with glycogen storage type 1a. *J. Lee[1], G. Kim[1], Y. Kim[2], J. Kim[2], B. Lee[1,2], H. Yoo[1,2].* 1) Medical Genetics Center, Asan Medical Center, Seoul, South Korea; 2) Dept. Pediatrics, Asan Medical Center, Seoul, South Korea.

Glycogen storage disease (GSD) Ia is caused by mutations in the G6PC gene. This condition is characterized by hepatomegaly, doll-like face, hypoglycemia, lactic acidosis, dyslipidemia and hyperuricemia. The aim of the study was to characterize clinical and molecular features and evaluate late complications in of Korean patients with GSD Ia. Thirty Korean patients (18 males and 12 females) from 26 unrelated families were diagnosed based on genetic and biochemical data between 1999 and 2013. Their medical records have been reviewed retrospectively. The mean age at diagnosis was 8.8 ± 9.9 years (range, 8 month to 42 years) and the follow-up period was 11.0 ± 7.0 years (2.4-43 years). Most patients (60%) presented with hepatomegaly and hypoglycemia during infancy. Dyslipidemia (80%) and frequent epistaxis (37%) were common findings as well. Serum lactic acid and uric acid levels at diagnosis were 27.44 ± 32.3 μmol/L and 7.91±2.4 mg/dL, respectively. Serum cholesterol and triglyceride levels were high (mean level: 222.8 ± 56.1 and 639.8 ± 507.1 mg/dL, respectively) as well. In genetic analysis, c.648G>T was the most common (45/52 alleles, 86.5%), followed by p.G122D (2 alleles), p.G222R (1 allele), p.S326P (1 allele), p.R83H (1 allele), p.F51S (1 allele) and p.Y128* (1 allele). Allopurinol and uncooked cornstarch were given to all patients. Fourteen patients (46.7%) received fibrates or HMG-CoA reductase inhibitors for hyperlipidemia. Iron supplementation was required in six patients (20%). Of 16 adults patients (10 males and 6 females), 11 patients (68.8%) had multiple hepatic adenoma and 7 patients (43.7%) had osteoporosis. In addition, 9 patients (56.3%) showed renal problems including microalbuminuria, gross hematuria, medullary calcinosis, ureter stones or renal insufficiency. Four patients (25%) suffered from gout despite preventive allopurinol medication. Two patients (12.5%) had pulmonary hypertension. The final adult heights were 156.1 cm (-2.94 SDS) in males and 155.5 cm (-1.02 SDS) in females. In conclusion, the most common mutation in Korean patients with GSG Ia was c.648G>T, suggesting founder effect. DNA analysis for this mutation is efficient for confirmatory diagnosis rather than liver biopsy or biochemical analysis.

2257T

Two new unrelated cases of Pyrroline-5-carboxylate synthase -new founder effect? *Y. Trakadis, M. Berry, S. Fox, A. Khan, C. St Martin, J. Morel, D. Buhas.* McGill U., Montreal, Canada.

Objectives: Pyrroline-5-carboxylate synthase catalyzes the biosynthesis of proline, ornithine and arginine. Its deficiency (P5CSD) is characterized by neurological and connective tissue abnormalities. 11 cases have been reported, with abnormal amino acids profile in only two families. Two new cases will be described here. **Methods**: A literature and chart review. **Results**: Both patients (B.A. and B.M.) are of Innuit origin and carry the same novel mutation in ALDH18A1 (homozygous c.544A>G). The two families are non-consanguineous and unrelated. B.M. has had a history of irritability, tremor and hypotonia since birth. He developed bilateral cataracts at 8 months. Plasma amino acids showed low Proline, Arginine, Citrulline and Ornithine, suggestive of P5CS deficiency. ALDH18A1 sequencing confirmed the diagnosis. At 2 years old he still had jittering episodes, mild developmental delay, short stature, and feeding problems. His brain MRI demonstrated an unusual T2 hyperintensity in the pons and bilateral widening of the temporal / insular CSF space. B.A. presented at 10 months old with bilateral subcapsular cataracts and recurrent vomiting. She had mild dysmorphic features and stretchy skin. The low level of Citrulline and Arginine, as well as, the low/normal Proline and Ornithine levels were evoking a defect in their synthesis, which was confirmed molecularly. At 17 months old she had hypotonia, joint/skin laxity and pes planus. Her brain MRI showed arterial tortuosity. **Conclusions**: Although rare, P5CSD should be in the differential diagnosis of metabolic causes for early cataracts. A founder effect may exist in the Innuit population of Quebec.

2258M

Phosphoglycerate kinase-1 (PGK-1) deficiency presenting as neonatal onset hemolytic anemia, rhabdomyolysis, and mild developmental delay. *Y. Watanabe[1,2], S. Ozono[1], T. Fukuda[3], H. Sugie[4], S. Yano[5], T. Matsuishi[1,2].* 1) Dept Pediatrics and Child health, Kurume Univ, Kurume, Japan; 2) Research Institute of GC/MS, Kurume Univ, Kurume, Japan; 3) Dept Pediatrics, Hamamatsu Univ, Shizuoka, Japan; 4) Dept Pediatrics, Jichi Medical Univ, Tochigi, Japan; 5) Genetics/Pediatrics, USC, Los Angeles, CA, United States.

Background: Phosphoglycerate kinase-1 (PGK-1) deficiency is an X-linked recessive condition and caused by mutation in the PGK1 gene. Highly variable clinical symptoms including hemolytic anemia, myopathy, seizures, developmental delay and renal failure as well as variable age of onset (infancy to adult) are known in this condition. Case report: The patient was a 2nd child born to a non-consanguineous Japanese couple (38 wk, 3410 g). Apgar scores were 81 and 95. He developed jaundice on the 1st day of his life (total bilirubin 12.1 mg/dl) requiring phototherapy and exchange transfusion. At age 9 months, he was referred to us for a chronic indirect hyperbilirubinemia and hemolytic anemia (Hb 8~9 g/dl). At age 15 months, he was hospitalized for acute onset generalized muscle weakness, worsening anemia, and myoglobinuria which were preceded by a fever. Laboratory studies showed anemia (Hb 5.4g/dL, MCV 95fL), increased reticulocytes (202‰), hyperbilirubinemia (TB 1.9mg/dL, DB 0.2mg/dL), high plasma levels of AST 163 (U/L), CK (4490 U/L), and LDH 1147 (U/L), and abnormal urinalysis with hematuria and myoglobulinuria. Peripheral blood smear showed RBC of variable size and shape. He received RBC transfusion and antibiotics for 8 days, and was discharged home. The activity of PGK-1 in RBC was found to be deficient confirming the diagnosis. Discussion: Neurologic involvement is associated in approximately 50% of patients with PGK-1 deficiency. The presented case has motor and speech delay. The mechanism of neurologic involvement is unclear although glutaminergic synaptic metabolism is thought to be disrupted by PGK-1 deficiency. The patient has been followed with MCT supplementation and has not developed another severe episode of hemolysis or rhabdomyolysis for six months.

2259T

NORRBOTTNIAN VARIANT OF GAUCHER DISEASE IN SOUTHER ITALY: LONG TERM FOLLOW-UP. *M. Grisolia[1], S. Simona[1], F. Ceravolo[1], E. Pascale[1], M. Filocamo[2], P. Strisciuglio[3], D. Concolino[1].* 1) Pediatrics, University "Magna Graecia", Catanzaro, Italy; 2) Centro di diagnostica genetica e biochimica delle malattie metaboliche, Istituto G. Gaslini, Genoa, Italy; 3) Depatment of Pediatrics, University Federico II, Naples, Italy.

The Norrbottnian type of Gaucher disease has been described many years ago, as due to a unique mutation which may have happened in or before the 16th century in northern Sweden and is a well defined nosological entity with a characteristic course of clinical manifestations. We report the results of a long-term follow-up of four patients affected by Gaucher disease type III. The patients, originating from South of Italy, present with clinical features and progression of disease, in particular for the skeletal involvement, comparable with the “Norrbottnian” Swedish phenotype. The patients (3M, 1F), median age 38,2 years (range 32-49 years) belonging to three different families, had been diagnosed at a median age of 3,4 years (range 1,1-8 years). The clinical manifestations at diagnosis were hepatosplenomegaly, thrombocytopenia, anemia and growth retardation. The [L444P]+[L444P] genotype was confirmed in all of them. Bone involvement consisting, primarily of bone pain, bone crisis, osteopenia and bone abnormalities (such as Erlenmeyer flask deformity) detected by X-ray, was present in all four patients. With age a progressive kyphoscoliosis, due to wedging of vertebral bodies, occurred in all patients together with a restrictive ventilatory defect. All of the four patients had cognitive delay and other neurological manifestations began to appear at a median age of six years (range 4-10 years), as horizontal gaze, intentional tremor, seizures for one patient and depression for two different patients. The patients started Enzyme Replacement Therapy (ERT) at a median age of 16,2 years (range 11-23 years) and were assessed during the years of treatment with regular clinical observations, biochemical tests and psychometric testing. The duration of ERT was associated with significant improvements in platelet count, haemoglobin, liver and spleen volumes, while neurological involvement seemed no to benefit of ERT so as skeletal manifestations.

2260M

GM2-gangliosidosis, AB variant: clinical, ophthalmological, MRI and molecular findings. *D. Renaud[1, 2, 3], M. Brodsky[1, 2, 4].* 1) Neurology, Mayo Clinic, Rochester, MN; 2) Pediatrics, Mayo Clinic, Rochester, MN; 3) Medical Genetics, Mayo Clinic, Rochester, MN; 4) Ophthalmology, Mayo Clinic, Rochester, MN.

GM2-gangliosidosis, AB variant is a very rare form of GM2-gangliosidosis due to a deficiency of GM2 activator protein, associated with autosomal recessive mutations in GM2A. Less than 10 patients, confirmed by molecular analysis, have been described in the literature. A 12 month old Hmong girl presented to the Neurometabolic clinic for evaluation of global developmental delay, hypotonia and cherry red spots. The parents are not known to be consanguineous. Her examination was remarkable for hypotonia with hyperreflexia and excessive startling. The head circumference was normal. An extensive neurometabolic evaluation was negative. Developmental regression began at 14 months of age. Retinal examination at 16 months of age disclosed 4+ cherry red / black spots with “heaped up” whitish infiltrate surrounding both foveae but no evidence of optic atrophy or peripheral retinal abnormalities. Repeat MRI scan at 17 months of age revealed delayed but interval myelination associated with abnormal signal intensity of the bilateral thalami presenting as T2 hyperintensity of the posterior thalami in the region of the pulvinar nuclei and T2 hypointensity in the anterior thalami. Sequencing of the GM2A gene revealed a homozygous c.160 G>T mutation, predicted to result in a premature protein termination p. Glu54*.

2261T

The role of inflammation in vascular disease in the MPS I canine model. *M. Vera[1], S. Le[1], S. Kan[1], P. Dickson[1], R. Wang[2].* 1) Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA; 2) CHOC Children's, Orange, CA.

Inflammation has come to be appreciated as an important pathophysiologic mechanism in the mucopolysaccharidoses (MPS). Inflammatory markers and pathways have been identified in the joints and cardiovascular systems in several MPS animal models. Undoubtedly, inflammation causes a significant portion of the morbidities and progressive decline suffered by MPS patients. Arterial disease has been well described in mucopolysaccharidosis type I (MPS I) and consists of intimal-medial thickening with proliferation of vascular smooth muscle cells and extracellular matrix remodeling. Evidence of inflammatory pathways involving Toll-like receptor 4 (TLR4) and transforming growth factor beta (TGFβ) signaling have been observed in these arterial lesions, though the inciting events stimulating these pathways have not been identified. In this report we have studied the role of macrophages as effector cells in the initiation of the vascular disease phenotype. Our model hypothesizes that macrophages are recruited to sites of vascular endothelial damage or dysfunction caused by lysosomal storage where they secrete cytokines that stimulate an inflammatory cascade.

2262M

A Novel mutation in the Methylenetetrahydrofolate Reductase (MTHFR) gene in a Turkish child: A Case Report. *M. ARSLAN[1], S. VURUCU[1], H.I AYDIN[2], B. ÜNAY[1], R. AKIN[1].* 1) GULHANE MILITARY MEDICAL SCHOOL, DEPARTMENT OF CHILD NEUROLOGY, ANKARA, Turkey; 2) TURGUT ÖZAL UNIVERSITY, DEPARTMENT OF PEDIATRIC METABOLISM DISEASES, ANKARA, Turkey.

Severe methylenetetrahydrofolate reductase deficiency is an autosomal recessive metabolic disorder of folate metabolism causing elevated levels of homocysteine and decreased levels of methionine. A 12-year-old boy with nonspecific developmental delay was admitted to our clinic with a 2-week of history of inability to walk, muscle weakness, speech disorder, confusion, and visual hallucinations. The patient was found to have high serum homocysteine and low-normal serum methionine, cerebral and cerebellar atrophy. Molecular genetic analysis identified homozygous p. F564V (c.1690T>G) MTHFR mutation in the patient.

2263T

Hypophosphatasia: a case report. *R. Ortega, J. Prieto.* Pontificia Universidad Javeriana, Bogotá, Colombia.

Introduction: Hypophosphatasia is a rare inborn error of metabolism that is characterized clinically by defective bone mineralization and biochemically by deficient activity of the tissue-nonspecific isoenzyme of alkaline phosphatase in serum and in tissues 1. Clinical expression is variable ranging from stillbirth without mineralized bone to pathologic fractures developing only late in adulthood2. Has been classified into six subtypes: perinatal lethal, prenatal benigna, infantile, childhood, adulthood, odontohypophosphatasia. The biochemical diagnosis is based on the determination of low serum alkaline phosphatase and serum or urine increased phosphoethanolamine, pyridoxal 5'-phosphate and inorganic pyrophosphate3, and molecular diagnosis with gene sequencing of ALPL. Case Report: Patient with short stature evident from 10 months of age, associated with severe chest deformity, marked and progressive genu valgus frontal bossing. At 2 years of life lost teeth that have not yet been replaced starts. Radiographs show metaphyseal dysplasia, generalized demineralization, and cranial synostosis. Serum calcium: 10.3 mg / dl. Chromatography of amino acids, protein, parathyroid hormone, phosphorus, renal function, cerebral MRI, IQ and karyotype were normal. Alkaline phosphatase 30u / L. Finding of elevated calcium and alkaline phosphatase suggest low child hypophosphatasia.To complete the study we request sequence analysis of gen ALPL that report a homozygous mutation c.892G>A (pGlu298Lys). Discussion: This patient has a phenotype that suggests inborn error of metabolism that is confirmed by biochemical analysis suggesting hypophosphatasia. The sequence analysis identified a c.892G>A change in the ALPL gene. This mutation has been previously reported in a Japanese patient. 1. P S Henthorn and M P Whyte. Missense mutations of the tissue-nonspecific alkaline phosphatase gene in hypophosphatasia. Clinical Chemistry 1992; v. 38, p.2501-5. 2. Mornet E. Hypophosphatasia: the mutations in the tissue-nonspecific alkaline phosphatase gene. Hum Mutat. 2000;15(4):309-15. 3. Alonso et al. Hypophosphatasia: new therapeutic approaches. Med Clin (Barc). 2009;132 (3):108-111110 4. Hideo Orimo, Zuisei Hayashi, Atsushi Watanabe, Tsunenori Hirayama, and Tsuneo Hirayama. Novel missense and frameshift mutations in the tissue-nonspecific alkaline phosphatase gene in a Japanese patient with hypophosphatasia. Hum. Mol. Genet. (1994) 3 (9): 1683-1684.

2264M

Autoimmune thrombocytopenia in a patient with Hunter Syndrome: a rare association. *A. Rufus, J. Flores, B. Croke, K. Fernandez, P. Sandiford, R. Antony.* Department of Pediatrics, University of Illinois College of Medicine, Peoria, IL.

Background:Individuals with Hunter syndrome develop coarse facial features, hepatosplenomegaly, cardiovascular disease, skeletal dysplasia, pulmonary disease, neurocognitive deficits and deafness from about 2 years with phenotypic variation. Hematological complications of Hunter syndrome are typically mild thrombocytopenia and neutropenia. We describe our experience managing a patient with Hunter syndrome with severe autoimmune thrombocytopenia/anemia. Case Report: A 4 year old boy with Hunter syndrome presented with fever, fatigue, poor oral intake and decreased urine output. Initial hematological indices: WBC 4600/microlitre, Hemoglobin 5.3g/dl, Platelet count 7000/microlitre, and ANC 1940/microlitre. Parvovirus B19, varicella, EBV, and CMV studies were negative. Blood and bone marrow studies (pancytopenia, dyserythropoiesis, no abnormal cell lines and no hemophagocytosis) ruled out hematological malignancy and HLH. Pericardial effusion was assumed to be viral or secondary to anemia. Despite multiple platelet transfusions his Hemoglobin (<6 g/dl) and platelet count (<10000/microlitre) failed to increment. Direct Antiglobulin Test positivity and antiplatelet antibody positivity led us to a working diagnosis of autoimmune hemolytic anemia and thrombocytopenia. IVIG (1gm/kg) did not have a beneficial effect so prednisone 15 mg q8 was commenced. The patient then developed acute intracranial hemorrhage and pulmonary hemorrhage necessitating oscillatory ventilation. Steroid dose was increased to IV methylprednisolone (MP) 12mg/kg/day and daily plasmapheresis was commenced (continued for 11 days) .Platelet infusions, Keppra, Labetalol and Clonidine were used for supportive care. His PICU course/ platelet count were as follows: Day 3: Platelets = 94000 (MP 2 mg/kg/day), Day 13: Platelets = 195000 and Hemoglobin 10.2 (MP 1mg/kg/day). Day 17: EVD/Ventilation discontinued. Day 45: MP stopped. On stopping the MP his platelet count slowly dropped to 147000 so MP restarted (30 mg Q8h) and IVIG 400mg/kg/day was administered for 4 days. This permitted us to wean and discontinue the MP by day 82. (Platelet =253000). Weekly enzyme replacement therapy was then commenced and patient's platelet counts has been consistently above 150000.Conclusion: With this report of the first case of Hunter syndrome associated with autoimmune thrombocytopenia we hope to add to the body of knowledge regarding the management of hematological complications of Hunter syndrome.

2265T

Bone Health in Children and Adults with Isolated Methylmalonic Acidemia. *J.L. Fraser, J.L. Sloan, E. Harrington, I. Manoli, C.P. Venditti.* National Human Genome Research Institute, NIH, Bethesda, MD.

Methylmalonic acidemia (MMA) causes multisystemic disease, characterized by both acute metabolic crisis and chronic, progressive end-organ damage, and requires lifelong therapy with protein-restricted diets. In our natural history cohort of patients with mut, cblA, and cblB MMA, we have observed patients with pathologic fractures and marked osteopenia/osteoporosis. We sought to explore the clinical features and potential biomarkers associated with decreased bone density in untransplanted MMA patients, as well as the effect of liver/kidney transplantation on bone health in our smaller transplantation cohort. A subset of patients enrolled in our MMA natural history protocol with *mut* (N=31), *cblA* (N=5), and *cblB* (N=8) MMA underwent whole body dual-energy X-ray absorptiometry (DXA) and comprehensive laboratory evaluation. Age and race-based DXA Z-scores were obtained and adjusted for patient height. These values were then used to query the relationships between bone health and body mass, electrolytes, minerals, vitamins, plasma amino acids, MMA-related metabolites, hematological indices, and renal function. Bone mineral density (BMD) Z-scores without adjustment for height suggest that two-thirds of our *mut* MMA cohort had a bone mineral density <-1.5 (osteopenia/osteoporosis). When adjusted for height, 40% of our cohort had BMD Z-score <-1.5, and the mean Z-score value with and without height correction was significantly different. Renal disease was a key determinant of growth retardation. After height correction, alkaline phosphatase levels, the percentages of neutrophils and lymphocytes, HDL cholesterol, triglyceride levels, and markers of oxidative stress, including F2 isoprostanes and oxidized LDL, but not serum MMA concentrations, remained associated with differences in BMD. The patients with the worst bone health (pathologic fractures, osteoporosis, height-adjusted BMD <-2.5) had essential amino acid deficiencies, significant growth retardation, and were more metabolically fragile. Transplanted patients showed improvement in bone health after liver/kidney transplant. Patients with MMA are at high risk for decreased BMD, poor bone health, and pathological fractures. Routine screening for bone mineral density using DXA with height adjustment of the resulting BMD Z-score should help identify patients who require more aggressive therapeutic intervention and reduce morbidity from associated renal disease, pathologic fractures, and poor bone health.

2266M

Optic nerve atrophy in methylmalonic acidemia (MMA): natural history, pathological findings and experience with anti-oxidant therapy. *I. Manoli[1], W.M. Zein[2], J.L. Sloan[1], E. Harrington[1], Y. Wang[3], J. Zhang[3], B.P. Brooks[2], A. Hamosh[4], C.C. Chan[3], C.P. Venditti[1].* 1) Genetics and Molecular Biology Branch, NHGRI, NIH, Bethesda, MD; 2) Ophthalmic Genetics & Visual Function Branch, NEI, NIH, Bethesda, MD; 3) Immunopathology Section, NEI, NIH, Bethesday, MD; 4) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD.

Optic nerve atrophy (ONA) is increasingly recognized as a complication of isolated MMA, caused by a deficiency of the mitochondrial enzyme methylmalonyl-CoA mutase (MUT). Multisystem manifestations include growth failure, pancreatitis and renal disease, but patients can also suffer acute neurological events such as bilateral globus pallidi strokes and ONA. The natural history and disease pathophysiology remain unknown. 26 patients with isolated MMA (17 *mut*, 4 *cblA* and 5 *cblB*, including 3 with organ transplant) underwent ophthalmologic evaluations, including serial optical coherence tomography (OCT), through a dedicated natural history protocol. A total of 56 OCT studies were performed, with 12 patients contributing 2-15 serial measures. Of 77 patients with isolated MMA evaluated, 10 mut patients (6M, 4F) presented with symptoms of acute or subacute visual loss. Age of presentation ranged from 7 to 27 years. Visual loss was bilateral but asymmetric and 8 patients were legally blind. Thinning of the retinal nerve fiber layer (RNFL) preceded symptomatic presentation in 2 patients. There were no apparent environmental triggers shared between patients, and MR imaging variably revealed thin optic nerves. Ophthalmic pathology, assessed by light and electron microscopy using autopsy samples from a 14.5y old mut patient with normal visual acuity, showed thinning of the macular outer nuclear layer, peripheral retinal gliosis and abnormal mitochondrial ultrastructure of enlargement, edema, and degeneration in the photoreceptor inner segment, the retinal ganglion cell and vascular endothelium, as well as corneal epithelium and keratocyte but not the retinal pigment epithelial cell. Coenzyme Q10, vitamin E, idebenone and intravenous N-acetylcysteine were employed with variable success in 5 of the patients. Thinning of the RNFL progressed with near complete loss of vision in all over the ensuing years. Our experience suggests that 1) ONA is a complication of isolated MMA and patients should undergo periodic eye exams, to include OCT, for pre-symptomatic diagnosis; 2) cell-intrinsic mitochondrial morphological changes in the retina, similar to what has been observed in the proximal tubule epithelial cells in the kidney, suggests that cellular autonomy may underlie the ONA in isolated MMA; 3) antioxidants, improved metabolic control and possibly organ transplantation may not prevent progression of disease, but can ameliorate the rate of loss of visual function.

2267T

Ophthalmic manifestations of Cobalamin C disease occur independent of metabolic control and prenatal treatment. *B. Brooks[1, 2], A.H. Thompson[1], N. Carrillo-Carrasco[3], J. Sloan[2], I. Manoli[2], W.M. Zein[1], C.P. Venditti[2].* 1) National Eye Institute, NIH, Bethesda, MD; 2) National Human Genome Research Institute, NIH, Bethesda, MD; 3) National Center for Advancing Translational Sciences, NIH, Bethesda, MD.

Cobalamin C (*cblC*) disease, an inborn error of intracellular vitamin B12 metabolism, is caused by mutations in the MMACHC gene and leads to impaired intracellular synthesis of adenosyl- and methylcobalamin, which produces both methylmalonic acidemia (MMA) and hyperhomocysteinemia. To examine the clinical course of ocular manifestations of *cblC* in relation to genotype and metabolic parameters over time we conducted a retrospective, observational case series study at the NIH Clinical Center. The 25 *cblC* patients ranged in age from 2 to 27 years at last ophthalmic visit, and follow-up ranged from 0 months to 83 months. There were a total of 69 visits; 14 patients were seen more than once, for a median follow-up duration of 36 months. Nystagmus was present in 64% of patients and strabismus 36%. Despite significant elevation of plasma homocysteine, lens dislocation was never observed. The median visual acuity at last ophthalmic visit of the 18 patients with measurable acuity was 1.15 LogMAR (20/280). Retinopathy with a prominent macular component was present in 72% of patients, with MMACHC c.271dupA homozygotes (14) showing the most extensive degeneration. Retinopathy was accompanied by optic nerve pallor and vascular changes in 68% and 44% of patients, respectively. Empiric treatment with prenatal vitamin B12 in two cases resulted in children with less cognitive dysfunction and better visual function than their affected older sibling. The retinopathy, while delayed in its progression compared to that of the older sibling, nonetheless progressed. These data suggest that maternal prenatal vitamin B12 therapy may be palliative, but does not prevent ocular symptoms or eventual ocular progression. Our longitudinal study reports the ocular status of the largest group of patients with *cblC* systematically examined at a single facility over an extended period of time. There are general differences in the progression and severity of macular degeneration, optic nerve pallor, and vascular attenuation between homozygous c.271dupA patients and compound heterozygote genotypes. Tempo and chronicity of the ophthalmic manifestations appear to be unrelated to metabolic status. The effects of prenatal treatment, the presence of patients with severe ocular disease in infancy, and the presence of comorbid developmental abnormalities (microcephaly, cardiac non-compaction), re-enforce that *cblC* disease displays a developmental as well as a degenerative phenotype.

2268M

GENOTYPE-PHENOTYPE CORRELATION IN FABRY PATIENTS DETECTED BY LYSOSOMAL NEWBORN SCREENING. *J. Navarrete[1], D. Cervantes[1], A. Limon[2], R. Del Valle[3], R. Delgado[4], A. Rivera[5].* 1) Dept Gen, Hosp Sur PEMEX, Mexico City, Mexico; 2) Pediatrics Division, Hosp Sur PEMEX, Mexico CIty, Mexico; 3) Gerencia Medicina Preventiva, Subdireccion Servicios de Salud PEMEX; 4) Pediatrics Division, Hosp Reg Villahermosa, Tab, Mexico; 5) Faculty of Medicine, Universidad Anahuac, Mexico city, Mexico.

The goal of newborn screening is an early detection of inborn errors of metabolism diseases. In Mexico we began newborn screening since 1977 with very few inborn errors of metabolism such as phenylketonuria, galactosemia, congenital hypothyroidism, sickle cell anemia and cystic fibrosis. Petróleos Mexicanos is a big governamental institution with approximately ten thousand workers and their families. Since 2005 a larger screening has been done to all newborns in this institution through all the country. We test for most aminoacidopathies including acidurias, hemoglobinopathies, G6PD deficiency, adrenal hyperplasia, cystic fibrosis, and biotinidase deficiency; since August 2021 we included primary immunodeficiencies, Gaucher disease, Niemann-Pick (A/B) disease, Pompe disease, Krabbe disease, Fabry disease, and MPS 1. We analyzed our results from August 2012 to May 2014. We have found 4 newborns with Fabry disease confirmed with enzyme activity and moleuclar analysis; and 4 patients with Pompe disease, three were pseudodeficiencies and ones was late onset presentation. Since the most lysosomal prevalent disease in our cohort was Fabry disease, we describe our findings and compared them with other populations, and we make a close follow up of all newborns with pathological mutations and study and analyze their families. We also look for biomarkers of the disease so we can start treatment depending of the mutation and the course of the disease as soon as possible.

2269T

Biochemical, Molecular and Clinical Heterogeneity in Very-Long-Chain Acyl-CoA Dehydrogenase Deficiency. The Atlantic Canadian Experience. *J. Gillis[1 2], S. Dyack[1 2], D. Skidmore[1 2], H. McDonald[1], J. Farrell[1], M. Chapman[1], N. Kureshi[1].* 1) Dept Medical Genetics, IWK Health Centre, Halifax, NS, Canada; 2) Dalhousie University, Halifax, NS, Canada.

Very long-chain acyl-CoA dehydrogenase deficiency (VLCADD, MIM #201475) is a recognized rare genetic disorder of fatty acid metabolism caused by mutations in the gene ACADVL. The disorder is classified into three forms depending on the time of onset and severity of illness. VLCADD demonstrates autosomal recessive disease inheritance. Accurate diagnosis and early detection are crucial for favourable clinical outcomes for these patients and for this reason VLCADD is now part of most expanded newborn screening (NBS) programs including the Atlantic Canadian provinces (Nova Scotia, New Brunswick and Prince Edward Island). Diagnosis is established through; acylcarnitine analysis by MS-MS of plasma or a dried blood spot specimen, molecular genetic analysis and measurement of residual enzyme activity in lymphocytes. However, very little remains known of the clinical course of this highly variable condition and there are no established guidelines for management. Since screening began in the Maritimes a decade ago, we have ascertained more true positive cases than expected, demonstrating a greater prevalence for the condition. We have identified novel and deleterious mutations in those ascertained via NBS, as well as in those born prior to NBS for VLCADD and clinically symptomatic; notably we have identified a novel splice-site variant in patients of Acadian ancestry. Our primary objective is to characterize the biochemical, molecular and clinical phenotype of patients with a diagnosis of VLCADD, including those identified via NBS in the Atlantic Maritime provinces. We have identified 17 patients known to our clinic ranging in age from birth to 25 years. Despite their asymptomatic presentation those ascertained on NBS were found to have two deleterious mutations and low level of residual enzyme activity (REA) on follow up analysis. Review of clinical data showed that ≤10% REA correlates with development of symptoms in our patients including; significant hypoglycemia, myopathy, cardiomyopathy and death. Moreover our data also suggest that Individuals with REA within 10-15% are also at risk to develop symptoms. Our study indicates that our current NBS method is sensitive and adequate to identify infants with VLCADD and that acylcarnitine profile on NBS may possibly predict biochemical phenotype correlated to REA and allow for earlier risk assessment for cardiac, hepatic and metabolic complications associated with this condition.

2270M

GAL-1-P levels and GALT Gene mutations in infants following abnormal newborn screening for galactosemia in South Florida. *S.A. Hosseini, C. Hung, B.J. Ilagan, G. Ghaffari, B. Johnson, O. Bodamer.* Division of Clinical and Translational Genetics, Dr. John T. MacDonald Foundation, Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, FL 33136.

Background: Galactose-1-phosphate uridyltransferase (GALT) deficiency (OMIM# 230400, Galactosemia) is an autosomal recessive inherited condition with an incidence of 1 in 30,000 to 60,000 newborn infants. Galactosemia is one of the 29 core conditions for newborn screening in the United States. No data are available for confirmatory testing in South Florida. Methods: We received blood samples from 42 infants following an abnormal newborn screening for low GALT activity from August 2012 until May 2014. All infants were born in the State of Florida and screened through the State Newborn Screening Laboratory in Jacksonville. Confirmatory testing samples were analyzed at the Clinical Biochemical and Molecular Diagnostic Laboratories, Department of Human Genetics at the University of Miami for erythrocyte galactose-1-phosphate (GAL-1-P) levels and for GALT gene sequencing, respectively. Standard laboratory protocols were employed. The laboratories are CLIA and CAP certified and participate in proficiency testing. Results: 16 infants had the D/G genotype (GAL-1-P: mean +/- SD; 21.8 mg/dL +/- 16.6), 18 infants were carriers for either D or G variants (GAL-1-P: 7.1mg/dL +/- 10.4), and 8 were without detectable mutation (GAL-1-P levels ranged from >1 to 1.4 mg/dl). GAL-1-P levels were significantly different between infants without mutation, carriers, and D/G infants ($p<0.005$). The Duarte variant c.940A>G (p.Asn314Asp) in cis with the promoter GTCA deletion (c.-116_-119delGTCA) as present on 25% of alleles. 22% of alleles carried the pathogenic mutations c.404C>T (p.Ser135Leu) or c.563A>G (p.Gln188Arg). Two alleles carried the LA variant c.940A>G (p.Asn314Asp) without, GTCA promoter deletion. We did not observe any infants with classic galactosemia. Discussion and conclusions: Confirmatory testing following abnormal newborn screening for galactosemia is straightforward provided GAL-1P can be analyzed within a few hours followed by timely molecular testing.

2271T

Implications for newborn screening by sequencing. *Y. Zou[1,2], S.E. Brenner[1].* 1) University of California, Berkeley, CA; 2) Fudan University, Shanghai, China.

Fatty acid oxidation disorders (FAOD), amino acid disorders (AAD) and organic acid disorders (OAD) are three groups of inherited metabolic disorders with an aggregate population frequency of roughly 1.4-5.0 per 10,000. These comprise several rare diseases that lead to severe morbidity or mortality if untreated. Therefore newborns are routinely screened, primarily using tandem mass spectrometry (MS/MS), to identify metabolites that are diagnostic of disease. With improvements in sequencing technology, it is becoming possible to imagine that whole genome/exome sequencing approaches may be applied in newborn screening. In this case, it will be necessary to identify the pathogenic variants in the genes responsible for these diseases.

As a foundation for such studies, we have examined the landscape of variants in 63 genes known to be associated with FAOD, AAD and OAD. As these diseases are believed to be recessive, we explored homozygous and compound heterozygous protein-altering variants in individuals whose genomes were sequenced as part of the 1000 genomes project. We observed that more than 800 protein-altering variants with minor allele frequency (MAF≤0.01 are present, and 76 protein-altering variants show homozygous (20 variants) or compound heterozygous (56 variants) state in at least one person. Several genes have higher homozygous frequency than expected from Hardy-Weinberg equilibrium. We found one homozygous or compound heterozygous protein-altering mutation with MAF≤0.01 in a FAOD, AAD or OAD related gene in 17, 28, and 19 people, respectively, among 1092 individuals. PolyPhen2 predicts that approximate 20% of homozygous protein-altering variants (4 variants) and 23% of compound heterozygous protein-altering variants (13 variants) are pathogenic. Moreover, based on the annotation of Human Gene Mutation Database (HGMD), 5 of 20 homozygous protein-altering variants observed in healthy people were validated to be linked to severe metabolic disorders like phenylketonuria or carnitine palmitoyltransferase II deficiency. As these greatly exceed the population incidence of these diseases, we infer that the vast majority of rare variation in these genes, including that flagged as pathogenic by either computational methods or databases, is likely benign. These findings indicate that it will be challenging to identify causal variants in screening the newborn defects of metabolic disorders by genomic sequencing technology.

2272M

Newborn screening and the incidences of inherited metabolic and endocrine disorders in the Arab Middle East. *A.M. Kinrich[1], P.A. Sanchez-Lara[2].* 1) Keck School of Medicine of University of Southern California, Los Angeles, CA; 2) Children's Hospital Los Angeles, Keck School of Medicine of University of Southern California, Los Angeles, CA.

The Middle East and North Africa (MENA) is home to over 300 million people spanning 17 countries and 4000 miles, with over 7 million annual births and some of the highest rates of consanguinity in the world. Several studies suggest that the incidences of inherited metabolic and endocrine disorders in MENA are higher than in Western countries. Early detection and treatment of these conditions can reduce morbidity and mortality, but newborn screening (NBS) can be sparsely performed in parts of these regions. We provide descriptive analysis of the national incidences of metabolic and endocrine disorders and assess the availability of NBS programs in each MENA country. Data was gathered from the Center for Transmission Genetics in Arabs database, indexed articles in PubMed, and through contact with regional experts. Newborn screening programs were categorized by number of conditions tested and how much of the population was screened in each country. Incidence rates were qualitatively compared to published incident rates of Californian neonates of Middle Eastern descent (Feuchtbaum et al 2012). NBS programs in the region were found to vary widely: seven wealthier countries with lower annual birth rates achieved nearly universal screening for a number of conditions, although these populations represented only 12.2% of MENA births. Two countries had limited national screening programs, and eight developing countries with higher annual birth rates had no universal screening program, which together comprised 87.8% of annual births in the region. Incidence rates were identified for 66 inherited endocrine and metabolic disorders in 10 of 17 MENA countries. Rates of disease were found to be higher in MENA countries than in Californian neonates of Middle Eastern descent. Published data on MENA NBS programs are insufficient—only a small number of MENA countries, accounting for 12.2% of regional births, have widespread neonatal genetic screening for multiple disorders. In most of the region, screening is limited to a few conditions in select portions of the population. As a consequence of insufficient access, the incidences of heritable diseases are likely underestimated in the MENA region. There is significant benefit in expanding national NBS programs in the MENA countries to avoid preventable morbidity and mortality associated with untreated inherited endocrine and metabolic disorders.

2273T

Providing more education to parents about newborn screening: not harmful and probably beneficial. *B.J. Wilson[1], B.K. Potter[1], J.C. Carroll[2], J. Little[1], D. Castle[3], P. Chakraborty[4,5], S. Craigie[1], H. Etchegary[6], L. Lemyre[7], F.A. Miller[8], G.A. Wells[1], J. Milburn[4], R. Rennicks White[9,10], G. Tawagi[9], M. Walker[9,10,11], CIHR Emerging Team in Genomics in Screening.* 1) Epidemiology & Community Med, University of Ottawa, Ottawa, ON, Canada; 2) Department of Family & Community Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada; 3) ESRC Innogen Centre, University of Edinburgh, Edinburgh, UK; 4) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 5) Department of Pediatrics, University of Ottawa, ON, Canada; 6) Clinical Epidemiology, Memorial University Newfoundland, St John's, NL, Canada; 7) School of Psychology & Institute of Population Health, University of Ottawa, ON, Canada; 8) Institute of Health Policy, Management and Evaluation, University of Toronto, ON, Canada; 9) Department of Obstetrics, Gynecology and Newborn Care, The Ottawa Hospital, ON, Canada; 10) Clinical Epidemiology Program, Ottawa Hospital Research Institute, ON, Canada; 11) Department of Obstetrics and Gynecology, University of Ottawa, ON, Canada.

Background: Effective parental education about newborn screening (NBS) is important even where screening is mandated. Potential benefits include promoting trust, mitigating anxiety about the NBS process and possible outcomes, and emphasizing the importance of follow up where needed. However, concerns are sometimes expressed that more comprehensive education may lead to parental anxiety and lower acceptance rates in consent-based programs. Published expert opinions suggest key NBS educational messages, but empirical evaluations are lacking. This study formally examined the effect of specific NBS education content on parents' decision-making. **Objectives:** To measure the effect of specific educational messages on pregnant women's decisional conflict about NBS. **Methods:** Women with low-risk pregnancies were recruited from routine second trimester ultrasound clinics in Ottawa, Ontario. Using a factorial survey, they were randomized to receive different combinations of messages: possibility of false positive/negative results; pain from the heel-prick; possibility for results of unclear clinical significance; storage/secondary use of bloodspots; and the nature of parental consent. In total, 32 (2^5) discrete message combinations were assessed. The primary outcome was the mean score on the validated Decisional Conflict Scale (DCS) (1-100, higher scores indicating higher decisional conflict). Secondary analyses explored associations with respondent characteristics, their understanding of key messages, and cognitive burden. **Results:** We received responses from 494 participants. The mean (95% CI) DCS score was 27 (25-29). We observed statistically significant differences in DCS score between recipients and non-recipients of two test messages. The heel-prick pain message ($p<0.01$) and the storage/secondary use of bloodspots message ($p<0.05$) were associated with lower DCS scores. Specific knowledge questions were more likely to be answered correctly if the participant had received the corresponding message. The mean DCS score declined significantly with increasing number of messages received (test for linear trend, $p<0.01$). **Conclusions:** The results suggest that providing information does not compromise NBS decision-making, and may enhance it. This study has high internal validity, but needs to be replicated in more diverse populations, using a broader set of outcomes. The factorial survey design is a useful approach in this context.

2274M

Evaluation of Behavior, Executive Function, Neurotransmitter Function and Genomic Expression in PKU "Nonresponders" to Sapropterin. *H.C. Andersson[1,2], A. Cunningham[1,2], K. Crivelly[1], Y. Li[2], D. Goldstein[3], T.J. Chen[1,2].* 1) Hayward Gen Ctr, Tulane Univ Med Sch, New Orleans, LA; 2) Dept. Pediatrics, Tulane Univ Med Sch, New Orleans, LA; 3) NINDS, NIH, Bethesda, MD.

Oral sapropterin (BH4) has been part of PKU care since FDA-approval in 2007. Sapropterin enhances residual metabolism of phenylalanine hydroxylase and 40-50% of PKU patients experience decreases in plasma phenylalanine (PHE). Anecdotal reports exist of patients who have not experienced changes in plasma PHE but have been noted to, or have reported, improved attention and mental clarity. Behavior, metabolic effects and gene expression has not been studied in PKU patients who were nonresponsive to sapropterin. We have studied 17 classical PKU patients (ages 8-39y) previously shown to be non-responders (NR) to Kuvan® and 4 patients subsequently shown to be responders (R). Prior to receiving Kuvan®, patients completed behavioral inventories of executive functioning and adaptive function (BASC, BRIEF), submitted 24-hour urine for catechol analysis and had gene expression arrays performed at baseline and after 4 weeks of Kuvan® (20mg/kg). Plasma amino acids during with the trial confirmed non-responsiveness to Kuvan®. Dietary diaries were used to assess stability of dietary PHE. RESULTS: On BASC, 2/10 NR patients and 1/1 R patient demonstrated significant improvement on multiple measures to normal score from significantly abnormal; 10 adults were too old for (BASC validated only to 21y). Only one patient demonstrated worsening on one measure. On BRIEF, 3/16 NR patients and 2/3 R patients showed significant improvement on multiple measures; 1/16 NR patient worsening on a single measure. Urine catechols showed no consistent pattern among NR or R patients. Gene expression arrays showed 171 diverse genes among NR patients which showed ≥1.5-fold up-/downregulation and R patients showed 316 diverse genes ≥1.5-fold up-/downregulation. Numerous canonical gene pathways not known to be directly involved in PHE metabolism were involved. In summary, a few NR patients demonstrated a clear effect of Kuvan® on behavior, suggesting a possible neurocognitive benefit of sapropterin beyond its effect to mediate PAH gene function.

2275T

TREATMENT OF PHENYLKETONURIA WITH A NEW FORMULA CONTAINING LNAA. *D. Concolino[1], I. Mascaro[1], M.T. Moricca[1], G. Bonapace[1], k. Matalon[2], P.V Patel[3], R. Matalon[3], P. Strisciuglio[4].* 1) Pediatrics, University "Magna Graecia", Catanzaro, Italy; 2) Health and Human Performance, University of Houston, TX; 3) Deptment of Pediatrics, University Texas Medical Branch, Galveston, TX; 4) Deptment of Pediatrics, University of Naples, Italy.

Background and objectives: Phenylketonuria is an autosomal recessive disease caused by deficient activity of phenylalanine hydroxylase. A low phenylalanine diet is used to treat PKU. The diet is very restrictive and dietary adherence decreases as patients get older. A new formula (PheLNAA) has been tested in this study with the purpose to improve compliance and lower blood phenylalanine. Patients and Methods: The formula has been tested for nitrogen balance and is nutritionally complete. It is fortified with more nutritional additives that can be deficient in PKU diet, such as B12, Biotin, DHA, Lutein and increased levels of Large Neutral Amino Acids to help lower blood Phe. The new formula has been tested on 12 classic PKU patients with a loading test of 4 weeks and the patients who initially responded to treatment continued the PheLNAA administration for an additional five months. Results: Fifty-eight percent of patients had a significant decline in blood Phe concentration from baseline throughout the study. The PheLNAA was well tolerated with excellent compliance and without illnesses or side effects. Conclusion: The new formula is suitable for PKU treatment and offers the PKU clinic a new choice for treatment specially in affected adolescence.

2276M

Optimizing treatment for neonatal hyperammonmia. *P.A. Levy.* Pediatrics, Children's Hospital at Montefiore, Bronx, NY. Albert EInstein College of Medicine.

Treatment of hyperammonemia in the neonate is at best problematic. Many factors complicate treatment. Discovering the hyperammonemia requires a good degree of suspicion. Current treatment requires removal of ammonia by Renal Replacement Therapy (RRT), either by hemodialysis or CVVH (continuous veno-venous hemofiltration). Confounding the treatment is the size of the patient (neonate) who usually weigh between 2.5 and 3 kg, and the placement of catheters suitable for dialysis. “The optimal RRT prescription for neonatal hyperammonemia remains unknown,” as was concluded by Spinale, et al, in a recent article in Pediatric Nephrology (2013). We decided to poll nephrologists as to their approach to RRT for hyperammonemia in a neonate. An informal telephone poll was conducted. Nephrologists were asked their approach to treating a 3 kg neonate with hyperammonemia. Findings: 1.The start of therapy was cited as the important factor for successful outcome 2. Catheter placement whether by ICU staff, interventional radiology or Pediatric surgery was not as much of a concern as was the speed with which the line could be placed so RRT could begin. 3. Catheter size was generally agreed to be a 7 French for this size infant. 4. Area of placement was optimally felt to be the internal jugular in most cases, but again speed of access would override the area of placement. 5. Hemodialysis vs CVVH: The division here seemed to depend on the experience of the nephrologists. Those more comfortable with hemodialysis preferred that mode of ammonia removal, while other felt that higher flows with CVVH could achieve similar results and was more easily accomplished with CVVH. Standardization of treatment is difficult because it depends on the preferences and experience of the treating nephrologist and factors related to each institution.

2277T

Melatonin and Dopamine as Biomarkers to Optimize Treatment in Phenylketonuria: Effects of Tryptophan and Tyrosine Supplementation. *S. Yano[1], K. Moseley[1], C. Azen[2].* 1) Genetics/Pediatrics, University of Southern California, Los Angeles, CA; 2) Clinical and Translational Science Institute, University of Southern California, Los Angeles, CA.

Objective: To determine if additional supplementation of tryptophan (Trp) and tyrosine (Tyr) improve serotonin and dopamine metabolism in individuals with phenylketonuria (PKU) treated with large neutral amino acid (LNAA) tablets. Study design: Ten adult individuals with PKU participated in a randomized double-blind placebo controlled crossover study consisting of three 3-week phases: washout, treatment with LNAA tablets plus supplementation with either Trp and Tyr tablets or placebo, and LNAA tablets plus the alternate supplementation. An overnight protocol to measure blood melatonin, a serotonin metabolite in the pinealocytes, and urine 6-sulfatoxymelatonin and dopamine in first void urine specimens was conducted after each phase. Results: Serum melatonin and urine 6-sulfatoxymelatonin and dopamine levels were increased in the LNAA phase (LNAA plus placebo) compared to the washout phase. Serum melatonin and urine 6-sulfatoxymelatonin did not increase in the active phase (LNAA plus Trp+Tyr) compared to the LNAA phase, although plasma Trp/LNAA ratios were increased compared to the LNAA phase. Among 7 subjects whose plasma Trp/LNAA ratios were above 0.03, a negative correlation between urine 6-sulfatoxymelatonin and plasma Phe levels was observed ($r = -0.072$). Urine dopamine levels and plasma Tyr/LNAA ratios were increased in the active phase compared to the LNAA phase. Conclusion: Melatonin levels did not increase with the larger dose of Trp supplementation although dopamine levels were increased corresponding to the larger dose of Tyr supplementation. Serotonin synthesis appears to be suppressed by high Phe levels at the Trp Hydroxylase level.

2278M

Induction of immune tolerance in MPS I patients initiating enzyme replacement therapy with Aldurazyme. *G.F. Cox[1], R. Giugliani[2], P.V. Novikov[3], S. Richards[1], Y. Xue[1].* 1) Genzyme, a Sanofi company, Cambridge, MA; 2) Medical Genetics Service/HCPA, Department of Genetics UFRGS and INAGEMP, Porto Alegre, RS, Brazil; 3) Department of Clinical Genetics, Moscow Research Institute for Pediatrics and Children Surgery, Moscow, Russia.

Objectives: Mucopolysaccharidosis I (MPS I) is a rare autosomal genetic disorder caused by deficiency of the lysosomal enzyme α-L-iduronidase and the subsequent accumulation of its substrates, the glucosaminoglycans (GAGs) dermatan and heparan sulfates. This results in progressive and debilitating multi-organ disease. Enzyme replacement therapy with laronidase (Aldurazyme®) is approved for the treatment of patients with MPS I. Studies in MPS I dogs have shown that antibodies to laronidase affect its biodistribution and ability to clear glycosaminoglycans (GAGs). In patients with MPS I, high antibody titers correlate with less urinary GAG reduction, but not with clinical response. This study was designed to determine whether an immunosuppressive regimen that induced immune tolerance in dogs could be used in patients with MPS I. Methods: This open-label clinical trial (NCT 00741338) in patients with severe MPS I (≤ 5 years of age) included a Tolerance Induction Period (TIP) followed by an Immune Challenge Period (ICP). The immunosuppressive regimen consisted of cyclosporine A (CsA), azathioprine, and low-dose weekly Aldurazyme (0.058 mg/kg). Following gradual discontinuation of immunosuppressants, patients received full-dose Aldurazyme (0.58 mg/kg) for a total of 24 weeks. Immune tolerance was defined as an anti-laronidase IgG antibody titer ≤1:3,200 at the end of the ICP. A sequential adaptive design allowed for modification of the immune tolerance regimen in a second patient cohort if the initial regimen was unsuccessful (<2 of 3 patients immune tolerant). Safety evaluations focused on signs of infection or organ toxicity, blood pressure elevation, and bone marrow suppression. Results: None of the 3 patients in Cohort 1 and 1 of 3 patients in Cohort 2 achieved immune tolerance. A second patient in Cohort 2 maintained a low IgG titer after 17 weeks of full-dose Aldurazyme, but discontinued the study due to infusion-associated reactions and was considered a treatment failure. CsA levels were variable and required frequent monitoring and dose titration. There were no deaths or other significant adverse events. Conclusions: The results of the study are inconclusive. The risk, benefit, and feasibility of immune tolerance induction should be reconsidered in patients with MPS I. This study was supported by the Genzyme/BioMarin Joint Venture.

2279T

Effects of pre-symptomatic initiation of enzyme replacement therapy for infantile-onset Pompe disease. *M. Kosuga[1], M. Tajika[2], Y. Miwa[2], K. Fujimaki[2], T. Matsuoka[2], K. Soga[2], Y. Umeda[2], S. Uemura[3], T. Fukuda[4], H. Sugie[5], T. Okuyama[1].* 1) Department of Clinical Laboratory Medicine, National Center for Child Health and Development, Tokyo, Japan; 2) Showa University Northern Yokohama Hospital, Children's Medical Center, Yokohama, Japan; 3) Showa University Northern Yokohama Hospital, Cardiovascular Center, Yokohama ,Japan; 4) Hamamatsu University School of Medicine, Department of Pediatrics, Shizuoka, Japan; 5) Jichi Children's Medical Center, Tochigi, Japan.

Pompe disease is an autosomal recessive lysosomal glycogen storage disease caused by a deficiency of acid α-glucosidase. Infantile-onset Pompe (IPD) disease is characterized by progressive cardiomegaly, skeletal muscle weakness, delay of motor function and respiratory insufficiency, and die before the first 1 year of life. Enzyme replacement therapy (ERT) has been shown to improve symptoms and life expectancy of IPD patients when it starts at early (pre-symptomatic) stage. We report on two siblings with IPD in this study. The siblings initiated ERT at 4 months (older sister) and 12 days (younger sister) of age, and we compared their outcomes after 6 and 3 years of treatment respectively. At the start of treatment, the older sister showed typical symptoms of IPD. Now she is 7 years old, bedridden and needs the invasive ventilation due to progressive muscular weakness. Younger sister was initiated ERT at 12 days after birth without any symptoms of IPD. She acquired natural motor development and has no symptoms of IPD at 3 years old. These results show that early ERT initiated at pre-symptomatic IPD is effective in preventing disease progression of IPD. Early diagnosis and ERT introduction can result in better clinical outcomes. Therefore, newborn screening is necessary to allow IPD patients to be treated as early as possible.

2280M

Expression and purification of human recombinant α-*N*-acetylglucosamine-6-sulphatase. *S. Le, S. Kan, R. Dokko, P. Dickson.* Dept Med Gen, LA Biomed Harbor-UCLA, Torrance, CA.

Mucopolysaccharidosis type III (MPS III, Sanfilippo syndrome) is a heritable lysosomal disorder of heparan sulfate degradation, divided into four types (A-D), depending on the enzyme deficiency. All four MPS III types are characterized by severe neurologic problems and relatively mild somatic syndromes. MPS IIID (Sanfilippo D syndrome) one of the rarest Sanfilippo syndrome and is caused by the deficiency of GNS (α-N-acetylglucosamine-6-sulfatase; EC 3.1.6.14). In this study, a full-length human GNS cDNA with a c-Myc purification tag was cloned into an expression vector, pCIneo, and expressed in Chinese hamster ovary cells. The secreted recombinant human GNS (rhGNS) was concentrated from culture medium and a one-step purification with c-Myc affinity chromatography was performed. The purified rhGNS demonstrated the expected molecular weight of the secreted form of GNS at approximately 110 kDa by SDS-PAGE. Deglycosylation assay with PNGase F showed that rhGNS expressed in CHO cells is highly glycosylated as expected. Further biochemical characterization and *in vitro* identification in MPS IIID fibroblasts will be performed to identify if this recombinant protein can be a feasible enzyme replacement therapy candidate for MPS IIID patients.

2281T

Clinical response to eliglustat in treatment-naïve patients with Gaucher disease type 1: Post-hoc comparison to imiglucerase in a real-world setting. *R. Mankoski, J.S. Taylor, J. Angell, M.J. Peterschmitt.* Genzyme, a Sanofi company , Cambridge, MA.

Introduction: Gaucher disease type 1 is a multi-systemic lysosomal storage disorder resulting from acid β-glucosidase deficiency. The standard of care for more than two decades has been enzyme replacement therapy. Eliglustat is a novel oral substrate reduction therapy in development for adults with Gaucher disease type 1.

Objective: To compare long-term treatment response to eliglustat and imiglucerase in treatment-naïve patients with Gaucher disease type 1.

Methods: Four-year data from eliglustat-treated patients in an open-label study (NCT00358150, N=26) and 18-month data from a randomized, double-blind, placebo-controlled study (ENGAGE: NCT00891202, n=20 in eliglustat arm) were compared to 75 matched imiglucerase-treated patients enrolled in the ICGG Gaucher Registry who had received at least 15 U/kg/2 weeks.

Results: At baseline, hematologic parameters were similar in the two groups but eliglustat patients had slightly larger spleens and livers. Time course and degree of improvement were similar for eliglustat- and imiglucerase-treated patients for most parameters. After 4 years, mean spleen volume decreased by 63% and 48%, mean liver volume decreased by 27% and 30%, mean platelet count increased by 95% and 99%, and mean hemoglobin level (g/dL) increased by 2.27 and 0.71 in eliglustat and imiglucerase patients, respectively. Improvements in lumbar spine and total femur Z-scores were consistently higher in the eliglustat group at all time points; however, bone data were limited from the imiglucerase-treated patients. The Z-score increases observed with eliglustat were higher than those observed by Wenstrup et al. (2007, J Bone Min Res) during low- to high-dose treatment with imiglucerase (0.06-0.13 Z-score/year) in patients who had similar mean baseline bone mineral density.

Conclusion: Although not a head-to-head trial, this post hoc analysis suggests that eliglustat, in treatment-naïve patients with Gaucher disease type 1, results in improvements in organ volumes and hematologic parameters that are comparable to those observed with imiglucerase in a real-world setting.

2282M

ENCORE: A randomized, controlled, open-label non-inferiority study comparing eliglustat to imiglucerase in Gaucher disease type 1 patients stabilized on enzyme replacement therapy: 24-month results. *T.A. Burrow[1], T.M. Cox[2], G. Drelichman[3], R. Cravo[4], M. Balwani[5], A.M. Martins[6], E. Lukina[7], B. Rosenbloom[8], L. Ross[9], J. Angell[9], A.C. Puga[9].* 1) Cincinnati Children's Hospital Medical Center, Cincinnati , OH; 2) University of Cambridge, Addenbrooke's Hospital, Cambridge, UK; 3) Hospital de Niños Ricardo Gutiérrez, Buenos Aires, Argentina; 4) HEMORIO, Rio de Janeiro, RJ, Brasil; 5) Mount Sinai Hospital, New York, NY, USA; 6) Universidade Federal de São Paulo, São Paulo, SP, Brasil; 7) Hematology Research Center, Moscow, Russia; 8) Cedars-Sinai/Tower Hematology Oncology, Beverly Hills, CA, USA; 9) Genzyme, a Sanofi company, Cambridge, MA, USA.

Introduction: Gaucher disease is an autosomal recessive disorder caused by deficient activity of the lysosomal enzyme acid β-glucosidase (glucocerebrosidase or glucosylceramidase), resulting in progressive substrate accumulation and a spectrum of debilitating visceral, hematologic, and skeletal manifestations. Eliglustat is a novel oral substrate reduction therapy in development for adults with Gaucher disease type 1. This open-label Phase-3 trial (ENCORE, NCT00943111, Genzyme, a Sanofi company) evaluated eliglustat and imiglucerase in patients who had reached pre-specified therapeutic goals after ≥3 years of enzyme replacement therapy. We report efficacy data from the 12-month primary analysis period (PAP) and the first 12 months of the extension period during which all patients received eliglustat. **Methods**: Patients were randomized 2:1 eliglustat: imiglucerase. The primary efficacy endpoint was percent of patients remaining stable on a composite of spleen, liver, hemoglobin, and platelet parameters. As this was a non-inferiority trial, efficacy analyses were performed on the per-protocol population (99 eliglustat, 47 imiglucerase patients). **Results**: Eliglustat was non-inferior to imiglucerase: after 12 months, 85% of eliglustat and 94% of imiglucerase patients maintained all four goals (lower bound of 95% CI of difference [-17.6%] within the pre-specified [-25%] non-inferiority margin). One hundred and forty five (91%) of the 159 patients treated in this study completed 24 months of treatment. Preliminary 12-month extension data demonstrate continued stability in spleen volume, liver volume, platelet count and hemoglobin level in most of the 99/106 patients who continued on eliglustat and most of the 46/53 patients who received imiglucerase in the PAP and then eliglustat in the trial extension. Most adverse events were mild or moderate in severity. In general, the overall number, seriousness and relatedness of adverse events for all eliglustat-treated patients during day 1 to month 24 is similar to the profile for eliglustat-randomized patients in the PAP as well as similar to the safety profile in patients who switched from imiglucerase to eliglustat after 12 months. **Conclusions**: In the Phase 3 ENCORE study, most patients maintained clinical stability while on eliglustat for 12 or 24 months.

2283T

Assessment of Pompe Patients at Reference Center of Inborn Errors of Metabolism (CREIM). *P. Feliciano, C.S. Aranda, C.S.C. Mendes, M.A. Curiati, C.M. Gonçalves, M.H. Rand, A.M. Martins.* Centro de Referência em Erros Inatos do Metabolismo, Universidade Federal de São Paulo, São Paulo, SP, Brazil.

Objectives: Pompe disease (PD) is a lysosomal storage disorder caused by deficiency of alpha-1,4-glucosidase which leads to an accumulation of glycogen in lysosomes. The aim of this study is to the evaluate the profile of patients with PD who are treated at the Reference Center of Inborn Errors of the Universidade Federal de São Paulo - Brazil (CREIM). Methods: Medical records of patients with confirmed diagnosis of PD were analyzed. Results: Six patients, one male (16.5%) and five female (83.5%) were analyzed. One patient (16.5%) died at seventeen years due to respiratory complications. Patients were divided into two groups: onset of symptoms in childhood (CG) and initial symptoms in adulthood (AG). There were no patients in this sample with neonatal presentation. The current mean age was 36 years (range 11 - 66yr). The mean age of early onset symptoms was between 9.6 years (CG) and 40.3 years (AG) and the mean age of diagnosis was between 13.2 years (CG) and 45.1 years (AG), indicating a delay in diagnosis of 3.6 yr and 4.8 yr respectively. The mean age of enzyme replacement therapy (ERT) beginning was 12 and 48.3 years. The initial symptoms were weakness, shortness of breath and fatigue. One year after eow ERT, two patients showed improvement of respiratory drive, with significant reduction in CPAP parameters. The other three patients never used respiratory support. Conclusion: Pompe disease is a rare disorder that difficulty in patient diagnosis reflects in delayed treatment. Improvement in knowledge about inborn errors of metabolism, especially of treatable lysosomal storage disease is mandatory to anticipate the diagnosis and start treatment to reduce sequel and improve the quality of life of patients.

2284M

CYP2D6 phenotype-based dosing of eliglustat. *S. Turpault, Z. Meng, S. Wang, L. Von Moltke.* Genzyme, a Sanofi Company , Cambridge, MA.

Introduction: Gaucher disease type 1 (GD1) is a multi-systemic lysosomal storage disorder resulting from acid β-glucosidase deficiency. Eliglustat, an experimental oral substrate reduction therapy for adults with GD1, is metabolized mainly by CYP2D6, a polymorphic cytochrome P450 isozyme. Clinical studies used a dose-titration scheme (ranging from 50 to 150 mg BID) to ensure plasma eliglustat steady-state pre-dose concentrations (C_{trough}) above 5 ng/mL; however, this method has limitations in clinical practice. An alternative dosing regimen of 100 mg BID was evaluated for CYP2D6 intermediate and extensive metabolizers (IM/EM), who comprise approximately 90% of patients.

Methods: A pharmacokinetic/pharmacodynamic (PK/PD) modeling approach using population-PK model predicted plasma exposure parameters of eliglustat and observed efficacy from eliglustat clinical studies was conducted.

Results: In the Phase 3 randomized placebo-controlled study of treatment-naïve GD1 patients (ENGAGE, NCT00891202), most of the CYP2D6 IM/EM patients (17/19, 89.5%) were dosed at 100 mg BID (maximum dose allowed) based on C_{trough}. Modeling confirmed that the percent change in spleen volume (primary endpoint) was similar whether patients were dosed by C_{trough} (-27.8%, observed) or if all IM/EM patients had received a 100 mg BID dose (-31.9%). In the Phase 3 randomized controlled study of GD1 patients switching from enzyme replacement therapy to eliglustat (ENCORE, NCT00943111), 55 of 88 IM/EM patients (62.5%) were dosed at either 50 or 150 mg BID (maximum dose allowed) based on C_{trough}. Modeling predicted that the change in spleen volume was similar whether patients were dosed by C_{trough} (-5.96%, observed) or if all IM/EM patients had instead received a 100 mg BID dose (-6.63%) and that this dosing regimen would not result in any additional treatment failures. For all patients, projected eliglustat exposure at the 100 mg BID dose was within the range observed in the clinical studies (i.e., would not be expected to affect safety).

Conclusion: A simplified dosing regimen for eliglustat based on CYP2D6-genotype-predicted phenotype (100 mg BID for IM/EM patients) would achieve exposure, efficacy, and safety results similar to dosing based on plasma eliglustat concentrations.

2285T

Hematopoietic stem cell transplantation in a very young child with Hunter Syndrome: four year follow-up. *A.L. Barth[1], A.C. Esposito[1], C. Bomfim[2], M.L.C. Oliveira[3], F. Scalco[3], A.P. Costa[1], A.B. Reis[1], J.C. Llerena Jr[1], D.D.G. Horovitz[1].* 1) Instituto Fernandes Figueira/Fiocruz, Rio de Janeiro, Brazil; 2) Hospital de Clínicas de Curitiba - UFP, Paraná, Brazil; 3) LABEIM, Departamento de Bioquímica - UFRJ, Rio de Janeiro, Brazil.

Enzyme replacement therapy (ERT) in Hunter Syndrome (MPS II) does not cross the blood brain barrier, limiting results in neurological forms of the disease. Hematopoietic stem cell transplantation (HSCT) may be effective for the non-neuropsychological symptoms only, although better outcome may be expected in young patients. We describe the four-year follow-up of an MPS II child submitted to HSCT with umbilical cord blood cells at 70 days preceded by ERT. He was investigated due to positive family history; undetectable iduronate-sulphatase (IDS) activity and the familial p.R88H mutation were detected in amniotic fluid. At birth slight hyper-reactivity and lumbar gibbus, with corresponding L3-L5 abnormality on X-ray were noted. Abdominal and cerebral ultrasound, echocardiogram, ophthalmology, audiology, pulmonary function, brain & spine MRI were normal. Lysosomal storage in placental cells and pericytes were observed by electronic microscopy. IDS activity in plasma [1,2nmol/4h/ml (ref=122-463)] and leukocytes [4,3nmol/4h/mg/ptn (ref=31-110)] re-confirmed MPS II. He received 6 infusions of Idursulfase from age 10 days without adverse reactions. HSCT with umbilical cord blood from unrelated donor was performed at 70 days of age, without additional ERT. Urinary glycosaminoglycans began to decrease during ERT and were in normal range 3 months post-transplant. Engraftment after 30 days revealed mixed chimerism with 79% donor cells. IDS activity 30 days post-transplant was low in plasma (22nmol/4h/ml; ref=122-463) and normal in leukocytes (49nmol/4h/mg/ptn; ref=31-110); 270 days, 2, 3 and 4 years after transplant the same pattern was observed. At age 4 years growth charts and motor development are normal; he is hyperactive and language is delayed. Very mild lumbar gibbus and signs of dysostosis multiplex in hands and lumbar spine are present. Follow-up studies after HSCT in MPS II are scarce, lacking information on longer term effects. Patients with the severe phenotype do not benefit from transplant when gross developmental delay is evident. We decided to go through with the transplant since no reports for children so young were available, and as the only "curative" option. ERT was used as an attempt to prevent storage during the elapsed time from birth to transplant/engraftment. Whether transplantation prior to the onset of neurological symptoms would prevent a severe outcome has yet to be proven. Longer follow-up and reports of similar cases are recommended.

2286M

Inhibition of Hepatic Mitochondrial Metabolism during Systemic Immune Activation. *P.J. McGuire[1], S. Matsumoto[2], K. Saito[2], J. Senac[1], S. Cologna[3], T.N. Tarasenko[1].* 1) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) National Cancer Institute, National Institutes of Health, Bethesda, MD; 3) National Institute of Child and Health Development, National Institutes of Health, Bethesda, MD.

Infection plays a significant role in precipitating life-threatening acute metabolic decompensation in inborn errors of metabolism (IEM). Acute metabolic decompensation is defined as a functional deterioration in metabolic status and may result in a range of hepatic perturbations including hypoglycemia, acidosis and hyperammonemia. We hypothesized that hepatic mitochondrial metabolic adaptations that normally occur during the immune response to infection may not be tolerated in IEM. To simulate systemic immune activation, mice were injected with poly I:C, a dsRNA analogue, for 3 days. During this systemic immune activation, reduced amounts of pyruvate dehydrogenase and components of respiratory chain were detected by immunoblot. In addition, pyruvate dehydrogenase (PDH) was hyperphosphorylated, suggesting inhibition of activity. To address functional perturbations in mitochondrial metabolic pathways, we profiled in vivo hepatic mitochondrial function by 13C-methionine oxidation breath testing and MR spectroscopy using hyperpolarized 13C-pyruvate. Following poly I:C treatment, while methionine oxidation was unaffected, significant increases in 13C-lactate/13C-pyruvate and 13C-alanine/13C-pyruvate were seen, consistent with PDH inhibition. The above changes occurred in the absence of mitochondrial morphological and number changes by electron microscopy. By employing a strategy of immunomodulation targeting macrophages, PDH activity improved. Overall, our findings suggest that systemic immune activation has direct effects on hepatic mitochondrial metabolism and that modulation of the immune system may be a viable target of intervention. These findings may have implications for patients with mitochondrial disease and other IEM with significant mitochondrial dysfunction. In addition, our findings may have broader implications for multiple organ dysfunction syndrome due to systemic immune activation (i.e. the Systemic Inflammatory Response Syndrome or sepsis).

2287T

Genotype and Phenotype of Vietnamese patients with ornithine transcarbamylase (OTC) deficiency. *K.Ngoc. Nguyen[1], D.Chi Vu[1], M.Chi Nguyen[1], H.Viet Dau[1], T.Anh Ta[1], G.H Kim[2], H.W Yoo[2].* 1) Department of Endocrinology, Metabolism, Genetics, National Hospital of Pediatrics, Hanoi, Viet Nam; 2) Medical Genetics Center, Aasn Medica Center, Korea.

Ornithine transcarbamylase (OTC) deficiency, a partially dominant X-linked disorder, is the most common inherited defect of the urea cycle. The disease present variable severity affecting both males and females. Molecular diagnosis of OTC deficiency was the first time in Vietnam. **Objectives:** Analyse the phenotype and genotype of three Vietnamese patients with OTC deficiency. **Methods:** Three patients were diagnosed OTC with the criteria of hyperammonemia, hypocitrullinemia and orotic aciduria. 10 exons and their respective exon- intron boundaries of the OTC gene were sequenced using genomic DNA. **Results:** There were 2 boys with newborn - onset form (the age of onset was 2 days and 8 days of age) and 1 girl with late - onset form (the age of onset was 18 months of age) in our study. One of two cases with newborn - onset form had two older brothers died at 5 and 7 days of age due to unknown comma. The rest were the 1st child. The initial symptoms of two newborn - onset cases were poor feeding, irritability, then comma and apnea. The initial symptoms of the late - onset case was vomiting, convulsion, comma and right hemiplegia. Management of hyperammonemic crisis of two newborn - onset cases included glucose infusion, L arginine, L carnitine, hemofiltration, restricted protein diet therapy (metabolic formula). Management of one late - onset case included glucose infusion, L arginine, L carnitine, restricted protein diet therapy (metabolic formula). One case of newborn - onset and one case of late on - set form survived (1 year old and 6years old). The rest died at 8 days old. Analysis of OTC gene detected 3 different mutations in 3 cases: c.77G>A (p.Arg26Gln), c.298+5G>C (IVS3+5G>C), c.422G>A (p.Arg141Gln) heterozygous. **Conclusion:** Newborn - onset OTC deficiency occur in Vietnamese male patients and late - onset one occur in Vietnamese female patients. The outcome depends on the time of diagnosis.

2288M

Precise targeted gene correction of arginase-1 deficiency using single-stranded oligodeoxynucleotides with TALENs or the CRISPR/Cas9 system. *YY. Sin, M. Morales, CM. McCracken, CD. Funk.* Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario K7L 3N6 Canada.

Arginase-1 (ARG1) deficiency, a rare autosomal recessive disorder leads to hyperargininemia with progressive neurological impairment, growth retardation and infrequent episodes of hyperammonemia. This disorder is caused by mutations found throughout the eight exons and several splice sites of the *ARG1* gene located on chromosome 6q23. Currently, there is no cure and pharmacological treatment is limited. Our study aims to correct *ARG1* mutant alleles through nuclease-induced homologous recombination (HR) in the presence of homologous donor repair templates. Using human cells [HEK293, fibroblasts, induced pluripotent stem cells (iPSCs)], we employed engineered transcriptional activator-like effector nucleases (TALENs) and the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)/Cas9 system, designed *de novo*, to target and modify specific DNA sequences at the *ARG1* gene locus. Here, we show that these genome editing tools efficiently cleave chromosomal DNA in a precise, predictable and robust manner, with mutation frequencies (indels) of up to 50%. Moreover, we designed short single-stranded oligodeoxynucleotides (ssODNs) as donor repair templates, which carry a unique novel restriction enzyme site and two flanking homology arms on both sides of the target region. Through selection-independent targeting, these ssODNs introduce desired DNA sequence edits in exons 2, 3 and 7 in conjunction with the delivery of TALENs or CRISPR/Cas9 and a guide RNA (gRNA). Correctly targeted clones are being isolated via limiting dilution, PCR screening and/or RE digestion. Overall, our results demonstrate the versatility of using both TALEN and CRISPR/Cas9 systems to provide proof-of-concept for gene manipulation in human cells and sequence alterations within a defined locus can be achieved simultaneously by introducing ssODNs alongside. iPSCs have been generated from dermal fibroblasts obtained from individuals with ARG1 deficiency. It is anticipated that HR-mediated direct gene corrections in ARG1 deficiency patient-derived iPSC lines may prove useful for autologous cell-based therapies.

2289T

Sodium phenylbutyrate decreases plasma branched-chain amino acids in patients with urea cycle disorders. *L. Burrage[1], M. Jain[1], L. Gandolfo[2], B. Lee[1,3], SC. Nagamani[1], Members of the Urea Cycle Disorders Consortium.* 1) Human & Molec Genetics, Baylor College Medicine, Houston, TX; 2) University of South Florida, Tampa, FL; 3) Howard Hughes Medical Institute, Houston, TX.

Sodium phenylbutyrate (NaPB), a nitrogen-scavenging agent, is used to prevent hyperammonemia in patients with urea cycle disorders (UCDs). Previous reports involving small numbers of patients with UCDs have shown that treatment with NaPB can result in lower plasma levels of the branched-chain amino acids (BCAA) but this observation has not been studied systematically. From a large cohort of UCD patients (n=591) enrolled in Longitudinal Study of Urea Cycle Disorders, a multicenter study of the Urea Cycle Disorders Consortium, we evaluated whether treatment with NaPB leads to a decrease in plasma BCAA levels. In our study population, the median levels of all three BCAA (leucine, isoleucine, and valine) were significantly lower (~ 30 %; $p<0.001$) in patients treated with NaPB across all age groups ($p<0.001$). Generalized linear model analysis showed that NaPB use independently affected the plasma BCAA levels even when other important covariates that could influence BCAA levels were taken into account ($p<0.00001$). In addition, NaPB use was not associated with a decrease in the plasma levels of other essential amino acids which indicates that the decrease in BCAA levels is likely due to a specific effect of the drug on BCAA metabolism. Furthermore, NaPB use increased the risk for BCAA deficiency by 2-8 fold. Our study, in an unselected population of UCD subjects, is the largest to analyze the effects of NaPB on BCAA metabolism and conclusively shows that NaPB use is associated with lower plasma BCAA. Our findings have significant clinical implications. First, they indicate that plasma BCAA levels should be monitored in UCD patients who are treated with NaPB and opens avenues to explore whether BCAA supplementation could lead to better metabolic control. Second, they lend further support to the hypothesis that NaPB could be a potential therapy for patients with maple syrup urine disease (MSUD), a disorder of BCAA catabolism characterized by elevations of BCAA. We have shown that NaPB alters the phosphorylation status of branched-chain keto-acid dehydrogenase thus increasing its enzymatic activity and BCAA catabolism. Currently, a randomized double-blind placebo controlled study is ongoing to investigate whether NaPB could be an effective therapy for patients with MSUD.

2290M

MEDNIK syndrome: clinical and biochemical delineation of the copper metabolism phenotype and response to zinc therapy. *D. Martinelli[1,2], M. Ben Tanfous[3], C.A. Drouin[4], C. Dionisi-Vici[2], N. Braverman[3,5], S.G. Kaler[1].* 1) Section on Translational Neuroscience, Molecular Medicine Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, Bethesda, MD, USA; 2) Division of Metabolism, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy; 3) Genetic Department, Montreal Children's Hospital, QC, Canada; 4) Department of Dermatology, Centre Hospitalier Regional Grand-Portage, Riviere-des-Loups, QC, Canada; 5) McGill University-Montreal Children's Hospital Research Institute, QC, CA.

MEDNIK syndrome (MIM 609313) - an acronym for mental retardation, enteropathy, deafness, neuropathy, ichthyosis, keratodermia - is caused by mutations in the *AP1S1* gene, encoding the sigma 1A subunit of adaptor protein complex-1 (AP-1). AP-1 plays a crucial role in the intracellular trafficking of transmembrane proteins, including the copper-transporting ATPases, ATP7A and ATP7B. On this basis, we recently identified disordered copper metabolism in MEDNIK syndrome, and documented clinical and biochemical signs of Menkes (MIM 309400) and Wilson (MIM 277900) diseases in MEDNIK patients (Martinelli et al. Brain 2013). Severe loss of function mutations in *ATP7A* or *ATP7B* genes cause the complete Menkes and Wilson phenotypes, respectively. We have performed serial clinical and neurological evaluations in three subjects with MEDNIK syndrome under zinc therapy (approved for the treatment of Wilson disease because oral zinc limits copper adsorption at the intestinal level). All patients presented with hypocupremia, increased serum free copper level, increased urinary copper excretion, and liver copper accumulation. The hepatic component of the illness has been shown in one patient to respond to zinc and the other two patients are currently receiving this treatment. Interestingly, plasma and cerebrospinal fluid (CSF) neurochemical levels, reliable biochemical hallmarks of Menkes disease, were normal in the one subject studied to date. This patient's CSF copper levels are pending. Skin abnormalities have persisted in all three patients. The above findings pose new questions about the clinical and biochemical effects of AP-1 dysfunction and how the resultant copper metabolic abnormalities should be managed. Chronic zinc treatment may protect affected patients from hepatic copper overload, but some copper may be required for normal brain growth and maturation, as illustrated in patients with Menkes disease. In addition to identifying rational treatment approaches for a novel disorder of copper metabolism, further clinical and translational research studies of MEDNIK syndrome will provide insight about AP-1 cargo specificity and expand our understanding of how adaptor protein complex subunit diversity contributes to the exquisite control of transmembrane protein trafficking exerted by these molecules.

2291T

Spinocerebellar Ataxias with Hypogonadism: PNPLA6 Mutations Broadens the Phenotypic Spectrum of a New Inherited Neurometabolic Disorder. *C. Lourenco[1], C. Sobreira[1], M. Hirano[2], S. Zuchner[3], W. Marques Jr[1].* 1) Neurogenetics, Univ Sao Paulo, Ribeirao Preto, Sao Paulo, Brazil; 2) Columbia University Medical Center, New York, USA; 3) University of Miami, Miller School of Medicine Miami, USA.

OBJECTIVES: To present the clinical data and molecular/biochemical studies of nineteen Brazilian patients with cerebellar ataxia and hypogonadism. BACKGROUND: The association between cerebellar ataxia and hypogonadism was first described in four sibs by Holmes in the beginning of the XX century, and has since become known as Holmes type ataxia. Since then, several syndromes with hypo/hypergonadotrophic hypogonadism and ataxia have been published, with remarkable clinical heterogeneity among them. DESIGN/ METHODS: All patients were evaluated in a neurogenetics clinics by geneticists, neurologists and endocrinologists. Brain MRI, ophtalmological exam, EMG/NCV, hormone and biochemical tests, muscle biopsy with respiratory enzyme assays and coenzyme Q10 measurement, molecular tests (Friedreich ataxia and SCAs 1, 2, 3, 6 and 7) were performed. Exome sequencing was performed in selected patients with a consistent phenotype. RESULTS: All patients had cerebellar ataxia, but the age of the onset was variable; it was worthy to note that ten patients had early onset ataxia . Consanguinity was noted in three families; 12 patients had hypergonadotropic hypogononadism. Intellectual Disability was seen in five patients with hypergonadotropic hypogonadism. None of the patients had chromosomal anomalies. Optic atrophy and retinochoroidal degeneration were found in five patients; axonal neuropathy was present in five patients. Cerebellar atrophy with pons or prominent vermis involvement was a constant feature. Coenzyme Q10 deficiency and CDG Ia were confirmed diagnosis in five patients. Boucher-Neuhauser syndrome (BNS) was identified in other 4 patients with mutation in the PNPLA6 gene identified. CONCLUSIONS: The association between cerebellar ataxia and hypogonadism comprise heterogenous entities. Screening for CDG and Coq10 deficiency should be done in such patients as a part of the work-up investigation. Identification of the molecular basis of BNS widens our comprehension of this rare group of the disorders, with involvement of glycosphingolipids metabolism in its physiopathology.

2292M

DNA diagnostics of MIDD and MELAS syndromes in Slovakia. *D. Gasperikova[1,2], M. Skopkova[1], I. Masindova[1], L. Valentinova[1], J. Stanik[1,3], L. Varga[4], M. Huckova[1,2], M.I.D.D. Study Group[1], M. Profant[4], I. Klimes[1].* 1) Diabgene Laboratory, Institute of Experimental Endocrinology, Bratislava, Slovak Academy of Sciences, Bratislava, Slovakia; 2) Centre for Molecular Medicine, Slovak Academy of Sciences, Bratislava, Slovakia; 3) First Department of Pediatrics, School of Medicine, Comenius University, Bratislava, Slovakia; 4) First Department of Otorhinolaryngology, School of Medicine, Comenius University, Bratislava, Slovakia.

Background and aims. Diabetes and deafness - the leading clinical features of MIDD (Maternally Inherited Diabetes and Deafness) arise from mutations in mtDNA, most often m.3243A>G. This mutation leads to different clinical symptoms according to heteroplasmy levels in different tissues. Usually, the first presentation of MIDD is a sensorineural hearing loss emerging in adolescence; the diabetes mostly develops between 20th-40th year of life. More severe clinical presentation of the mutation is the MELAS syndrome (Mitochondrial Myopathy, Encephalopathy, Lactic Acidosis and Stroke-like episodes). The aim of our study was to search for m.3243A>G mutation among patients sent for MIDD or MODY testing, who lack mutations in the most common MODY genes (i.e. GCK, HNF1A or HNF4A). Materials and methods. Unrelated probands from 257 families fulfilling at least one of the following criteria, i.e. matrilinear inheritance, diabetes plus hearing impairment, diabetes development after 20th year of life, or progressive hearing loss, were tested for m.3243A>G mutation by the RFLP and/or Real-Time PCR. The heteroplasmy was evaluated for peripheral blood and/or buccal mucosa. DNA testing was also extended to the family members of probands carrying the mutation. Results. The m.3243A>G mutation was found in 18 patients from 8 families (3%). Probands' phenotypes varied from diabetes as the sole symptom to a complex picture of the MELAS syndrome (in one proband). Diabetes or impaired glucose tolerance developed all of the probands (diabetes onset ranged from 21 to 52 years), but only 3 of 10 relatives with the mutation. Five probands (62.5%), and 4 (40%) of the relatives with the mutation had hearing impairement. The heteroplasmy was higher in buccal swab samples compared to the peripheral blood (27.9±18.6% versus 12.7±17.3%). In one case, the heteroplasmy was detected in the buccal DNA only, while the blood DNA samples were repeatedly negative. Conclusion. Among 257 probands with the clinical suspicion on MIDD or MODY, 8 (3%) had the m.3243A>G mutation of the mitochondrial DNA. At the time of testing, only 5 (62.5%) of the probands were diagnosed with typical combination of symptoms, i.e. diabetes and hearing loss. Therefore, DNA testing for MIDD seems to be reasonable also in diabetes patients without hearing impairement, particularly using DNA from the buccal mucosa. Supported by APVV 0187-12, „KCMM" (ITMS 26240220071).

2293T

Screening of Adults for Cholesterol ester storage disease (CESD). *M. Beal, D. Roy, J. Kaplan, M. Taylor.* University of Colorado Denver - Anschutz Medical Campus, Aurora, CO.

Background: Cholesterol ester storage disease (CESD) due to lysosomal acid lipase deficiency represents the adult presentation of the more severe and uniformly fatal Wolman disease. The development of enzyme replacement therapy (now in clinical trials) motivates efforts to identify latent cases that may benefit from treatment. To this end, some have recommended that a combination of low high-density lipoprotein (HDL) and high liver aspartate aminotransferase (AST) may be markers of CESD. **Methods:** We undertook a chart review of medical records from a large academic medical center adult hospital of patients with HDL < 15 (mg/dl) and AST > 39 (IU/L) values collected on the same day between 2006 and 2010. Paper and electronic medical records were abstracted by two medical students and evaluated for clinical history, known diagnoses, imaging results and any biopsy reports. **Results:** 104 patients met the HDL and AST criteria (74 men and 30 women; mean age 49.9 +/- 13.6 years). The mean HDL and AST values for the cohort were 11.4 mg/dl and 377.7 IU/L, respectively. No cases of confirmed CESD were found in this cohort, although no patients had evidence of CESD testing performed. Coronary artery disease and high-cholesterol diagnoses were found in 16.3% and 36.5% of patients, respectively. The most common symptoms reported were: abdominal pain (38.5%), joint pain (16.3%), diarrhea (25.0%), nausea (28.8%), shortness of breath (25.0%), and vomiting (26.0%)Potentially confounding diagnoses, common to the cohort, included Alcoholism (20.1%), 'hepatitis' (30.8%), and HIV/Aids (27.9%). Abdominal ultrasound showed hepatomegaly in 32% of patients imaged (27/84). Twenty-six patients had been referred for liver biopsy and 15 of those were available for review by our hepatopathologist (JK). None of the biopsies showed clear pathological features of CESD and only one had minimal evidence of microvesicular steatosis. As DNA was not available from this chart-review cohort, no assessment of LIPA gene mutation status was performed. **Conclusions:** A Low HDL and high AST screen identified a heterogeneous cohort of patients unlikely to have CESD. Hepatic biopsy in a subset of our cohort did not reveal changes consistent with a high proportion of undiagnosed CESD. The addition of exclusion criteria (alcoholism, HIV/Aids, and hepatitis) might increase the sensitivity of these screening criteria by removing confounding diagnoses.

2294M

Meta-analysis of antibody titers, safety, and treatment outcomes in MPS I patients receiving enzyme replacement therapy (ERT) with laronidase (Aldurazyme) in clinical studies. *S. Fallet, Y. Xue, S. Richards, A. Mahmood, G. Cox.* Genzyme, a Sanofi company, Cambridge, MA.

Objectives: Laronidase (Aldurazyme®) is an enzyme replacement therapy approved for the treatment of patients with mucopolysaccharidosis I (MPS I), a genetic disorder caused by deficiency of α-L-iduronidase and subsequent accumulation of glycosaminglycans (GAG) throughout the body. Most MPS I patients treated with laronidase in clinical trials have developed anti-laronidase antibodies. Theoretically, antibodies might alter the biological effects of laronidase by inhibiting its activity, changing its turnover, limiting its uptake into target tissues, and/or causing an allergic reaction. The objective of this meta-analysis was to determine if relationships exist between anti-laronidase antibodies and clinical outcomes, hypersensitivity reactions, and urinary GAG reduction in MPS I patients receiving laronidase. Methods: Data from 73 patients enrolled in 4 clinical trials of laronidase were used for this meta-analysis. Patients received treatment with the labeled dose of laronidase (0.58 mg/kg/wk) for 26 to 208 weeks. Efficacy parameters included 6-MWT, % predicted FVC, liver volume, range of motion, left ventricular mass, investigator global assessments, and uGAG levels, as appropriate, in individual studies. Safety and immunogenicity parameters included infusion-associated reactions (IARs), anti-laronidase IgG antibodies (seroconversion and time to seroconversion, peak titer, overall exposure over time), enzyme activity inhibition and enzyme cellular uptake inhibition assays, and hypersensitivity testing (i.e. IgE, complement activation, and serum tryptase). Results: Based on the analysis of selected clinical outcome measures, there was no apparent correlation between anti-laronidase IgG antibodies and efficacy outcomes. Likewise, there was no apparent relationship between antibodies and the occurrence of potential allergic reactions or IARs. There was, however, a statistically significant inverse correlation between uGAG reduction and antibody titer (p=0.0001), consistent with individual study results. Conclusions: This meta-analysis describes the most comprehensive, systematic review of immunogenicity data from clinical studies of Aldurazyme (including long-term follow-up), and involves patients with all MPS I phenotypes. Although the presence of antibodies correlated with impaired urinary GAG clearance, no impact on the clinical efficacy or safety of laronidase was observed. This study was funded by the Genzyme/BioMarin Joint Venture.

2295T

Best Policy for helping LSD patients and Families: Collaboration between Charity Foundation, MOH, and pharmaceutical companies. *F. Hadipour[1,2], A. Hadipour[2,3], A. Torabi[2,3], Z. Hadipour[1,2], Z.Hadipour.* 1) Medical Genetics, Sarem Cell Research Center & Hospital, Tehran, Tehran, Iran; 2) YASNA Charity Foundation, Tehran, Iran; 3) Golden Daric CO, Tehran, Iran.

The lysosomal storage disorders (LSDs) are due to deficiencies of lysosomal enzymes caused by mutations of genes that encode the enzyme, proteins and related cofactors. Lysosomal enzymes degrade most biomolecules. This process is crucial for cells. LSDs result in accumulation of products in lysosomes. More than 45 LSDs are known, and occur in about 1 in 5,000-7,700 live births. LSDs have recessive inheritance. The presence of "founder" mutation in Middle East with a high degree of consanguinity is expected to lead to high prevalence of LSDs. Treatment varies across the LSDs. The ERT of LSDs is expensive comparing with the minimum wage of Iranian; so most of patients could not afford it. Charities play a vital role in facilitating treatment of rare disease especially LSD. Charities exist to help people, serve in areas where there are little or no resources available, and provide aid by educating, setting up programs, providing necessary medical supplies. YASNA is a non-profit, non-political, and non-governmental charity supported by people. YASNA strive to care for patients with rare disease. Main goal of YASNA include: Diagnosis and treatment, prevention, through improvements in public health, and education. When Registration and subsidizing of ERT for each kind of LSD is done in MOH; ERT would be provided for patients in special centers. Iranian MOH subsidized 90% of the price of ERT for LSD. But remaining 10 % should be paid by patients. According to The minimum wage of Iranian which is 248 $ per month; most of patients could not afforded to pay it, and they discontinued their treatment. During 3.5 years YASNA charity Foundation could covered 50% of price of ERT for 70 MPS type 1 patients; and 70% of it for 45 pompe patients whom are living in territory of IRAN(for Iranian or resident in Iran). YASNA's missions are providing access to medical care, and afforded for them regular counseling and training meeting, Genetic counseling, Genetic testing and sometimes prenatal diagnosis. For example genetic testing is done for 12 / 160 patients of Iranian Cystinosis. For countries with situation like IRAN, high rate of consanguinity marriage leads to high rate of rare diseases; and The minimum wage is less than price of ERT per month. The best solution for helping Patients and families is collaboration between pharmaceutical companies, MOH, and charity foundation. MAHAK foundation & YASNA foundation are 2 good examples of such collaboration in Iran.

2296M

Antisense U1-snRNAs-based correction of g.9273C>T and g.9331G>A *GLA* deep intronic mutations which cause Fabry disease. *L. Ferri[1,2], A. Caciotti[2], G. Covello[3], D. Perrone[4], R. Guerrini[1,2], M.A. Denti[3], A. Morrone[1,2].* 1) Department of Neurosciences, Psychology, Pharmacology and Child Health (NEUROFARBA), University of Florence, Florence, Italy; 2) Paediatric Neurology Unit and Laboratories, Neuroscience Department, Meyer Children's Hospital, Florence, Italy; 3) Centre for Integrative Biology, University of Trento, Trento, Italy; 4) Department of Biology and Evolution., University of Ferrara, Ferrara, Italy.

Mutations in the *GLA* gene cause enzyme deficiency of alpha-galactosidase A (α-GAL A) which leads to the X-linked Fabry disease, a lysosomal storage disorder of the glycosphingolipids metabolism. *GLA* produces two transcripts, the major mRNA which encodes α-GAL A and a minor mRNA whose function is unknown. The minor mRNA isoform arises by an alternative splicing and carries a criptic exon between exons 4 and 5 which contains a premature stop codon. Two *GLA* deep intronic mutations, the g.9273C>T and g.9331G>A, have been described being the cause of a pathological unbalanced expression of these two *GLA* transcripts, leading to α-GAL A deficiency and Fabry disease. Antisense-induced Exon Skipping represents a promising approach to correct such kind of mutations and can be obtained by the use of antisense oligonucleotides (AONs) or antisense modified U1snRNAs (U1asRNA). Hence, our research focuses on the setting up of an Exon Skipping based approach to correct the deep intronic mutations of the *GLA* gene. We designed and produced a set of AONs and of U1asRNAs specific to silence both g.9273C>T and g.9331G>A mutations. We also designed three *GLA* minigenes to express the two studied mutations and the wild-type. We tested our antisense molecules against the minigenes by transfection of COS cells. We evaluated the effect by α-GAL A enzyme assay, by Real-Time PCR and by Western Blot analysis. U1asRNAs that we designed promoted the recovery of the α-GAL A activity up to the wild-type level for both g.9273C>T and g.9331G>A *GLA* mutations. Western blot analysis confirmed an increase in the amount of α-GAL A protein produced as a consequence of the treatment with antisense molecules. The analysis of transcripts showed the antisense-induced exon skipping of the criptic exon with an increment of the major *GLA* mRNA. Our study suggests that a new therapeutic strategy based on antisense molecules can be developed and potentially used in the correction of splicing defects in Fabry disease. *Authors wish to express their gratitude to: Fondazione Meyer ONLUS, Florence, Italy (Young Researcher grant to Lorenzo Ferri); Fondazione Meyer ONLUS and Regione Toscana, Italy (grant POR CRO FSE 2007-2013); Italian Ministry of Health and Regione Toscana (Ricerca Finalizzata-2011-02347694); AMMeC Foundation, Italy..*

2297T

Nutrient Intake and Growth Patterns among Adult and Pediatric Subjects with Urea Cycle Disorders (UCDs) Participating in Glycerol Phenylbutyrate (GPB) Clinical Trials. *D. Hook[1], G.A. Diaz[2], B. Lee[3], J. Bartley[1], N. Longo[4], W. Berguist[5], C. Le Mons[6], I. Rudolph-Angelich[7], M. Porter[7], B.F. Scharschmidt[7], M. Mokhtarani[7].* 1) Miller Children's Hospital , Long Beach , CA; 2) Icahn School of Medicine at Mt Sinai, New York, NY; 3) Baylor College of Medicine, Houston, TX; 4) University of Utah, Salt Lake City, UT; 5) Stanford University, Palo Alto, CA; 6) National Urea Cycle Disorders Foundation, Pasadena, CA; 7) Hyperion Therapeutics, Brisbane, CA.

Objective: To analyze protein and caloric intake among adult and pediatric UCD patients as well as growth in pediatric patients participating in the GPB (HPN-100, RAVICTI®) clinical trials. Methods: Forty-four adult patients completed a 28 day, randomized, double-blind, placebo-controlled crossover comparison of 24-hr ammonia exposure on GPB vs. sodium phenylbutyrate (NaPBA). Weekly 3-day diet histories were collected. Prescribed and actual protein and caloric intakes were compared to those for healthy individuals (HI) derived from the National Health and Nutrition Examination Survey and current recommendations for UCD patients (UCD-REC). Forty-nine pediatric patients ages 2-mo to 17-yrs received GPB for 12 months at a dose equivalent to their previously prescribed NaPBA dose. Prescribed protein, calorie, height, weight and BMI were assessed at enrollment (baseline) and throughout the 12 month open-label study. Results: In adult patients, mean protein intake was 53-56% of that for HI, greater than UCD-REC, and accounted for 10% of caloric intake vs. 15% for HI. Total caloric intake was ~70% of that for HI and 20-25% less than UCD-REC. There was a positive correlation between blood urea nitrogen (BUN) and total protein intake (r=0.55; p=0.0003). In pediatric patients, mean prescribed protein and caloric intake were 40% to 64% and 92 to 100% of HI and 54% to 97% and 68% to 87% of UCD-REC, respectively. Mean height, weight and BMI z-scores were within 2 standard deviations of normal population at baseline and did not change during the study. BMI correlated negatively with prescribed protein or calories expressed as g/kg/day or kcal/kg/day, respectively, in both adults (protein; r=-0.641; p = 0.007; calories; r=-0.798 p<0.001) and peds (protein: -0.329; p=0.022; calories: r=-0.362 p=0.024), but positively and weakly when expressed as total daily protein (g/day) or calories (kcal/day) in both populations. After 12 months of GPB dosing, the percentage of patients with normal BMI (5-85% percentile) increased from 46% to 59%, those considered obese (>85% percentile) decreased from 50% to 39% and those considered underweight (<5% percentile) decreased from 4% to 2% compared to baseline. Conclusions: Both pediatric and adult UCD patients ingested less protein and fewer calories than HI and more protein, but fewer calories than UCD-REC. Patients' BMI tended to normalize during this 12 month open label and uncontrolled study.

2298M

Predictive Value of Blood Ammonia and Glutamine to Hyperammonemic Crises in Patients with Urea Cycle Disorders. *B. Lee[1], G.A. Diaz[2], W. Rhead[3], U. Lichter-Konecki[4], A. Feigenbaum[5], S.A. Berry[6], C. Le Mons[7], J. Bartley[8], N. Longo[9], S.C. Nagamani[1], W. Berquist[10], R.C. Gallagher[11], D. Bartholomew[12], C.O. Harding[13], M.S. Korson[14], S.E McCandless[15], W. Smith[16], S. Cederbaum[17], D. Wong[17], J.L. Merritt II[18], A. Schulze[5], J. Vockley[19], D. Kronn[20], R. Zori[21], M. Summar[4], D.A. Milikien[22], M. Marino[13], D.F. Coakley[23], M. Mokhtarani[23], B.F Scharschmidt[23].* 1) Baylor College of Medicine, Houston, TX., USA; 2) Icahn School of Medicine at Mt Sinai, New York, NY; 3) Medical College of Wisconsin, Milwaukee, WI; 4) Children's National Medical Center, Washington, DC; 5) University of Toronto, Toronto, Canada; 6) University of Minnesota, Minneapolis, MN; 7) National Urea Cycle Disorders Foundation, Pasadena, CA; 8) Miller Children's Hospital, Long Beach, CA; 9) University of Utah, Salt Lake City, UT; 10) Stanford University, Palo Alto, CA; 11) Children's Hospital Colorado, Aurora, CO; 12) Nationwide Children's Hospital, Columbus, OH; 13) Oregon Health Sciences, Portland, OR; 14) Tufts Medical Center, Boston, MA; 15) Case Western Reserve Univ. Medical Center, Cleveland, OH; 16) Maine Medical Center, Portland, ME; 17) University of California, Los Angeles, CA; 18) Seattle Children's Hospital, Seattle, WA; 19) Univ of Pittsburgh, Pittsburgh, PA; 20) Westchester Medical Center, Westchester, NY; 21) University of Florida, Gainesville, FL; 22) Accudata Solutions Inc., Lafayette, CA; 23) Hyperion Therapeutics, Brisbane, CA.

Limited information is available regarding the optimal timing of blood ammonia measurements in patients with urea cycle disorders (UCDs), target ammonia levels, or the relationship of ammonia or glutamine to hyperammonemic crises (HAC). Methods: More than 1000 blood samples obtained from 100 UCD patients chronically treated with sodium phenylbutyrate (NaPBA) who participated in short and long-term glycerol phenylbutyrate (GPB) trials were analyzed with respect to the relationship between fasting ammonia, glutamine and occurrence of HACs over 12-month treatment with GPB. Results: Nineteen of 100 UCD patients (49 pediatric, 51 adults) treated with GPB for 12 months experienced a total of 27 crises compared to 30 patients reporting 53 crises on NaPBA during the 12 months before enrollment. There was a strong positive relationship between fasting ammonia and 24-hr total ammonia exposure (r = 0.764, p < 0.001) and a weak but significant correlation between baseline ammonia and glutamine levels (r=0.27; p = 0.008) that was primarily driven by patients with high ammonia. Baseline fasting ammonia was significantly lower in patients without vs. with a crisis [mean (SD) of 26 (18) μmol/L vs 42(24) μmol/L (p=0.0013)], whereas glutamine was not [mean (SD) of 782 (246) vs 732(186) μmol/L; (p=0.150)]. Similarly, during 12 months of GPB dosing, ammonia values were significantly lower in patients without vs. with a crisis whereas glutamine values were not. For every 10 or 25 μmol/L increase in total ammonia exposure over 12-months, the relative risk of a HAC increased, by 50% and >200% (p<0.0001) respectively. Among patients ≥ 6 years of age (N=77) in whom fasting ammonia can be reliably measured, the relative risk of a HAC was 20 times higher (p = 0.009) in patients with fasting ammonia ≥ 1.0 ULN vs. < 0.5 ULN. Patients with or without a crisis did not differ in terms of other baseline variables often associated with severity of illness such as age, UCD subtype and gender. Conclusions: As compared with glutamine, fasting ammonia correlates strongly and positively with total daily ammonia exposure as well as the long-term risk and rate of HAC. These findings suggest that tight ammonia control (<0.5 ULN) deserves further study as an approach to improving long-term outcome in UCD patients.

2299T

Gene trap of *NDUFS4* in mouse is a viable model of mitochondrial complex I deficiency. *B. Graham[1], T. Donti[1], Y. Lai[2], M. Ge[1], S. Ather[3], R. Lucero[1,4], R. Paulter[3], J. Noebels[1,4], X. Wehrens[3], W. Craigen[2].* 1) Dept Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dept Pediatrics, Baylor College of Medicine, Houston, TX; 3) Dept Molecular Physiology & Biophysics, Baylor College of Medicine, Houston, TX; 4) Dept Neurology, Baylor College of Medicine, Houston, TX.

Mitochondria are essential for many fundamental cellular processes, including energy production, ROS metabolism, Ca2+ homeostasis, and apoptosis. Dysfunctional mitochondria frequently result in multiorgan disease that can include neuronal, cardiac and muscular pathology. Mitochondrial complex I is comprised of approximately 45 subunits and is most often implicated in mitochondrial disease. NADH:ubiquinone oxidoreductase iron-sulfur protein 4 (NDUFS4) is one of the nonenzymatic, nuclear-encoded subunits of complex I that is essential for assembly. Defects in *NDUFS4* cause complex I deficiency and a Leigh-like neurological phenotype in humans that typically result in death within 3-16 months of birth. Heart dysfunction is also a common phenotype in primary mitochondrial disease, typically presenting in the context of multi-systemic disease and can manifest as a cardiomyopathy and/or as cardiac arrhythmias. From the International Gene Trap Consortium, mouse embryonic stem (ES) cells that contain a gene trap insertion in *Ndufs4* were obtained and used to generate transgenic animals. Homozygous mutant mice show reduced levels of NDUFS4 protein in various tissues by Western blot. While *Ndufs4* mutants exhibit partial embryonic lethality, surviving adult mutants are able to survive for at least eighteen months. *Ndufs4* mutant animals exhibit partial complex I deficiency, increased oxidative stress in brain and heart, and defects in cellular and mitochondrial respiration. They also exhibit neuronal dysfunction manifested by constitutively abnormal electroencephalograms (EEGs) associated with intermittent myoclonic seizures, as well as increased sensitivity to kainate-induced seizures. In addition, three-month-old mutant and matched wild type *Ndufs4* mice were analyzed by electrocardiography and transthoracic echocardiography. These mutant mice exhibited cardiac arrhythmias combined with left ventricular dysfunction. Mutant animals also exhibited a 30% increase in heart weight normalized to body weight. In summary, *Ndufs4* mutant mice homozygous for a gene trap insertion represents a viable model of mitochondrial complex I deficiency that exhibits increased oxidative stress with neurological and cardiac dysfunction. Comparing the cardiac phenotype of this constitutive *Ndufs4* mutant model with a previously published cardiomyocyte-specific *Ndufs4* conditional knockout (PMID 23931755) suggests that a non-cell autonomous mechanism may be playing a pathogenic role.

2300M

A Molecular Genetic Study of Indian Patients with Glycogen Storage Disorders reveals Six Novel Mutations. *S. Shetty[1], V. Tejaswini[1], M. Bhat[1,2], R.K. Sanjeev[4], S. Jagdeesh[3].* 1) Centre for Human Genetics, Bangalore, India; 2) Indira Gandhi Institute of Child Health, Bangalore, India; 3) Fetal Care Research Foundation, Chennai, India; 4) Base Hospital, Delhi, India.

Glycogen Storage Disorders (GSDs) are a set of autosomal recessive conditions caused by mutations in genes that govern different aspects of the glycogen metabolic pathway. Depending on the enzyme that is deficient, the GSDs are classified into 12 different types. GSDs have been a very complex set of diseases to deal with due to the variation in the presenting clinical features, which tend to overlap between types. There are differences not only in the reports of incidence rates in the different types of GSDs but also in establishing which type might be the most commonly occurring/observed GSD. GSD-I, III and VI seem to be the most common among the GSDs and their overlapping phenotypes leads to their grouping which then requires us to take on a differential diagnosis approach with genetic testing. GSD-1 is further divided into Ia and Ib. GSD-Ia is caused by a deficiency of the microsomal glucose-6-phosphatase (*G6PC*) in the liver, kidney, and intestinal mucosa and is indicated as such by an excessive accumulation of glycogen in these organs. Patients with GSD-Ia often suffer from hepatic adenoma/hepatoma, renal insufficiency, and pulmonary hypertension. Distinguishing Ia from Ib necessitates an assay of G6Pase activity in a liver biopsy. GSD-Ib, caused by mutations in the Glucose-6-phosphate translocase (G6PT or *SLC37A4*) shows similar symptoms to GSD-Ia with a few patients exhibiting neutropenia, inflammatory bowel disease and/or recurrent infections. Deficiency of the Glycogen debranching enzyme causes GSD-III. The variable phenotype encountered is explained by differences in tissue expression of the defective enzyme coded by the *AGL* gene and when deficient in both liver and muscle, it is sub-grouped as GSD-IIIa but when the enzyme deficiency is observed only in the liver it is considered as type IIIb. Our initial studies on a set of forty Indian patients, who could all come under one of the three GSD groups named above, have revealed six novel mutations. Four in patients with GSD-1a, one in a patient with GSD-Ib and one with GSD-III. Genetic testing has been fraught with problems encountered in classifying GSDs, as clinical features are not clearly distinguishable between types and genotype-phenotype correlations remain largely indistinct. Our study also illustrates that the Indian population may harbour a much greater variation in mutations as we did not find any of the common or hitherto reported mutations.

2301T

Stable Isotope Breath Tests to Assess Metabolite Flux in Methylmalonic Acidemia (MMA). *E.A. Harrington[1], I. Manoli[1], Y. Ktena[1], S. Smyth[2], J. Hattenbach[2], J. Senac[1], J. Sloan[1], K. Chen[2], C.P. Venditti[1].* 1) GMBB, NHGRI, Bethesda, MD; 2) Clinical Endocrinology Branch, NIDDK.

Isolated methylmalonic acidemia (MMA), caused primarily by a defect in the mitochondrial enzyme methylmalonyl-CoA mutase (MUT), is characterized by metabolic instability, multiorgan pathology, and a poor prognosis for long-term survival. Elective liver (LT), kidney (KT) or combined liver kidney transplantation (LKT) are used to stabilize severely affected patients. Secondary mitochondrial dysfunction, characterized by aberrant ultrastructure and decreased complex IV activity, is observed in patient and murine tissues. Mice that express the *Mut* gene under the control of a muscle-specific creatine kinase promoter ($Mut^{-/-}$;$Tg^{INS\text{-}MCK\text{-}Mut}$) were used to probe metabolism of various ^{13}C-isotopomers. These animals are rescued from neonatal lethality but display severe metabolic perturbations, growth retardation, and a hepatorenal mitochondriopathy. 1-^{13}C-propionate oxidation was used to assess *in vivo* Mut activity, while 1-^{13}C-methionine, glycine, α-ketoisocaproic acid and octanoate oxidation were studied as secondary pathways dependent on mitochondrial function. $Mut^{-/-}$;$Tg^{INS\text{-}MCK\text{-}Mut}$ mice displayed decreased propionate (P=0.003), methionine (P=0.03) and glycine (P=0.008) oxidation. We subsequently administered 1-^{13}C-propionate to 6 controls (age range 35-62y) and 16 patients with isolated MMA (*mut* N=12, *cblA* N=3, *cblB* N=1, age range 9-41y). Resting energy expenditure was measured and serial breath samples were collected following an oral or G-tube bolus of the isotopomer. Decreased 1-^{13}C-propionate oxidation was observed in all MMA patients vs. controls (P<0.0001). The most severe *mut* patients, who carried two stop mutations (N=3), showed almost no metabolism of label, despite isolated KT in one patient. LKT recipients (N=2) showed a complete restoration of oxidation rates to control levels (P=0.004 compared to non-transplanted patients), while only minimal oxidation was evident in a recipient after a auxiliary liver allograft and KT. B12-responsive patients with KT (N=2) showed near normal recovery. ^{13}C breath testing with varied isotopomers represents a non-invasive method to assess whole body *in vivo* MUT activity and resulting secondary metabolic perturbations, and could be used as prognostic or outcome measure to study interventions, such as B12 supplementation, transplantation or gene therapy in MMA patients. These methods could be readily applied to other disease states where branched-chain amino acid oxidation and mitochondrial function are disturbed.

2302M

Diagnosis of Lesch-Nyhan syndrome in an 8-year-old male referred by nephrology for elevated serum uric acid. *L. Pugliesi, P. Gupta, L. Ely.* Genetics Division, St. Joseph's Regional Medical Center, Paterson, NJ.

Lesch-Nyhan syndrome (AHC [MIM 300322]) is an x-linked genetic condition characterized by elevated uric acid production, neurological abnormalities, and distinct behavioral changes. We present an 8 year old male initially referred to our clinic by nephrology for elevated serum uric acid levels. Medical history is significant for cognitive delay and diagnosis of cerebral palsy made at eleven months of age. The patient displayed finger picking and head banging at four years of age and first presented with lip and finger biting at seven years of age. He was consequently diagnosed with an anxiety disorder due to the self-injurious behaviors. He utilizes a wheel chair and elbow restraints and is post-surgical bilateral hip osteotomy. Family also reports a history of abnormal sleep patterns. Nephrology evaluation noted intermittent discoloration of urine, urethral discharge, and abnormal kidney ultrasound identifying echogenic kidneys with acoustic shadowing in the cortical and pelvic area. Physical evaluation was significant for dysmorphic features, mutilated lower lip, removed incisors, mild scoliosis, tight heel cords, increased tone, abnormal deep tendon reflexes, hyper-reflexive elbows, and bilateral ankle clonus. Family history reveals recurrent miscarriage in his mother, and a maternal uncle with intellectual disability of unknown etiology. Upon genetics evaluation, Hypoxanthine-Guanine Phosphoribosytransferase enzyme testing and *HPRT1* (AHC [MIM 308000]) sequence analysis were completed both confirming a diagnosis of Lesch-Nyhan Syndrome. We present this case to demonstrate how misdiagnosis of individual symptoms may lead to a delayed diagnosis of an over encompassing genetic syndrome. We also aim to bring awareness to the common findings associated with Lesch-Nyhan syndrome to aid health care providers in identifying early signs and symptoms and thus to achieve a diagnosis in affected individuals.

2303S

The UDP Self Study Short Course for Genome Wide Analysis. *T.C. Markello[1,2], C.F. Boerkoel[1,2], W.A. Gahl[2], E. Valkanas[1,2,3], E.D. Flynn[2], W.P. Bone[2], A.E. Links[2], C.J. Markello[2], P.J. Pemberton[2], D.R. Adams[1,2], NIH Intramural Sequencing Center, NIH Genomic core, and UDP Clinicial Team.* 1) Med Gen Branch, NIH/NHGRI, Bethesda, MD; 2) NIH Undiagnosed Diseases Program, Office of the Director Common Fund and NHGRI, NIH, Bethesda, MD; 3) Clinical Director, NHGRI/NIH, Bethesda, MD.

Since 2008 the NIH Undiagnosed Diseases Program has found many types of Mendelian and non-Mendelian genetic and molecular mechanisms for human diseases in patients referred to the program. These include: Homozygous recessive cases of identity by descent with parental consanguinity, compound heterozygous recessive alleles including double point mutation pairs, point mutation/deletion pairs and double deletion null allele pairs, unique de novo dominant new mutations, de novo dominant with transmission to the third generation mutations, moasicism including both fixed ratio and progressive telomere shortening, duplications, and a form of variant discovery we call extreme novel findings. These discoveries all occurred in single nuclear families with one exception and typically with only one or two probands. Many are CLIA certified and some are published, with the remainder in various stages of cell biological validation. To teach the techniques that systematically find these types of variants in single families, we have taken 20 such cases and generated a set of self-study exercises, along with a set of 4 narrated power point presentations, and a total of 150 pages of exercise and reference manuals. These cases include annotated VarSifter (and/or .vcf) files, .bam files, take up approximately 800GB of storage, but to run these files currently requires a Windows based 64bit machine with a minimum of 64GB of RAM. We are currently anonymizing this data, and other ongoing interesting other UDP cases, to allow general distribution. We hope that these cases can be used to instruct future generations of genome wide analysts in the UDP agnostic search strategies gained from our real world examples, and practical experiences.

2304S

Effectiveness of Personalized Genetic Education Modules. *A. Jenks, S. Darabi, J. Eggert.* Clemson University, Clemson, SC., USA.

Effectiveness of Personalized Genetic Education Modules Applied genetic therapies and personalized medicine urge the necessity of genetic education for clinicians. Specifically there is a need for education about how to utilize available genetic technologies for practice and how to translate available genetic information into specialty practice. In 2012 Powell et al. reported that only %15 of clinicians feel knowledgeable enough to utilize genetic medicine in their clinic. If genetic medicine is to converge with personalized medicine, then doctors in specialty practice need to be equipped to better utilize and interpret genetic technologies. The purpose of this study was to determine the effectiveness of specialty department personalized genetic education modules for clinicians in hospital settings using pre and post-test measures. For the project, two researchers provided independent presentations on specialty focused genetics in the clinic of two different departments in a large community hospital in the Southeast region of the United States. Before the presentation, both researchers met with department staff to discuss need, observed the assigned departments and developed the presentation content based on identified needs. Pre and post-tests were designed based on the presentations and given to the staff in each department to determine the effectiveness of the presentation to increase clinicians' understanding of genetics and attitudes toward genetic medicine. Overall the researchers found little increase (%9.61) in genetic knowledge. This research is a baseline study to determine the effectiveness of specialty department based personalized education modules for clinicians. Results from this project indicate that healthcare professionals are already being taught the theory of genetic medicine but that application based modules are necessary for clinicians to be open to the importance of personalized genetic medicine to their specialty practice. Reference Powell, K. P., Cogswell, W. A., Christianson, C. A., Dave, G., Verma, A., Eubanks, S., & Henrich, V. C. (2012). Primary care physicians' awareness, experience and opinions of direct-to-consumer genetic testing. Journal of genetic counseling, 21(1), 113-126.

2305S

"Hey Doc, here is my DNA sequence, what do you think?". *M. Kriek[1,2], H. Sminia[2], C.E. van der Meer[3], I. Arnold[3], E. Aten[1], M. Koopmans[1], A. van Haeringen[1], M.H. Breuning[1], T. Vrijenhoek[4], J.T. den Dunnen[2,1].* 1) Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 3) Workgroup Promotion of expertise General Practitioners Rijnland, Netherlands; 4) Center for Genome Diagnostics, Department of Medical Genetics,-University Medical Center Utrecht.

When my General Practitioner (GP) fell from his chair when I opened my laptop to check whether the drug prescribed matched with my genome, I realized there was a problem. Although DNA-based medicine is at the doorstep, GPs are not trained in this direction and therefore they lack the knowledge to correctly interpret complex clinical data. To bridge this gap, we developed an unique educational project that prepares physicians for their role in the near future; they were trained to translate genome-wide analysis data to the patient's specific clinical situation. The educational module was set up in close collaboration with members of the target group (e.g. GPs) to help us define the most informative and comprehensive way of communication about this subject. The module consisted of: 1. refreshment of genetics through polling 2. case studies 3. genetic gift card 4. practical (bitter tasting and genetic variants) 5. future perspectives. A total 210 GPs participated in the 195-minute genetic training session and afterwards an evaluation form was completed by each participant. There was a difference in appreciation of the educational module between the ' old and the new' generation GPs; the overall rating was 7.9 (scale from 1-10). The majority of the participants was especially interested in Pharmacogenomics (drugs/genes) and pedigree based clinical decision making. And, of great importance, the training lowered the threshold to contact the clinical geneticist. We will elaborate on the content of educational module prior to disclosing the results of the evaluation by the 210 GPs and, finally, we will summarize the lessons learned.

2306S

Teaching Genomic Medicine to Physicians - this is our responsibility as medical geneticists !!! *I. Maya[1], L. Basel-Vanagaite[2,3,4], E. Taub[1], A. Koifman[5,6], DM. Behar[7], R. Tomashov-Matar[1], R. Sukenik-Halevi[8], A. Reches[9], M. Hubshman Weiss[1,3], N. Orenstein[1,3], D. Marom[2,10], M. Shohat[1,2].* 1) Rabin Medical Center, Petah Tikva, Israel; 2) The Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 3) Pediatric Genetics, Schneider Children s Medical Center of Israel, Petah Tikva, Israel; 4) The Raphael Recanati Genetic Institute and Felsenstein Medical Research Center, Rabin Medical Center, Petah Tikva, Israel; 5) Institute of Human Genetics, Soroka University Medical Center, Beersheba, Israel; 6) Ben-Gurion University of the Negev, Beersheba, Israel; 7) Rambam Medical Center, Haifa, Israel; 8) Meir Medical Center, Kfar Saba, Israel; 9) Souraski Medical Center, Tel Aviv, Israel; 10) Pediatrics A, Schneider Children s Medical Center of Israel, Petah Tikva, Israel.

Background: Due to the transition of genetic knowledge from research laboratories into clinical practice the new field of "Genomic Medicine" has emerged. Primary care practitioners are facing this era with inadequate knowledge and skills in medical genetics and many are unaware of the technical, ethical, legal and psychosocial implications of genetic testing. The challenges of translating genomic technologies into health care practice require novel approaches to educate existing and future clinical physicians from all medical fields. Methods: We initiated a "genomic education" program for the purpose of providing physicians from different medical fields and from all phases of their medical carrier an advanced knowledge in genomic medicine . In the last 2 years since we initiated the program we are organizing 4 courses a year, both in Rabin Medical Center and in Tel Aviv University. We emphasized the main take-home messages for physicians, defined as: risk calculation for various genetic diseases, recognition of the mode of inheritance from pedigrees, guidelines for decision-making on which technique to use, interpretation of test results and their clinical implications. Results: To date, 111 physicians have participated in 7 courses of our "genomic education" program, which included lectures, workshops and guided tours in genetic laboratories. The physicians that participated in our program came from 14 different hospitals and 2 HMO in Israel. Two thirds where senior physicians defined as specialists for more than 10 years, and 1/3 were junior physicians (both residents, fellows and young specialists). In the "pre-course" examination the average score was 56%, the median was 55% (range 20%-80%), whereas in the "post-course" examination it was 78%, Median 80% (range 40%-100%). The average improvement in score as a result of the course, both for junior and senior physicians was 20% (range 0%-80%). The physicians who participated in our program reported a very high level of satisfaction from the theoretical and practical knowledge they acquired as well as the concept of a "one-week update course". Conclusion: A one-week "genomic education" program is an effective strategy to update physicians regarding the advances in Genomic Medicine in order to improve their care of patients.

† Trainee Award Finalist

2307S

Medical Student Confidence and Knowledge about Hereditary Retinoblastoma. *J. Govindavari, A. Nath, W. Blazey, D. McMahon, T. Chan, D. Tegay, B. Krishnamachari.* NYIT College of Osteopathic Medicine, Old Westbury, NY.

Background: In the pediatric population, hereditary cancer predispositions are associated with ~5-10% of all cancer diagnoses. Early identification and treatment of hereditary cancer syndromes often significantly improves likelihood of survival. The level of physician knowledge regarding pediatric hereditary cancer syndromes is currently unknown. Assessing pediatric hereditary cancer knowledge amongst physicians and physicians in training, including medical students, is crucial. The goal of this study was to assess perceived confidence in knowledge, and actual knowledge, of pediatric hereditary cancer syndromes in medical students, with a focus on retinoblastoma (RB). Methods: Medical students answered questions regarding confidence in knowledge about hereditary cancer and knowledge-based question derived from national clinical guidelines and recently published literature. 239 students completed the survey. Results: When analyzed by year in school, the four years differed in confidence of their knowledge of hereditary cancer (p=.02), with the highest percentage of students feeling confident in year 2 (42.0%). However, there was no statistically significant difference between class years in terms of answering knowledge questions correctly. Students who felt confident in their knowledge of hereditary cancer were more than 2 times more likely to know that the risk for extra-ocular primary neoplasms is increased in individuals with hereditary RB [OR 2.30 (1.08, 4.91)] and were more than 3 times more likely to know "that genetic screening of the RB1 gene should be considered in any child with bilateral RB" [OR 3.42(1.11, 10.54)], after adjusting for age and years in school. Conclusion: Confidence was associated with knowledge, but knowledge was not associated with increased years of schooling. Students felt most confident in their second year, and less so in their third and fourth years. The increased confidence during year 2 may have been due to students preparing to take their initial board exam. More research is needed to understand the cause of the differences in confidence and knowledge between the four years of medical school classes with extension to assessment of knowledge and confidence in medical residents and attending physicians.

2308S

Creation of a genetics education program for OB/GYN resident physicians. *C.M. Osborne[1], E. Hardisty[1], C. Vladutiu[2], N.L. Vora[1].* 1) The University of North Carolina at Chapel Hill, Department of Obstetrics and Gynecology, Division of Maternal-Fetal Medicine, Chapel Hill, North Carolina; 2) The University of North Carolina at Chapel Hill, Department of Epidemiology, Chapel Hill, North Carolina.

Background: Educators in genetics have previously acknowledged that practicing physicians have limited knowledge about genetics, genetic testing options and the implications for medical practice. This deficit is expected to increase as the breadth of genetics knowledge continues to increase exponentially. Prior studies have suggested that a greater emphasis on genetics education is needed at all levels of medical education, including OB/GYN resident programs. Methods: Three hours of educational workshops were developed for OB/GYN resident physicians at the University of North Carolina at Chapel Hill. The educational workshops were intended to increase the residents' exposure to, and knowledge of, clinical and reproductive genetics. The seminars were presented over a three week period, with a one-hour lecture being administered each week. Curriculum content included common fetal aneuploidies, structural chromosome rearrangements, pedigree analysis, Bayesian analysis, risk analysis, application of Hardy-Weinberg equilibrium, ethnicity-based carrier screening and hereditary cancer syndromes. Lectures were created with input from reproductive genetic counselors, Maternal-Fetal Medicine specialists and clinical geneticists. A 14-question, multiple choice pre- and post-test were administered to all resident physicians to determine if overall understanding and knowledge of clinical genetics increased over the three week course. Answers were collected and analyzed anonymously. Pre- and post-test scores were compared using the Wilcoxon Mann Whitney test. Results: Thirteen pre-tests and 13 post-tests were submitted for analysis. The median (interquartile range, or IQR) scores were 8 (6-8) for the pre-test and 10 (9-12) for the post-test. The difference between underlying distributions of the scores from the pre-test and scores from the post-test groups were found to be statistically significant ($z = -2.6888$, $p = 0.0126$). Conclusions: Creation of an intensive genetics curriculum for OB/GYN resident physicians appeared to enhance the understanding of clinical and reproductive genetics among participants. As the complexity and utilization of genetic testing continues to increase in the prenatal setting, implementation of genetics education for the OB/GYN resident physicians should be considered.

2309S

Preparing future physicians to practice genomic medicine: Lessons from the first two years of the University of Miami Master of Science in Genomic Medicine program. *W.K. Scott[1,2], S.E. Hahn[1,2], S.C. Sacharow[1,2], M. Tekin[1,2], E. Rampersaud[1,2], B. Johnson[1], G. Ghaffari[1], E.M. Bendik[1,2], J.M. Vance[1,2].* 1) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 2) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL.

Needs Assessment: One barrier to the implementation of genomic medicine in the clinic is the concern among physicians that they do not have the knowledge needed to use genomic information in patient care. Lowering the barrier requires training at three levels: medical school, residency, and continuing education. While medical schools should be the leaders in this educational effort, curricula are saturated with topics required for subsequent board exams. **Rationale**: To address this barrier in medical school, we have created a 4-year Master of Science program in Genomic Medicine (MSGM) that is completed concurrently with the medical curriculum. Students apply to the program in the first semester of medical school and instruction begins in the second semester. **Course Content**: MSGM topics were selected by faculty with expertise in genomic medicine, genetic counseling, clinical and human genetics, epidemiology, statistics and laboratory science. Topics include ethical, legal, and social issues in genomic medicine; computational methods; bioinformatics; clinical laboratory methods and technology; pharmacogenetics, and research ethics. The first 3 semesters also contain courses in clinical applications of genomic medicine that examine cases from physiological systems studied in the medical curriculum. **Teaching Strategies**: A *flipped classroom* model is used the first 2 years. Didactic materials (lectures, media, interactive modules) are provided online for self-directed learning. Lessons are reinforced by weekly small-group discussions of core topics and examples from medical literature. The final 2 years include a 4-week clinical genetics clerkship and a practicum consisting of clinical experience and research performed with a faculty mentor. **Outcomes**: In total 17 students (5 in the first class, 12 in the second) have enrolled in the MD/MSGM program. All have continued to be successful in the medical curriculum. All reported satisfaction with the current structure and curriculum and believe the knowledge acquired will enhance their future clinical practices regardless of anticipated specialty. A frequent suggestion was to increase coverage of clinical applications and to reduce research-oriented discussions. This suggests students chose the program to acquire knowledge in genomic medicine, and not as a path to research careers. We believe this is a viable approach to prepare future physicians to incorporate genomic medicine into their daily clinical practice.

2310S

Building Integrated, Individualized, Themed Genetics Electives for Resident Trainees at Johns Hopkins. *T. Wang[1, 2], J. Bodurtha[1], P. Sosnay[3], H. Vernon[1], J. Fahrner[1,2], N. Green[4], C. Thomas[1,2], D. Augustyniak[4], B. Shapiro[4], D. Valle[1,2].* 1) Inst Gen Med, Johns Hopkins Univ, Baltimore, MD; 2) Medical Genetics Residency Program, Johns Hopkins University, Baltimore, MD; 3) Department of Internal Medicine, Johns Hopkins University; 4) Kennedy Krieger Institute, Baltimore, MD.

Needs and Objectives: With the rapid advances in clinical genomics and individualized medical care, genetic contributions to all diseases are increasingly recognized. Genetics residents and other healthcare providers at all levels need enhanced genetics education. The existing one-size-fits-all genetics elective provides only partial exposure to the broad range of genetics conditions. More than 40 clinics, centers, and clinical programs exist at Johns Hopkins to care for patients with individual genetics conditions. The educational potentials of these resources are much under-utilized for residency training due to lack of coordination and organization. We aim to organize campus-wide clinical genetics resources at Johns Hopkins to build integrated, individualized, and themed electives in medical genetics for resident trainees in genetics and other specialties. **Setting**: Specialized clinics, centers, and programs caring for genetics patients at Johns Hopkins and Kennedy Krieger Institute (KKI). **Participants**: Second year residents in medical genetics programs at Hopkins (n=5). **Descriptions**: We built four themed electives including Adult Genetics, Biochemical Genetics, Clinical Genomics, and Syndromes & Malformations. Each resident will (1) meet with elective directors to discuss their background, interests, and training goals, (2) select 4-6 clinics of their interest among participating clinics at Johns Hopkins and KKI to build his/her own individualized, themed elective plan, (2) participate in all clinical educational activities during the elective (2-4 weeks), (3) study online teaching materials related to the elective theme, (4) directly interact with elective directors and attending physicians (5) contribute to 1-2 case summaries or presentations, and (6) participate in online evaluations before and after the elective to assess progress toward objectives. **Evaluation**: direct feedback, structured survey, and online evaluations by attending **Lessons Learned**: All participating residents are very enthusiastic about these electives. They particularly like that these electives are tailored to individualized interests and learning objectives, and are more efficiently support them to meet their educational goals. 90-95 percent of the residents consider these electives from “very good” to “excellent” and will recommend to other trainees in genetics. Online teaching resources and evaluation tools are being developed.

2311S

Personal Pharmacogenetic Testing in the Medical School Curriculum: Student Knowledge and Attitudes. *S.C. Sanderson[1], S.A. Scott[1], N. Schrager[1], M. Zweig[1], L. Shi[1], A. Kasarskis[1,2], G.A. Diaz[1].* 1) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 2) Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029.

As increasing numbers of personalized medicine programs have begun to pre-emptively genotype pharmacogenetic variants, the utilization of clinical pharmacogenetic testing is escalating. The recent publication of clinical practice guidelines has facilitated this process, but concern over the lack of expertise among clinicians to interpret pharmacogenetic test results is often cited as a barrier to uptake. To assess the effects that direct experience with pharmacogenetic testing have on knowledge and attitudes about pharmacogenetics, medical students at the Icahn School of Medicine at Mount Sinai were offered the option of being anonymously genotyped for actionable variants in CYP2C9, CYP2C19, CYP2D6, and VKORC1 through the CLIA-certified Mount Sinai Genetic Testing Laboratory as part of their medical genetics and genomics training. After an introductory presentation, all students were given the option of receiving personal or example genotype reports, and simultaneously offered the opportunity to participate in a research protocol assessing knowledge and attitudes on pharmacogenetics using a structured survey before and after the educational module. In the first year of the program, 93 out of 140 (66.4%) second year medical students opted for personal pharmacogenetic testing. Among those who answered the pre-course questionnaire (86.0%), decisional conflict scores were low, indicating that this decision was made easily by most students. Almost half (45.7%) believed the information would not be medically useful or change medical decisions, and 42.9% felt that pharmacogenetic testing costs too much. The strongest reasons in favor of testing were general curiosity (91.0%), to understand what a patient may learn or experience (91.4%), and to help get a better understanding of human genetics (91.4%). Importantly, knowledge about pharmacogenetics was significantly higher after compared to before the course (p=0.03), and the proportion of students who stated that physicians have a professional responsibility to help individuals understand pharmacogenetic testing results, even if the physician has not ordered the test, significantly increased from 8.6% before to 14.3% after the course (p= 0.021). These results suggest that given the option to undergo pharmacogenetic testing and receive their personal results as part of their training, the majority of medical students opt to do so, and that this may have a positive impact on their educational experience.

2312S

Dissecting the genetic susceptibility to sporadic molar pregnancies and mechanisms of their formation. *Y. Khawajkie[1], NMP. Nguyen[1], W. Buckett[2], P. Sautier[3], M. Breguet[3], R. Slim[1].* 1) McGill, Montreal, Canada; 2) Obstetrics and Gynecology, McGill University Health Center, Montreal, Quebec, Canada; 3) Service de Gynécologie Oncologique, Hôpital Notre-Dame, Université de Montréal, Montréal, Québec, Canada.

A hydatidiform mole is an abnormal human pregnancy characterized by absence of, or abnormal, embryonic development, excessive trophoblastic proliferation, and hydropic degeneration of placental villi. The common types of moles are sporadic and affect 1 in 600 pregnancies in western countries. Among patients with one mole, up to 20% have a second reproductive loss, which is higher than the incidence of recurrent SAs in the general population (2-5%). This indicates that these patients are genetically susceptible to reproductive loss. Based on the parental contribution to sporadic molar tissues, there are three main genotypic types of moles: diploid androgenetic monospermic, diploid androgenetic dispermic, and triploid dispermic. However, it is not known whether these three genotypic entities have the same genetic predisposition to reproductive loss. In addition, it is not known whether the other forms of reproductive loss in women with one mole are caused by the same or by different mechanisms. To address these questions, we compared the reproductive histories of patients with the three genotypic types of moles with each other and with those of patients with recurrent spontaneous abortions and no moles. We then used flow cytometry, fluorescent microsatellite genotyping, and fluorescent in situ hybridization on a total of 69 pregnancy losses from 50 patients to determine the parental contribution to all their available products of conception. At the meeting, we will present our comprehensive analysis and discuss the genetic susceptibility and mechanisms of pregnancy losses in patients from the three categories of sporadic moles and patients with recurrent spontaneous abortions and no moles.

2313S

Kyoto Model of developing a genetics education program in Japan. *N. Akiyama[1], M. Torishima[1], T. Katayama[2], T. Wada[1], S. Kosugi[1].* 1) Kyoto University Graduate School of Medicine, Kyoto, Kyoto, Japan; 2) Osaka Prefectural Hirakata High School.

[Introduction] Genetic technology has been advancing rapidly and genetic tests are becoming increasingly common in Japan. The curriculum of primary, junior high, and high schools in Japan includes little information regarding human genetics. In order to improve genomic literacy in Japan for this dawning genomic era, we have started developing a genetics educational program for ordinary people, schoolchildren in particular. [Methods and Results] (1) Exhibition of human genetics at several events at universities or in society: (a) A two-hour exhibition with more than 60 groups from all faculties in a one-day event at Kyoto University. We exhibited a DNA model (zometool®), illustrated books about genetics, a quiz game in a PowerPoint file, and extraction of DNA from a chicken liver. We discussed genomic topics with participants who ranged from young children to adults. (b) At a two-day Kids Festival event at Kyoto railway station. More than 3000 kids and their families participated in the event, and more than 500 come to our section on human genetics among eight scientific exhibitions from Kyoto University. Families observed human chromosomes with a microscope and studied DNA and chromosomes with genetic counselors. (2) Lectures in a high school biology class: We conducted three 50-minute lectures on human genetics. We discussed several genetic subjects, including genetic tests for HBOC and prenatal diagnosis for Down syndrome by NIPT. The lecturer is a student earning their master’s degree in genetic counselling. Students studied the basic of human genetics and how to draw a family chart. (3) Seminars on human genetics for primary school pupils: We plan to have several seminars for 20 schoolchildren each time during weekends or on summer vacation. Each four-hour seminar will be composed of lectures on the basics of human genetics, an exercise on DNA extraction, drawing of the family tree, and discussion of genetic topics. [Conclusions] Our project is unique in that master’s students or certified genetic counselors are leading the project. The purpose of our project is to familiarize children and parents with human genetics, to enable them to appreciate genetic information so they can better manage their health. It is important that we consider how and what we should teach about human genetics to children and students. We have just started the program, and it will take time to evaluate our project.

2314S

Attitudes and Response to Genetic Risk Information Among Adolescents. *C. Bloss[1], B. Darst[1], B. Daley[2].* 1) Scripps Genomic Medicine, Scripps Translational Science Institute, La Jolla, CA; 2) High Tech High, San Diego, CA.

Genome sequencing with pediatric populations is currently being performed, however there is little data on the impact of genetic risk information on younger individuals. Furthermore, most work in this area is focused on specific clinical populations, and the voices of minors themselves are often overlooked with most studies focused on parents. The current, school-based study assessed attitudes and response to genetic risk information in group of tenth graders before and after an educational seminar on genomics and personalized medicine. Students recruited from a local group of San Diego charter schools attended an educational seminar on genetics and genomics and then received information about a test of genetic risk for complex disease and drug response traits within the context of an analogue study design. Specifically, each participant was randomly assigned to one of four conditions, which varied by the number of diseases or drug response traits that showed a high risk result. Students completed demographic, attitudinal, and behavioral assessments before and after the seminar and exposure to the analogue vignette. 140 students across three biology classrooms were offered participation. Of those, 80 students enrolled and completed both the baseline and follow-up assessments. All participants were in the 10th grade and ranged from 15 to 17 years of age. There were no significant differences in emotional response between students in the four genetic risk conditions. Analyses of follow-up data also showed significantly lower anxiety towards genomic testing than the already low levels reported at baseline before the seminar. Further, after attending the genomics seminar, students' knowledge of genomics increased significantly and a large majority of students said that they would consider undergoing genetic testing if it seemed warranted. Our findings suggest a lack of adverse emotional responses among adolescents exposed to genetic risk information. While analogue methods cannot perfectly replicate reality, previous studies suggest that responses to risk information observed in such studies are similar to those seen in actual clinical situations. Although studies with larger samples are needed, these results also highlight the utility of education programs for minors to increase genomic literacy. Such efforts are critical given that the youngest members of our society are likely to be increasingly exposed to genomic medicine as they develop into adulthood.

2315S

Teaching the relationship between genotype and phenotype with public data and Molecule World™ on the iPad. *TM. Smith, SG. Porter.* Digital World Biology LLC, Seattle, WA.

It is impossible to grasp fundamental concepts in biology without understanding the relationship between sequence, structure, and function. Modern data collection technologies are creating enormous data resources that can be used to help students' understand these relationships, but currently are underutilized. This is due to the fact that easy-to-use tools that meet teachers' needs for clear instruction are not yet commonplace. Filling the gap between the embarrassment of data riches and practical classroom use requires three things: user-friendly tools, content that demonstrates specific applications with interesting stories, and packages that combine instruction, assessments, and inquiry-based investigations.

Digital World Biology is addressing this need with its on-line courses and mobile apps. The on-line courses increase students' computer literacy while using standard tools like Cn3D, Blast, ORF finder, and multiple databases, in directed and exploratory ways, and help students better understand biology as well gain a better appreciation for the value of the data and the field of bioinformatics. In response to nearly two hundred interviews with K-12 and college teachers and students, we created Molecule World and the Molecule World DNA Binding Lab™ iPad apps to display 3D-data from multiple structure databases (MMDB, PDB, and PubChem) using a novel rendering engine that allows us to uniquely highlight chemical properties and sequence orientation. The ability to display and highlight sequences within molecular complexes enables exploration into the relationships between sequence, structure, and function in new ways. Preliminary data collected in professional development workshops, and many demonstrations, supports the hypothesis the being able to view and simultaneously interact with data improves teaching capabilities and student engagement. This work has been supported by NSF grant IIP 1315426.

2316S

Teaching the genome generation, a laboratory intensive high school genomics and ethics course. *C. Wray[1], M. McKernan[1], D. Waring[2].* 1) Genomic Education, The Jackson Laboratory, Bar Harbor, ME; 2) Personal Genetics Education Project, Harvard Medical School, Boston MA.

To increase genetic and genomic literacy and train the next generation of scientists, we have developed, tested and assessed an intensive Genomic Biology course, entitled Infinite Variations, Personalized Medicine and Genomics. The course is multi-faceted and includes bench molecular biology experiments, bioinformatics exercises and focused bio-ethics discussions. The goals of the course are four-fold. First the course introduces the complexity of human genetic and genomic variation and stresses the importance of individual variants. Rather than discussing mutant and wild-type alleles, the course highlights the fact that every genome is rife with single nucleotide polymorphisms (SNP), insertion-deletions (indels), and structural or copy number variants (CNV). Second, at the hands-on skill level, the course instructs students in the use of molecular assays used to detect genetic variation. Four experimental modules and a panel of de-identified human genomic DNA samples are used to demonstrate and assay for SNPs, indels, and CNVs; simple PCR-restriction digest assays are taught and compared to higher throughput molecular beacon PCR techniques and traditional sequencing. The third goal of the course is to introduce students to Bioinformatics and the use of public databases to learn about genetic diseases and genetic variation in humans. The fourth, final and vital goal of the course is challenge students with relevant ethical dilemmas that anchor genomic biology within real-world situations. Students are engaged in discussions covering genetic privacy, genetic pre-determinism, prenatal genetic testing, and biomedical research ethics. Over the most recent three year period a pre-course and post-course assessment strategy has been used to measure learning outcomes. In all three years the student participants have shown significant gains in genomic and genetic literacy and knowledge of bioethics; on average, % correct answers on the post-course assessment increase by 24%. The course is currently being re-designed as a teacher professional development exercise in order to increase dissemination across New England high schools.

2317S

Learning a language through Genome Education: Spanish School Experience. *S. C. Zapico[1], S. Martinez de Castro[2].* 1) Anthropology, Smithsonian, NMNH, MRC 112, Washington, DC; 2) Consejo de Residentes Españoles en Washington, CRE, Washington, DC.

Nowadays, genetic research, and the application of these studies to everyday life, is more attractive to the public, especially as it can be used to find cures for diseases. Also as scientists, it is important to increase the interest for science in new generations to ensure future advancements of the field. The Spanish School of Washington DC promotes the study and practice of the Spanish language for children born in the United States, but with parents of Spanish origin. This education takes place through extracurricular classes and cultural activities, conducted in Spanish, outside of their English-speaking schools. One such activity was a visit to "Genome: unlocking life's code," an exhibit at the National Museum of Natural History. Children were divided into three groups of 12, each group ranging in ages from 7 to 15 years old. The explanation in Spanish started with the basic concepts of DNA, which included how DNA is processed to form proteins, the history of DNA, and the sequencing of the whole genome. Once the students understood these concepts, the next step was to teach the significance of recent advances in sequencing in keeping the balance of the ecosystem. After that, the rest of the lesson was focused on human genetics; from how migrations are marked in our mitochondrial DNA, to how the advances in genetics let us diagnose diseases and look for new drugs and treatments. Although some of the concepts were difficult to explain to some of the younger students, the approach was to take examples from real life, connecting our genotype with our phenotype. To illustrate the concept of DNA, at the end of the lesson the students were given the opportunity to work as scientists. We performed an experiment that consisted in extracting DNA from strawberries and bananas through a protocol that used common household items. Both parents and students were involved in this activity, and at the end many of the students were successful in the extraction of DNA. The students and parents reacted positively to the lesson and especially the experiment: they kept the protocol and many expressed their desire to try it again at home. In summary this activity achieved our two primary objectives: to promote science in new generations, and make young students familiar with science concepts in Spanish, increasing their knowledge of this language.

2318S

Accelerating public awareness of personal genetics. *D. Waring[1], M. Gelbart[2].* 1) Depart of Genetics, Harvard Medical School, Personal Genetics Education Project Boston, MA; 2) Marnie Gelbart Harvard Medical School Department of Genetics Personal Genetics Education Project Boston, MA.

The Personal Genetics Education Project (pgEd.org) is using a unique and expanding platform of strategies to address the growing gap between what researchers are learning at the frontiers of genetics and what the public understands. Here, we focus on our most recent accomplishments in the domains of policy, entertainment, gaming, and education. First, in 2014, pgEd conducted a Congressional briefing to draw policymakers' attention to scientific advances in personal genetics and the importance of raising public awareness. Second, pgEd is expanding its emphasis on advancing awareness through television and film through partnerships with Hollywood, Health & Society and the Science & Entertainment Exchange. Third, pgEd is continuing to develop the on-line game, Map-Ed (map-ed.org) for spreading awareness. Our 2014 expansion of Map-Ed's topics include the microbiome and genetic discrimination. To date, Map-Ed has 5,400 players on all 7 continents. Fourth, pgEd is building our up-to-date and freely available on-line curriculum that addresses the benefits and implications of personal genetics. Finally, pgEd's efforts training teachers to engage young people on the interdisciplinary topic of personal genetics are extending beyond the biology classroom and into health, social studies, and literature. pgEd is continuing to set new goals, including engaging faith-based communities, and is seeking partners and collaborators to expand our vision and reach.

2319S

Celebrity disclosures and information seeking: The case of Angelina Jolie. *R. Juthe[1], A. Zaharchuk[2], C. Wang[3].* 1) Office of Communications and Education, National Cancer Institute, Rockville, MD; 2) iDoxSolutions, Inc., Bethesda, MD; 3) Department of Community Health Sciences, Boston University School of Public Health, Boston, MA.

Background: On May 14, 2013, actress Angelina Jolie disclosed that she had a deleterious *BRCA1* mutation and underwent a preventive bilateral mastectomy to reduce her inherited cancer risk. Her disclosure generated media frenzy and much discussion about her potential "effect." We sought to document the impact of her disclosure on information-seeking behavior, specifically regarding online cancer genetics and risk reduction resources available from the National Cancer Institute (NCI). **Methods:** Using Adobe Analytics, daily page views for 11 online resources, including resources written for the public and health professionals, were tracked over a 9-week period from April 23, 2013 through June 25, 2013. Online usage data were also obtained for 4 resources over a 2-year period (2012-2013). Source of referral by which viewers located a specific resource was also examined. **Results:** Jolie's disclosure led to a dramatic and immediate increase in traffic to NCI's online resources. The *Preventive Mastectomy* fact sheet written for the public saw the largest increase with 69,225 page views on May 14, representing a 795-fold increase compared with the previous Tuesday. The Cancer Genetics Services Directory for the public saw a 31-fold increase in page views from May 7 to May 14. A fivefold increase in page views was observed for the health professional Physician Data Query (PDQ)® *Genetics of Breast and Ovarian Cancer* summary in the same timeframe. Health professional PDQ® summaries on the genetics of other inherited cancers, including skin cancer and colorectal cancer, received increases of up to 65% from May 7 to May 14. Two-year data demonstrated the magnitude and longevity of Jolie's effect. A substantial increase from 0% to 49% was seen in referrals from news outlets to NCI's breast cancer resources from May 7 to May 14. **Conclusions:** Celebrity disclosures, such as Jolie's disclosure of her *BRCA1* mutation status and her risk management decision, can dramatically influence online information-seeking behaviors. Efforts to capitalize on these disclosures to ensure easy access to accurate information are warranted.

2320S

Public Perspectives on Genetic Ancestry Testing. *C.D. Royal[1,2], C.M. Wolpert[1,2], A. Hoffmeyer[3], J.D. Powell[1,2], K. Haynie[1,4], R. Agans[3].* 1) Department of African & African American Studies, Duke University, Durham, NC; 2) Institute for Genome Sciences & Policy, Duke University, Durham, NC; 3) Carolina Survey Research Laboratory, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Department of Political Science, Duke University, Durham, NC.

Little is known about the general public's familiarity with and views of genetic ancestry testing (GAT). Preliminary data in the United States (n= 761) suggest that 68% (±4%) of the adults surveyed had heard of using genetic testing to find out about one's ancestry, and furthermore, 62% (±6%) said they would be interested in knowing their genetic ancestry. When examining minority and non-minority status differences, both non-Whites (68%; ±7%) and Whites (67%; ±5%) were equally familiar with GAT, but non-Whites were more interested in knowing their genetic ancestry (66% vs 60%, respectively) (r = .136, p=.059). In addition, non-Whites were more likely to consider having GAT (41.2%; ± 2%) in comparison to Whites (23%; ±2%). Overall, 73% (±1%) of the respondents believe that their genetic ancestry is related to their health and 86% (±1%) would tell their doctor about their GAT results. At present, there are no White/non-White differences here. The general public seems to be familiar with GAT and interested in knowing their genetic ancestry; however, when minority and non-minority status are considered, respondents who self-identified as non-White reported greater interest in both knowing their genetic ancestry and having GAT. Further research is ongoing and is expected to inform efforts to better engage the general public in order to explore their views on GAT and their perceptions of the relationship between genetic ancestry and health.

2321S

REU Site: Research Experiences for Underrepresented Minority Undergraduates in Basic Science and Genetic Research at Louisiana State University Health Sciences Center (LSUHSC). *K. Foley, A. Umrigar, A. Musto, H. Farris, F. Tsien.* LSUHSC, New Orleans , LA.

The Louisiana State University Health Sciences Center (LSUHSC) in New Orleans provides research training for a period of ten weeks for up to ten Louisiana underrepresented minority (URM) undergraduates interested in exploring graduate school and biological research careers, particularly in genetics. In the New Orleans area, Blacks account for 60.1% of the population, Hispanics 5.0%, and Native Americans 0.3%, compared to12.6%, 16.3%, and 0.9%, respectively in the US. The field of genetics is constantly evolving and it is therefore imperative for students who are pursuing biological science careers to be proficient in hypothesis-driven scientific problem solving, current methodology, data analysis, and communication skills. The objectives of the proposed REU program in basic biology research include: 1) Scientific problem solving 2) Learning state-of-the-art laboratory techniques 3) Developing scientific communication skills including teaching at a genetics workshop to high school students 4) Acquiring resume and CV writing and career education resources 5) Participate in a number of seminars and interactive workshops. Students present posters at an end of the summer symposium which are judged by local scientists. Additional seminars regarding graduate school professional development are presented by currently enrolled genetics graduate students and invited faculty. Students complete anonymous pre- and post-internship surveys designed to measure their self-assessed research interest, demographics, and achievement. The students will be followed up for a period of at least 3 years following the completion of the program to gauge the degree to which their experiences have had a lasting influence on their respective career paths. Students will also use an on-line REU assessment tool. We will stay in close contact with the students via a yearly electronic survey. Following the REU program, interns also have the opportunity to teach genetic concepts to New Orleans area schools during the academic year, an important skill that will enhance their future careers as scientists.

2322S

Gene hunting with IMG-ACT: Integrating genomics research into the undergraduate biology curriculum. *K. Moitra.* Dept of Biology, Trinity Washington University, Washington DC, DC.

Genome annotation is the process of linking biological information to gene sequences. It involves annotating predicted genes and connecting these genes to protein functions. The integration of genomics research into the undergraduate biology curriculum provides students with the opportunity to become familiar with state-of-the-art bioinformatics tools and answer original research questions. Our purpose with this research project was to upscale the research experience so that large numbers of students could have access to research. One way to achieve this is through integration of research with the classroom experience. Students in an introductory genetics class annotated predicted ABC (ATP Binding Cassette) genes of Methanothermobacter thermautotrophicus using the Integrated Microbial Genomics-Annotation Collaboration Toolkit (IMG-ACT) developed by the Joint Genome Institute. This research project is linked to the much larger scientific question: What can the sequence of an organism's genome tell us about its overall biology? A variety of pre-course and post-course tests and surveys were conducted to assess if the project engaged students and generated interest in scientific research. Preliminary results suggest that it did. This session will discuss: initiation and overview of the project, integration of research into the biology curriculum, up-scaling research projects for involvement of larger numbers of students, student learning outcomes and data collected from student surveys/assessments, including how future career choices were impacted by this research project. The session will also highlight building and scaffolding research skills in a research-intensive genetics course and the integration of research into the curriculum.

2323S

Development of a molecular test and human pedigree analysis on Alpha-1 Antitrypsin Deficiency. *A. Olson, D. Caporale.* Biology, University of Wisconsin-Stevens Point, Stevens Point, WI.

Alpha-1 Antitrypsin Deficiency (Alpha-1) is a monogenic recessive disorder affecting roughly 1 in 4,000, making it 1 of 3 of the most common genetic diseases to cause death among adult Caucasians in the United States. The deficiency is caused by a mutation of the SERPINA1 gene, located on chromosome 14, which codes for the serpin peptidase inhibitor alpha-1 antitrypsin. Inheriting two mutated variants can result in Emphysema, COPD, and/or liver disease between ages 40-60. In one case study, an individual was recently diagnosed with this disorder and was found to be homozygous for the affecting variant "Z", rendering each parent a carrier. This raised several concerns for other family members, who then wished to seek genetic testing. Current genetic tests include restriction digestion of amplicons, which can be less reliable or repeatable due to partial digestion issues. The objective of my undergraduate research was to design a novel genetic test for the SERPINA1 "Z" single-nucleotide polymorphism (SNP) based on DNA sequence comparisons - that is reliable, repeatable, noninvasive and relatively easy to perform - to facilitate the detection and diagnosis of Alpha-1 patients. Since the "Z" variant is a result of a transition missense mutation, whereas a guanine (found in the wild type form denoted as "M") is replaced by an adenine, our methods included designing a forward and reverse primer pair that would successfully amplify this SNP region and generate DNA sequences representing an individual's genotype that is easily interpreted. Eighteen members of the Alpha-1 affected family were used as controls to verify the reliability of this molecular test. Amplified products were cycle-sequenced by the fluorescent Sanger method and electrophoresed using an automated capillary system. After protocol optimization, homozygous wild types (MM), heterozygous carriers (MZ), and homozygous mutants (ZZ) became easily discernable among the respective electropherograms. Heterozygote electropherograms displayed equal representations (concentrations) of adenine and guanine at the polymorphic site. In addition, a genogram was constructed that successfully illustrated the congruence of their genotypes with their respective phenotypes of this disorder. We conclude that our molecular method for identifying M/Z genotypes associated with Alpha-1 Antitrypsin Deficiency is a reliable test that can be easily adapted for research or diagnostic purposes.

2324S

Using a PyMOL activity to reinforce the connection between genotype and phenotype in an undergraduate genetics laboratory. *A. Ribes-Zamora, T.K.T. Nguyen, A.D. Simmons.* Biology Department, University of St Thomas, Houston, TX.

With the purpose of developing activities that would help clarify and reinforce genetic concepts that undergraduates find challenging, we designed a 3-hour teaching module using the PyMOL software, which can be downloaded for free for educational purposes, plus other genomic browsers. We believe this teaching module reinforces the concept of mutation and the connection between genotype and phenotype. Pre-lab handouts with background information on a specific protein or protein family are provided to the students the week before the day of the activity. Students are also given a homework assignment that allows them to begin exploring some of the tools of the PyMOL software. On the day of the lab session, teams of four students are given different handouts that prompt students to find the location of the gene in the human genome, see how many introns and exons it has, and investigate the structure and function of a specific protein. They are also given a few disease-associated amino acid changes that lead to a particular disease and asked to predict how the change can ultimately alter overall protein shape or function. Teams are given two hours to complete the activity and then give a 10-minute presentation on their findings and show their models. After the activity we were interested in measuring if there was a better understanding of mutation among students that were enrolled in the laboratory and the lecture concurrently, as compared to students who were only enrolled in lecture. Students were asked to complete an anonymous post-quiz. In all three lab sections, the average post-quiz grade was slightly higher than the average post-quiz grade for students that did not take the lab. Also, there were 6 questions in which there was a higher percentage of students that took lab and lecture concurrently that obtained the correct answer as compared to students that only were enrolled in the lecture. While this data is encouraging, the differences we observe are not statistically significant. We believe this is the case because we have small class sizes and we would probably have to pool data from various sections to achieve significance.

2325S

Orchid DNA-Barcoding: a guided research project for undergraduate genetics students. *A.D. Simmons[1], C. Ribadeneyra[1], K. Foss[1], E.J. Baenziger[2], A. Ribes-Zamora[1].* 1) Biology Department, University of St Thomas, Houston, TX; 2) Department of Modern and Classical Languages, University of St Thomas, Houston, TX.

Recent studies have found that undergraduate STEM courses in which students are taught using problem-based approaches exhibit greater learning gains than do students in traditional courses. Studies done by the National Academies and The American Association for Advancement of Science indicate that undergraduate education can be enhanced significantly when students are involved in authentic research. DNA barcoding is a method by which organisms of various species can be characterized by obtaining a specific short genomic sequence. In Fall 2013, our genetics laboratory course was modified so that students were taught lab skills in the context of conducting a research project using the DNA barcoding technique. Our hypothesis was that structuring the lab course around an actual research project would be a more effective way to teach the scientific method and increase the student's future interest in the scientific field, while also producing potentially publishable data. We had three sections of lab, each with 9-15 students. Lab teams were composed of 3-4 students, and each team collected leaf samples from 3 different species for a total of 33 different species. Throughout 6 of the 11 lab sessions students worked on the project, beginning with an introduction to DNA barcoding and its applications, then moving on to basic lab techniques, performing the data analysis using DNA Subway and ending with presentations of the phylogenetic trees students produced. At the beginning of the semester, students were given a survey that asked them how confident they were in various aspects of experimental design and presentation after taking the prerequisite course, which employed a standard lab approach. At the end of the semester, the students were asked to fill out the same survey, this time ranking how they felt after completing our research-oriented lab. All surveys were anonymous. After taking this lab, over 95% of students felt confident or very confident at applying the scientific method, compared to only 27% before the lab. While approximately 60% of students were confident or very confident in their abilities to organize and present scientific data before the lab, an additional 35% of students shared this confidence after completing the lab.

2326S

Sex and Gender Differences in Health: Educational and Collaborative Outreach to Genetics Researchers, Clinicians, and Students. *M.R. Tennant[1], M.E. Edwards[2], H.F. Norton[2], N. Schaefer[2].* 1) Health Science Center Libraries/Genetics Institute, University of Florida, Gainesville, FL; 2) Health Science Center Libraries, University of Florida, Gainesville, FL.

Purpose: Historically, medical and scientific thinking has been based largely on male anatomy, disease presentation, and response to therapy. This focus presents ethical and scientific challenges. We embarked on an outreach project to promote equitable research sensitive to sex and gender differences in the basic and clinical sciences, including genetic research. Given the NIH policy to reduce gender bias in research, this project provides timely information to clinicians, researchers and students at our institution. Through National Library of Medicine/NIH Office of Research in Women's Health funding, a team of librarians partnered with researchers and clinicians to promote awareness of the need for research in sex and gender differences in health, facilitate collaboration among such practitioners and students, and make research in this area more accessible. Setting/Methods: A large land grant institution with contiguous main and health center campuses. "Collaborating with Strangers" (CoLAB) workshops were held to facilitate collaboration among participants interested in this discipline. Affiliations represented the gamut of health, life, and social sciences, as well as clinical and research realms. Instructional sessions related to the science of sex and gender differences in health were performed for a number of cohorts, including graduate and undergraduate students in genetics and genomics. In order to make sex and gender differences research results more accessible within and external to the institution, an open access publication fee fund was created. Researcher records in the campus finding tool (VIVO) were augmented with keywords, and publications arising from the university's sex and gender differences researchers were included in the Institutional Repository (IR). In the second year of funding the team collaborated in the university's outreach to high school students and teachers, catching students early in the pipeline when they contemplate careers in science and medicine. Results/Conclusions: In its first year the team hosted two CoLabs with 37 attendees; a third session has been held in Year 2. Evaluations suggest that participants connected with other researchers and gained potential collaborators. Instructional sessions for genetics students yielded cohorts ready to enter medical and graduate school with a greater understanding of sex and gender differences research, and how those differences relate to genetics, health and disease.

2327S

Studying Ourselves: Using SNP Genotyping to Engage Undergraduates in Genetics and Research. *E.E. Murray[1], M.D. Hineman[2], A. Hinshaw[2], M. Harrison[3].* 1) Integrated Sci & Technol, Marshall Univ, Huntington, WV; 2) Chemistry, Marshall University, Huntington, WV; 3) Biological Sciences, Marshall University, WV.

The human genome project has made it possible for students to interrogate their own DNA with PCR and to answer questions about human diversity for many non-health related biological characteristics including taste, athletic ability, stress management and hair and eye color. Over the last four years, we developed SNP genotyping assays for polymorphisms as research class projects in an undergraduate human genetics class. Successful projects were a springboard for senior capstone projects using larger biology or forensic science classes as sample populations while providing those classes with a human genetics laboratory. This permitted student researchers to take on the role of co-PI for the SNPs they chose to investigate. Students isolated their own DNA using Epicentre MasterAmp or BuccalAmp kits. PCR was performed using Promega 2X mastermix. Lonza Flash gels were used to rapidly test PCR amplification and in some cases, RFLP polymorphisms. Lonza PAGEr Gold precast gels were used to resolve some RFLP polymorphisms. In some human genetics class SNP projects, published primers and/or PCR cycles were available (PTC tasting, ACTN3 and athletic ability, COMT and test anxiety, HVR 1 and 2 and ethnic origin) while in other projects, students used NCBI database and dbSNP to locate the relevant SNP(s) and design primers (Red hair, cilantro tasting, blue eyes) and either RFLP assays, qPCR genotyping or direct DNA sequencing. Detailed protocols for all assays will be provided. Results: New assays were developed in four annual human genetics classes. Three students from the human genetics classes continued this research in larger student populations as senior capstone projects. Capstone students were required to write an IRB submission and obtain informed consent from research participants. DNA samples were tracked with a number rather than a student name to maintain confidentiality, and samples were discarded after the conclusion of the project. Students in the Cell Biology Class that served as capstone subjects for Cilantro tasting project were surveyed using a Likert scale (with 1 being strongly disagreed and 5 being strongly agreed). Students ranked learning about PCR and SNPs highest as 4.47 and learning about bioinformatics lowest at 3.94. 81% of students surveyed stated that more experiments like this would be valuable in Cell Biology laboratory. Capstone students were uniformly satisfied with the projects.

2328S

Statewide Education Regarding Personalized Genomic Medicine, the Iowa Experience. *C.A. Campbell[1], C. Nishimura[1], M.A. Mansilla[1], T. Bair[1], S. Maddhi[1], A.E. Kwitek[1], K.L. Knudtson[1], O. Shchelochkov[1], S. Gunstream[2], R.J. Smith[1].* 1) Iowa Institute of Human Genetics, University of Iowa, Iowa City, IA; 2) Integrated DNA Technologies , Iowa City, IA.

As Personalized Genomic Medicine (PGM) is incorporated into medical practice, we are challenged with educating the health care team and public about genomic medicine, as well as developing a work force to support the growing genomic needs of the population. The Iowa Institute of Human Genetics (IIHG) provides an interface for, and support of, state-wide activities related to human genetics with a mission to educate Iowans and others about human genetics. To accomplish this mission we've introduced multiple educational activities. Programs for *students* include; Careers in Human Genetics Day, and Personalized Genomic Medicine: Careers in Bioinformatics & Big Data. Following these events, 97% and 83%, of attendees report an increased interest in careers in PGM, respectively. Attendees report; "I have many more options than I thought"; "It helped me realize what kind of environment I'd like to work in"; "I didn't realize the magnitude of teamwork"; "I wish I could've attended earlier because it could've steered me in the right direction sooner". A genetic counseling internship provides students with clinical experience and is the only undergraduate genetic counseling internship in the country. The IIHG also offers a "Topics in Human Genetics" summer course, and outreach to Iowa colleges and high schools. Opportunities for *researchers and clinicians* include; Next Generation Sequencing and Bioinformatics User Groups Monthly Conferences; Bioinformatics Short Course; and Exome Analysis for Rare Disease program. For the *general public*, we have hosted seminars for the general public on genetics in art, cinema, medicine, and ethics. Our plain language patient brochures and website provide educational materials for patients, students, researchers and health care providers.

Formal collection of feedback on all activities is used to gauge the interests and needs of our target populations for future program development. The average satisfaction score is 4.59 out of 5 for all of the offerings. Following our events, students have applied for bench and computational human genetics laboratories jobs, and graduate school. This demonstrates our education programs are effective in motivating students to pursue careers in PGM. We are committed to educating and engaging the entire community by creating a public awareness of personalized genomic medicine as an investment in our future.

2329S

Molecular Genetic Studies in Indian Patients with Ectodermal Dysplasia. *s. kushmakar.* pediatrics, All India Institute of Medical Science, New Delhi, India.

Introduction: Ectodermal Dysplasia (ED) is a group of about 170 heritable disorders that affect the ectoderm, the outer layer of tissue. Each of the roughly 170 ED syndromes represents a different combination of abnormalities. Inheritance pattern are variable include autosomal dominant, autosomal recessive, X linked dominant and X linked recessive disorders. Here we report our experience with X-linked hypohidrotic or anhidrotic ectodermal dysplasia. The EDA gene which causes this disease is present on Xq12 - q13 and comprises of 12 exons. Eight of which encode the transmembrane protein ectodysplasin-A and 4 exon are non coding. More than 179 mutations have been reported from different groups, including missense, splicing, small deletions and insertions. Objectives: Our objectives were to identify the spectrum of mutations in Ectodermal Dysplasia patients and to plan a strategy for molecular diagnosis so that optimal management, genetic counselling, and prenatal diagnosis can be offered. Material and methods: Thirty (25male and 5 Female) patients with clinical features suggestive of ED were included in the study. In two families mothers also had mild features suggestive of ED. Four of the families were consanguineous.DNA was extracted from 30 patients and their parents. PCR amplification was carried out using 8 pairs of primers and sequencing was done . All the patients who were found negative for EDA gene sequence variation were subjected to screening of EDAR gene. Results: 14 patients out of 30 were reported with sequence variant and out of 14 ,8 were reported and 6 patients were found to be a reported mutations, c.2 T>A (exon 1) ,c.463 C>T (exon 3) , c.466 C>T (exon 3) ,c.1045 G>A (exon 9) and c.820 T>A (exon8) were found in one patient each. Five novel mutations, c.502+1 G>T, c.924+2 T>G (exon 9) c.371A>G (exon 1) c.292_304del17 (exon-1) and c.906 C>A were identified in six patient each. Two patients were found to have a splice site mutation.. Protein prediction softwares, polyphen 2/SIFT predicted that one of the novel changes has a pathogenic effect. In EDAR gene 10 patients were found positive for different sequence variation involving exon 4,8,12 and IVS11. Conclusion: We detected variation of EDA and EDAR genes with variable phenotype of Ectodermal Dysplasia. This study should be useful for genetic counseling and prenatal diagnosis for the families enrolled in study and in expanding the database on EDA and EDAR gene.

2330M

Patient Perspectives on the Use of Electronic Health Records for Research: The MI-GENES Study. *R.A. Haddad[1], H. Jouni[1], I.N. Isseh[1], J.B. McCormick[2], C.G. Chute[3], I.J. Kullo[1].* 1) Division of Cardiovascular Diseases, Department of Internal Medicine, Mayo Clinic, Rochester, MN; 2) Division of Biomedical Ethics, Departments of Internal Medicine and Health Sciences Research, Mayo Clinic, Rochester, MN; 3) Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN.

Background: The use of electronic health records (EHRs) is increasing due to federal mandates. The myocardial infarction genes (MI-GENES) study, an eMERGE genomic medicine implementation pilot, is exploring the integration of genomic results into the EHR. The views of participants in genomic studies on the use of EHRs for research are not known. **Methods:** The MI-GENES study randomized adults aged 40-65 years, at intermediate risk for coronary heart disease (CHD) and not on statins, to receive either the 10-y risk of CHD based on conventional risk factors versus the 10-y risk based on a genetic risk score and conventional risk factors. We assessed participants' perspectives regarding the use of EHRs for clinical care and research using the health information national trends survey (HINTS). This survey was completed by all study participants at the initial study visit. **Results:** The survey was completed by 212 study participants recruited from the community (mean age 58.8±5.1 y, 47% male). The majority (n= 208, 98.1%) were in support of investigators having the ability to access and review EHRs for research; 99% of participants (n=210) felt that physicians should be able to share medical information in EHRs and 204 (96.2%) felt confident in the mechanisms in place to maintain confidentiality of their EHRs. Most participants (n=191, 90.1%) felt it is important to be able to access information in their own EHRs; 201 (95.2%) had internet access but only 119 (56.1%) used the internet to communicate with their medical providers. A significant proportion (n=154, 72.6%) was aware about the availability of direct-to-consumer genetic testing. **Conclusion:** Among participants in a genomic medicine implementation study, the majority supported the use of EHRs for research and felt confident in the mechanisms in place to maintain confidentiality of their EHRs.

2331T

A systematic review of individuals' perspectives on broad consent and data sharing in the United States. *N.A. Garrison[1,2], M. McPheeters[3,4], N.A. Sathe[3], R.R. Walden[5], E.W. Clayton[1,2].* 1) Center for Biomedical Ethics and Society, Vanderbilt University, Nashville, TN; 2) Department of Pediatrics, Vanderbilt University, Nashville, TN; 3) Vanderbilt Evidence-Based Practice Center, Institute for Medicine and Public Health, Vanderbilt University, Nashville, TN; 4) Department of Health Policy, Vanderbilt University, Nashville, TN; 5) Eskind Biomedical Library, Vanderbilt University Medical Center, Nashville, TN.

Objective: The purpose of this systematic literature review was to synthesize and evaluate evidence related to individuals' willingness to provide broad consent for biobank research and data sharing. Study Design: We searched bibliographic databases (MEDLINE® via the PubMed interface, Web of Science, National Reference Center for Bioethics Literature databases (EthxWeb, GenETHX), and Dissertation Abstracts International) for original research conducted in the United States and published since 1990. Two reviewers independently screened studies against predetermined inclusion criteria using DistillerSR™, an online systematic review tool. Next, the two reviewers independently assessed the quality of studies using a tool designed for the review. Data were extracted using a predefined instrument. Any disagreements that arose between reviewers were resolved through discussion. Results: The final sample consisted of 46 articles. Most studies involved surveys (n=26), followed by focus groups (n=8), mixed methods (n=9), analyses of consent forms (n=2), and interviews (n=1). Overall, women were more often included in these studies than men. The quality of studies ranged from good (n=18), to fair (n=26), and poor (n=2). This presentation highlights our major findings. First, individuals were more willing to provide broad consent if the samples are de-identified. Respondents' preference for broad consent over other types of consent (such as tiered or study-specific) became stronger when they learned more about the logistics and costs associated with maintaining large biobanks and when the privacy concerns were adequately addressed. Second, most respondents would allow data derived from their samples to be shared with many researchers. Preferences about data sharing, however, varied across demographics; individuals from under-represented minority backgrounds tended to be less willing to have their data shared. Furthermore, respondents were less willing to provide broad consent if pharmaceutical companies had access to the data, regardless if it is de-identified or not. The participants who were less willing to consent to data sharing cited privacy and confidentiality concerns. Conclusions: Further research is needed to explore factors that affect willingness to participate in research by demographics and biobank model in order to address remaining barriers, seek a wider diversity of opinions, and enhance public support of this research.

2332M

Attitudes and concerns related to placing genomic information in the electronic medical record: a survey of biobank participants. *A. Fiksdal[1], J. Olson[2], K. Maschke[3], JB. McCormick[1].* 1) Biomedical Ethics Program, Mayo Clinic, Rochester, MN; 2) Health Sciences Research, Mayo Clinic, Rochester, MN; 3) The Hastings Center, Garrison, NY.

Introduction: Biobanks are repositories of biological specimens and health information that provide researchers with the resources to carry out large-scale genomic research studies that would not be otherwise possible. As genomic testing becomes increasingly incorporated into clinical settings, biobanks now have the potential to serve a greater role in supporting clinical genomics in the context of personalized medicine, with some researchers envisioning placing genetic information directly into the electronic medical record (EMR). Although the integration of genomic information into EMRs and everyday medical practice holds the potential to improve care, ethical implications of such practices must be addressed as the process unfolds. Methods: Mayo Clinic recently conducted a study that involved testing samples of participants in the Mayo Clinic Biobank for known pharmacogenetic variants and automatically placing the results in the electronic medical record. We sent a follow-up survey to individuals who had either consented, refused, or did not respond to an invitation to take part in that parent study. Our survey was designed to assess the factors influencing decisions to participate in the parent study and evaluate potential concerns related to placing genomic information in the EMR. Results: Our sample consisted of 643 total respondents, including 285 individuals who consented, 178 who refused, and 180 who did not respond the invitation to take part in the parent study. Concerns related to placing genetic information into the medical record varied by group. Those consented to take part in the parent study were less likely to report such concerns than those in the refused group (25% vs. 58%). Respondents in the refused group were also less likely to report feeling comfortable with sharing their genetic information with outside researchers than those who had consented (50.1% vs. 73.3%). A majority of all respondents in all groups (86.9%) reported being concerned if their genetic information was made available to insurance companies. Conclusion: Although most respondents reported feeling comfortable with placing genetic information into the medical record, concerns related to who has access to that information remain for a notable minority of individuals. Researchers utilizing biobanks should be mindful of those concerns as they establish data-sharing agreements and develop informed consent procedures.

2333T

Research Use of STored samples (RUST)-Community Perspectives from a Developing Country. *S. Ramalingam.* Molecular Medicine, PSGInstitute of Medical Sciences and Research, Coimbatore, Tamilnadu, India.

Introduction: Human biological samples including blood, tissue and other body fluids are very important for biomedical research. Rapid development in the newer technologies increases the scope of bio banking in developing countries. In countries like India with a large number of diseases there are large volumes of left over samples and tissues which are being used for biomedical research. However there are lot of ethical and legal implications in the reuse of these samples for research. Bio banks are often sought for International collaborative research to generate newer findings that are generalizable across populations in the world. Lacks of clear guidelines compounded with lack of awareness of biomedical research especially with stored samples are important problems in countries such as India. Paucity of data on this in a populated country such as India with a huge potential for bio banking makes this study an important one. Aims: To study the perspectives and attitudes of the community towards the use of stored sample for biomedical research. Methodology: A cross sectional survey from patients and general population at a tertiary care hospital in South India and the field practice area of community medicine was carried out. Data was analyzed using the SPSS 19 software. Results: There were 225 respondents with 52.4% females and 47.6% males. 62% of the population did not want to store their samples for research. 88% opined that they will consent if their treating doctor informs them about the use of samples for research and explains about the scope of research. There was a significant association between age and consenting for research. the younger the age, more likely that they are consenting for use of their leftover samples. ($p<0.05$) but this difference was not observed between the sexes.Though they agreed for use in general research, many did not agree for genetic research. They also wanted to know the results of the study for which their samples are being used and most of them (61%) felt that the left over samples of their children will not be used for research. Conclusions: Community should be made aware about the scope of using the left over samples for research -both biomedical and genetic research. A uniform policy on research using stored sample should be drawn based on the literacy and education status and also keeping in mind the cultural context.

2334M

Genes for Good: An Online Community Study of Genes, the Environment, Health and Disease. *J.R. Forster[1], W. Li[1], M. Zawistowski[1], J. Wu[2], K. Brieger[3], S.I. Vrieze[1], G.R. Abecasis[1].* 1) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, School of Public Health, Ann Arbor, MI; 2) School of Information, University of Michigan, Ann Arbor, MI; 3) Medical School, University of Michigan, Ann Arbor, MI.

Genetic studies with traditional recruitment strategies using in-person assessments, tissue sampling, and medical evaluations are precise, but also very expensive and exclusive. Not everyone can join these traditional studies both due to limitations on available time and geography, among others. Genes for Good is a research study conducted at the University of Michigan and designed to break down these research participation barriers. We have developed a database driven dynamic survey and passive monitoring platform for large scale genetic association studies that uses the internet and social media to recruit participants and provide online surveys that can be completed by participants when convenient for them. The software is currently available as a Facebook App, allowing us to engage thousands of participants quickly and cost-effectively through a popular and widely-used social media platform.

Participants that complete a minimum number of surveys about their health and behavior are eligible for DNA analysis. We send participants spit kits to collect saliva in a simple and non-invasive manner. Extracted DNA is genotyped with a reference set of GWAS and coding variants. To motivate participants to contribute and promote literacy about genetic information, we have developed web-based software applications that provide participants information about their genetic ancestry and their survey results. Users can download their survey results in full and receive their own raw genetic data. At the time of abstract submission we are about to launch, pending an NIH certificate of confidentiality, and expect to present data on participation rates, attrition, measurement accuracy, and cost efficiency of the research design. Participants will be snowball sampled, in a process that will be started by the investigators using twitter and posts on their Facebook wall. Our software platform will then rely on existing Facebook social networking to encourage active participants to invite their Facebook friends to participate and to advertise the study by posting to their Facebook wall. We ultimately want to open our platform up to any researcher with a relevant scientific question. They will be able to use this research tool, including the genetic and phenotypic data we have collected, to answer a wide range of questions about the links between genetic information and health.

2335T

Practical Solutions for Protecting Individual Genomic Privacy. *J. Fellay[1, 2], J.L. Raisaro[3], Z. Huang[3], M. Humbert[3], E. Ayday[3], P.J. McLaren[1, 2, 4], A. Telenti[4], J.P. Hubaux[3].* 1) School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) School of Computer and Communications Sciences, Ecole Polytechnique Fédérale de Lausanne, Switzerland; 4) Institute of Microbiology, University Hospital and University of Lausanne, Lausanne, Switzerland.

The increasing availability of genomic data has major implications for personal privacy. The issues raised by genomic privacy reside at the crossroads of medicine, computer science, legislation and public policy. We here describe the design and development of new privacy enhancing technologies that aim to find the optimal balance between usability and privacy of genomic data in clinical care and in biomedical research. First, we propose a privacy-preserving algorithm for genetic risk testing in clinical care that uses homomorphic encryption and proxy re-encryption. After genomic data (e.g. sets of variants from whole genome sequencing) are generated by a certified institution, they are encrypted and stored at a centralized “storage and processing unit” (SPU). Our architecture, while preserving data privacy, enables a medical unit to retrieve the encrypted information from the SPU and to use it for individualized care. We deployed this solution in a pilot pharmacogenomics study of 180 patients participating in the Swiss HIV Cohort Study. Retrieval and processing of encrypted genotypes, for a test using 50 markers (SNPs), take less than 1 second on commodity hardware. An interim analysis showed this to be acceptable by clinicians as both usability and privacy of genomic data are preserved. Second, we developed a system to protect the privacy of mapped short reads (e.g. bam files). Millions of sequencing reads from individual genomes are stored at a SPU in encrypted form. Our scheme allows a medical unit to privately retrieve a subset of the reads without revealing the nature of the request to the SPU. In addition, the SPU can mask particular parts of the retrieved reads if they are not in the requested range or not consented by the patient (e.g., regions revealing sensitive diseases). Finally, we provide a method, based on graphical models and belief propagation, to estimate the erosion of genomic privacy of an individual when genomic data of some of his/her relatives are publicly available. We showed how a target genome can be reconstructed by relying on Mendel’s laws and linkage disequilibrium. As a result of this inference attack, we proposed different possible definitions of genomic privacy metrics.

2336M

Data Sharing in Human Pluripotent Stem Cell Research: Developing a Principled, proportional process. *R. Isasi, B. Knoppers.* Centre of Genomics and Policy, McGill University, Montreal, Quebec, Canada.

Biobanks serve as the primary resource for access to authenticated, quality-controlled and ethically sourced human pluripotent stem cell (hPSC) lines. Their scientific value is predicated on the quality of the data associated with their curated cell lines. Collections of hPSC lines including well-annotated genomic, epigenomic, and donors’- research participants’ phenotypic and demographic data are essential, given that genetic and epigenetic variations contained in the lines could have an impact on the value of their associated data. Comprehensive data curation facilitates disease modelling, drug development, and further contributes to the understanding of genetic variation and its role in normal cell behaviour. Moreover, next generation sequencing technologies - from microarrays to whole genome analysis - provide data on a wide range of healthy and sick participants with the potential to contribute to the clinical translation of cell based-therapies and personalized medicine. Several scientific and social developments are prompting reconsideration of the ways in which the scientific and ethical imperative of data sharing and security are not only conceptualized but also implemented. Data sharing is envisaged as a tripartite responsibility of data producers, users, and funders. These factors, together with reports of the ease of re-identification in the scientific literature and popular press, contribute to changing public attitudes regarding the meaning of individual privacy and attendant expectations about the In the context of hPSC research specifically, on behalf of the International Stem Cell (Funders) Forum this presentation will propose guidance for research consortia, stem cell banks, and registries, in the drafting of data sharing agreements that are proportional to the nature of cell line derivation (e.g. hESC vs. iPSCs). We offer policy recommendations for determining thresholds for the release of individual and summary data (e.g. genomic, epigenomic, phenotypic, and demographic) as well as identity profiles (e.g. STR, SNP, etc.) associated with an hPSC line to bona fide researchers.

2337T

Utilizing Online Consent and Data Collection in Studying Genetic Changes Related to Autism. *B. Smith-Packard, S. Martin, W.A. Faucett, Simons Variations in Individuals Project.* Genomic Medicine Institute, Geisinger Medical Ctr, Danville, PA.

The Simons Variations in Individuals Project (Simons VIP) was designed to characterize the natural history of individuals with copy number variants (CNVs) related to autism (e.g., del 16p11.2) by collecting detailed phenotypic information through phone and in-person evaluations of probands and family controls. In September 2010, a website (www.simonsvipconnect.org) was developed to serve as a patient registry, to provide families with community support and resources, and to function as a novel method of online study recruitment and engagement. In “Phase 1” of the study, the website was primarily used for recruitment and community engagement. Members of the online community who indicated interest in research submitted their genetics lab report to be evaluated by the Simons VIP study team for eligibility. Medical and family history information was collected by phone, and comprehensive, in-person evaluations at one of five collaborating medical centers throughout the United States.

Given the success of online subject recruitment and retention, “Phase 2” of the study was launched in February 2014 to provide an online method of data collection by utilizing remote assessments (primarily online and by phone and mail as needed) to collect phenotypic information. In transitioning to Phase 2, the community support and resource features of the website were retained while a direct portal to online research participation was designed and incorporated. Through this online portal, subjects can learn about the study, opt-in (or out) of research participation, complete online consent and assent, and complete online research surveys, allowing longitudinal contribution of phenotypic data from anywhere around the world. Since the launch of Phase 2 in February 2014, there have been 66 families consented and enrolled in Phase 2 (206 individuals total; 55 families re-enrolled from Phase 1, 11 new families). Online research surveys have been assigned to 121 individuals, of whom 99 (82%) have completed the assigned surveys. Of those completed, 81 (82%) finished in less than a week, with about one quarter (26%) of these individuals finishing surveys in less than 2 days. Our experience has shown that families are able to successfully enroll in an online research study and complete online consent/assent for research. We have found that once online research surveys are assigned, the majority of subjects will complete the surveys in a timely manner.

2338M

Whole genome sequencing of children: Consent, parental choice, and the hunt for secondary variants. *N. Monfared[1], C. Shuman[2], J.A. Anderson[9], R. Hayeems[1], M. Szego[6,7,8], R. Zlotnik Shaul[5,6,9], M.S. Meyn[1,2,3,4,5], S. Bowdin[1,4,5].* 1) The Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 4) Division of Clinical and Metabolic Genetics, Dept of Paediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 5) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 6) Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada; 7) McLaughlin Centre, University of Toronto, Toronto, ON, Canada; 8) Centre for Clinical Ethics; Providence Healthcare, St. Joseph's Health Centre and St. Michael's Hospital; Toronto, ON, Canada; 9) Department of Bioethics, The Hospital for Sick Children, Toronto, ON, Canada.

Background: Traditionally, consent processes preceding single or multigene genetic tests have focused on preparing individuals for test outcomes related to clinical presentation and/or family history. WES and WGS can also reveal medically important variants unrelated to the primary diagnosis (secondary variants) thereby challenging existing genetic counseling models. As part of The Hospital for Sick Children's Genome Clinic, a research project designed to model routine use of WGS in paediatric genomic medicine, we are studying the process of informed consent for WGS of children and the choices parents make regarding return of secondary variants (SVs) and carrier variants. Methods: We offer WGS to children undergoing clinical molecular testing for a suspected genetic disorder. Prospective participants (child and parent trios) are counseled regarding the nature of WGS and the opportunity to learn about the child's pharmacogenomics as well as childhood and adult-onset medically actionable variants (MAVs) and carrier status variants in over 2800 disease genes listed in NIH's Clinical Genomic Database. Parents are also offered the option to undergo targeted testing for adult onset MAVs and carrier status variants identified in their child. Results: Of the 321 families approached to date, 54% agreed to participate; 42% declined. The initial consenting session averages 45-60 minutes with parents and children participating in the decision making process. The mean lag between the first counseling appointment and enrollment is 9 days with most families requiring more than one appointment in this interval.There was no consensus among families regarding return of SVs. 58% of those enrolled chose to learn about all secondary adult-onset MAVs and 63% chose to learn about carrier variants, while 17% elected not to learn of any SVs. The most common concerns raised by families declining to learn about SVs included the psychological burden of identifying SVs in the child and themselves, potential impact on other family members, and/or fear of life insurance discrimination. Among parents who decided to learn of their child's adult-onset MAVs and carrier variants, 68% chose to learn their own status for the same risk variants. Conclusion: Our findings highlight the need for genetic counseling models, decision aids and resources to evolve in order to support the complex decision making processes necessary for clinical use of WES/WGS.

2339T

Pediatric whole genome sequencing: the benefit-risk calculus of parents and health care providers. *R.Z. Hayeems[1,2], J.A. Anderson[3], M.J. Szego[4], M.S. Meyn[5,6,7], C. Shuman[5,6,7], N. Monfared[6,7], S. Bowdin[6,7], R. Zlotnk Shaul[3].* 1) Child Health Evaluative Sciences, Hospital for Sick Children, Toronto, Canada; 2) The Institute for Health Policy Management and Evaluation, University of Toronto, Toronto, Canada; 3) Department of Bioethics, The Hospital for Sick Children, Toronto, Canada; 4) Centre for Clinical Ethics, St. Joseph's Health Centre, The Centre for Applied Genomics, The Hospital for Sick Children; Department of Family and Community Medicine, University of Toronto. Toronto, Canada; 5) Department of Molecular Genetics, University of Toronto, Toronto, Canada; 6) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada; 7) The Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, Canada.

Objectives Implementing pediatric whole genome sequencing (WGS) requires attending to parents' and health care providers' (HCPs) experiences with its use. The Hospital for Sick Children's Genome Clinic analyzes genomic data for both primary and secondary variants and uses a choice-based model for managing return of secondary variants (SVs) to parents of index children. In this context, we elicited parental responses to the consent process, their reasons for pursuing or declining receipt of SVs, and HCPs' perceptions of WGS. **Methods** Parents of children enrolled in the Genome Clinic were invited to participate in semi-structured interviews following completion of an informed consent discussion and pre-test counseling. HCPs who referred patients to the Genome Clinic and those peripheral to it, were also invited to participate in interviews. Transcripts were analyzed thematically, using techniques consistent with interpretive description. **Results** 58% of the parents of 174 enrolled children elected to receive adult onset medically actionable variants (MAVs) identified in their children, while 42% did not (26% declined, 16% undecided). For all 25 parents interviewed, pursuing WGS was motivated by their child's diagnostic quest and by their belief in new genomic technologies. All were satisfied with the detailed consent process; 18 opted to receive all SVs (including adult onset MAVs), 3 opted to receive only carrier results, and 4 opted to receive only childhood onset MAVs. Those who opted to receive all SVs were more enthusiastic about the potential for SVs to prevent future health risks for their child and were less worried about an over-protected childhood, psychological burden, and insurance discrimination than those who opted to receive only some SVs. Views of 6 initial HCPs interviewed are both optimistic and guarded. Upon receipt of SVs from a laboratory, HCPs feel obligated to return results to families. They orient to the best interests standard from a broad, family-based perspective. **Conclusion** Our findings suggest that parents' preferences regarding receipt of SVs range widely and are motivated by marked differences in definition and balance of best interests-oriented values. HCPs have mixed views about WGS that are informed by clinical and ethical imperatives. Given the primacy placed on achieving a diagnosis by all participants, value-based tensions related to SVs, that are experienced by parents and HCPs warrant further deliberation.

2340M

Attitude toward genetic research on children and informed assent. *Z. Yamagata[1], I. Ishiyama[2], K. Muto[3], J. Minari[4], G. Yoshizawa[4], K. Kato[4].* 1) Dept Health Sciences, School of Medicine, Univ Yamanashi, Yamanashi, Japan; 2) Teikyo-Gakuen Junior College, Yamanashi, Japan; 3) Dept Public Policy, Inst Medical Science, Univ Tokyo, Tokyo, Japan; 4) Dept Biomedical Ethics and Public Policy, Graduate School of Medicine, Osaka Univ, Osaka, Japan.

[Aim] The aim of this study is to determine the attitudes of the general Japanese public toward genetic research on children and to clarify factors related to such attitudes based on nationwide surveys conducted in 2014. [Methods] From the general Japanese population, 2,400 people (age = 20-69) were selected using a stratified two-phase sampling method. In a mail survey administered in February 2014, the participants were surveyed regarding the following topics: (1) their attitudes toward genetic testing, the genetic testing of children, and obtaining blood donations from children for research; (2) their perspective regarding informed consent; (3) their level of scientific literacy regarding genomics; and (4) their demographic information and socioeconomic status. [Results] The response rate was 56.4%(1,354/2,400). We calculated genomic literacy scores by considering the participants' (1) knowledge of genomic terminology, (2) contextual understanding of genomic terminology, and (3) awareness of the benefits and risks of genomic studies. Conducting genetic testing on children for disease susceptibilities was favored by 58.4% of participants. Regarding obtaining blood donations from children, 53.2% approved, 11.3% disapproved, and 34.8% were undecided. A higher proportion of participants with high genomic literacy levels approved of obtaining blood donations from children. The multiple logistic analysis odds ratio regarding genomic literacy was 1.28 (95% confidence interval = 1.05-1.50). Regarding whether seeking consent from children is appropriate, 47.0% of people answered that seeking consent from a child is acceptable if the child understands the details of the research and 42.9% answered that seeking consent from a child is acceptable if the child can judge the pros and cons of participating in the research. These responses were associated with genomic literacy. [Discussion] Many birth cohort studies have been conducted worldwide. The results of this study suggest that people's genomic literacy is essentially related to people's perspective on genomic research on children.

2341T

EuroGentest Guidelines for Diagnostic Next Generation Sequencing. *P. Bauer[1], G. Matthijs[2], M. Alders[3], A. Corveleyn[2], S. Eck[4], I. Feenstra[5], V. Race[2], H. Scheffer[5], E. Sistermans[6], E. Souche[2], M. Sturm[1], M. Weiss[6], H. Yntema[5], Participants of the EuroGentest Workshop on Diagnostic NGS Guidelines.* 1) Institute of Medical Genetics and Applied Genomics, Univ Tuebingen, Tuebingen, Germany; 2) Center for Human Genetics, Leuven, Belgium; 3) Department of Clinical Genetics, Academic Medical Centre (AMC), Amsterdam, Netherlands; 4) Center for Human Genetics and Laboratory Diagnostics, Dr. Klein, Dr. Rost, Martinsried, Germany; 5) Department of Human Genetics, Radboud University Medical Center (RUMC), Nijmegen, Netherlands; 6) VU University Medical Center, Amsterdam, Netherlands.

The use of Next generation sequencing (NGS) for clinical diagnostics is increasingly recognized. While the exploitation for research purposes has made tremendous advancements in human genetics, the diagnostic use brings challenges at different levels including data production and storage, and interpretation of results. Although several diagnostic NGS guidelines have been issued by the American, Australian, Dutch and British genetic professional societies, a couple of relevant topics have not yet been addressed and therefore, a EuroGentest expert group has been working on compiling, integrating and completing these guidelines. Among other statements, the guideline delivers three major definitions: "diagnostic utility", "quality scoring for NGS tests", and "reportable range". We believe that defining the 'diagnostic utility' of the NGS test is the laboratory's first duty when preparing to offer diagnostic NGS. Our scoring system for the different NGS assays depends on quality and comprehensiveness. With this system, referring physicians, patients, and stakeholders in the health system will be enabled to compare different tests offered at the market. This scoring system is new, as it does not feature in any other guideline. As far as 'reportable range' is concerned, we propose the use of 3 specific percentages depending on the reference (technical target, coverage of transcript in a gene panel, coverage with reference to the genome) which will again allow to compare individual results within runs, between tests and between laboratories. The guidelines propose a generic template for reporting NGS results as well. While dealing with informed consent, unclassified variants and unsolicited findings, again from the laboratory standpoint, is already addressed in aforementioned published guidelines, the distinction of diagnostic and research is a very relevant topic when it comes to the "duty to recontact". The latter is elaborated with a practical solution.

2342M

Attitudes toward direct-to-consumer genetic testing and its regulation in Japan. *K. Muto[1], A. Nagai[1], A. Kamisato[1], B. Zhao[1], H. Hong[1], Z. Yamagata[2].* 1) Dept Pub Policy, Inst Med Sci, Univ Tokyo, Tokyo, Japan; 2) Dept Hlth Sci, Sch Med, Univ Yamanashi, Chuo, Japan.

[Background] Direct-to-consumer genetic testing (DTCGT) in the US and China has been suspended due to alerts by the authorities in these countries. However, DTCGT can still be purchased in Japan. The Ministry of Health, Labour and Welfare (MHLW) in Japan has ignored these international alerts for a long time. The Ministry of Trade and Industry (METI) have conducted market research three times since 2007, to create a new market and to educate consumers in this field. The METI considered the regulations governing DTCGT and attached importance to quality control, scientific evidence, and informed consent. There is at present no ban on genetic discrimination or legal protection of personal genomic information in Japan. Several surveys have been conducted to clarify questions regarding public or consumer attitudes to DTCGT; however, there is still little data available in East Asia. [Purpose] The aim of this study is to clarify public attitudes towards DTCGT and the unique regulations in Japan. Method: We conducted a web-based questionnaire survey to investigate general perceptions in 2014. In total, 7,350 Japanese citizens completed the questionnaire (RR = 30.5%). We analyzed this data and compared it with past datasets. [Results] Of the respondents, 21.8% knew about DTCGT, and just 1% had purchased it before. Regarding actionable diseases, 50% of the respondents reported that they would like to receive their results via their physicians. Only 10% wanted to receive the results by post or on a website. We asked respondents to report their willingness to undergo 4 types of genetic testing based on scientific evidence. About 50% reported that they would like to undergo genetic testing on drug susceptibility genes or genetic susceptibility to actionable diseases with certain scientific evidence. However, 25-26.4% responded that they would undergo such testing even if there was no scientific evidence. More than 70% of the respondents reported that they needed ban on DNA theft and usage without permission. The perceived need for regulation on genetic discrimination in employment and in insurance was 59.7% and 52.8% respectively. [Discussion] A limitation of our study was that the percentage of respondents who reported "I can't decide" was relatively high. Compared with past studies in Korea and Taiwan, Japanese respondents showed less interest in DTCGT. Before promoting DTCGT in Japan, legislation for protection of personal genomic information is needed.

2343T

Development of the Clinical NGS Industry in a Shifting Policy Climate. *M.A. Curnutte[1], K.L. Frumovitz[1], J.M. Bollinger[2], A.L. McGuire[1], D.J. Kufman[2].* 1) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 2) Genetics and Public Policy Center at Johns Hopkins University, Washington, DC.

There is active debate about appropriate regulation of the rapidly evolving clinical next generation sequencing (NGS) industry. Interviews with industry leaders (n=19) and a web-based search of NGS companies (n=94) suggest the industry is developing services along the NGS pipeline partly in response to this regulatory uncertainty. Currently, few companies offer the full range of technology, software, and services needed to perform clinical testing. Rather, the industry tends to sell laboratories Research Use Only (RUO) products that do not stand alone as a clinical test. Clinical laboratories with expertise in laboratory standards, reimbursement, and clinical reporting then assemble these components into medical NGS assays that are classified as laboratory-developed tests (LDTs), subject to little NGS-specific regulation. Manufacturers of components of the LDTs have few clear standards to meet. This is occurring against a backdrop of evolving policy, technology, and insurance coverage. A better understanding of how the NGS business is developing in response to the current policy environment should inform regulatory efforts.

2344M

The protection of genetic privacy in the European Union and the proposed data protection reforms. *A.C. de Paor.* Centre Disability Law & Policy, Natl Univ Ireland, Galway, Galway, Galway, Ireland.

With developments in genetic science and technology, genetic information is becoming more widely accessible. However, the increasing availability of genetic information raises many ethical and legal issues that may threaten advancing science, in the absence of sufficient regulation. One such issue is the violation of genetic privacy. Genetic information is sensitive personal information that can reveal intimate details about an individual and an individual's family. In light of the sensitive nature of genetic information, the potential abuse is clear, as is the desire to protect such information from access and disclosure. Therefore, there are compelling reasons to maintain the privacy and confidentiality of genetic information. In the European Union (EU), the Data Protection Directive (introduced in 1995) provides a framework that protects the privacy of personal information. However, there is currently no substantive provision for genetic data and no specific reference to this category of information, indicating a gap in the privacy protection for genetic data in the EU. Recent developments indicate an intention to include genetic data within the scope of EU data protection laws, with the proposal of a new Regulation. With this new Regulation, the European Commission aims to develop an updated data protection framework. It is committed to reform data protection legislation, in line with the realities of today's society, and changing norms. The draft Regulation identifies 'genetic data' as a category of personal data designated for special protection. 'Genetic data' is defined broadly to include 'all data, of whatever type, concerning the characteristics of an individual that are inherited or acquired during early prenatal development', thereby presumably incorporating all genetic data including family medical history. Although this draft Regulation has encountered delay and a lack of consensus, it will enhance the protection of personal information (including genetic information) in the EU, when it is introduced. As regards protection of genetic privacy in third party contexts, this draft Regulation is welcomed as expressly recognising genetic information as a category that deserves protection. This paper analyses the issue of genetic privacy in the EU. It examines the current data protection framework, as well as the proposed reform of the EU data protection framework and the impact that these reforms will have on the protection of genetic data.

2345T

Towards an ethics "safe harbor" for global genomic research. *E.S. Dove[1], M.H. Zawati[1], E. Lévesque[1], J. Simard[2], B.M. Knoppers[1].* 1) Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Molecular Medicine, Laval University, Quebec City, Quebec, Canada.

Genomic researchers are becoming increasingly globally connected and collaborative. Though global data-driven research holds great promise for disease discoveries, the underlying research ethics review system in much of the world challenges improvements in human health and paradoxically may not improve respect for persons who participate in research. Case reports illustrate that the current system is costly, fragmented, inefficient, inadequate, and inconsistent. There is an urgent need for criteria to determine whether there is "substantial equivalency" between the principles and procedures, ethics review and governance to be able to share data across jurisdictions. Building on the international privacy "safe harbor" model that was developed following the adoption of the European Data Protection Directive in 1995, we propose a federated "Safe Harbor Framework for International Ethics Review Equivalency" that could facilitate the harmonization of ethics review of specific types of data-driven international genomic and/or disease research projects, while respecting globally transposable research ethics norms and principles. The Safe Harbor Framework would consist in part of a newly constituted organization (provisionally called the International Federation for Ethics Review, or IFER), formed by a voluntary agreement among countries, granting agencies, philanthropies, institutions, and healthcare, patient advocacy, and research organizations. IFER would be both a central ethics review body and also a forum for review and follow-up of policies concerning ethics norms for international research projects. It would be built on five principle elements: 1) registration, 2) compliance review, 3) recognition, 4) monitoring and enforcement, and 5) public participation. A Safe Harbor Framework would create many benefits for researchers, countries, and the general public. Research participants would enjoy uniform adequate protection, while researchers would enjoy ensured ethics review expertise, proportionate ethics review, and a reduction in cost, time, administrative hassle, and redundant regulatory hurdles. Most importantly, society would enjoy the maximization of the potential benefits of genomic and disease research.

2346M

Analysis of ethical and social issues of large scale genome cohort/biobanking projects in Japan. *J. Minari[1], Y. Hanabusa[1], S. Toda[2], F. Nagami[2], K. Kato[1,3].* 1) Graduate School of Medicine, Osaka University, Suita, Osaka, Japan; 2) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 3) Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto, Japan.

In the last decade or so, many large scale genome cohort/biobanking projects have been developed all over the world. Their aim is to establish infrastructures for disease and heath research. Many of the projects collect a large number of blood samples and associated life style information from healthy people, while some projects target patients with a variety of diseases.

In Japan, several large scale projects have been established starting from early or middle 2000s. One well-known example is the Biobank Japan project which has a collection of DNA and blood serum samples of 200,000 patients. Other projects started relatively recently. The Tohoku Medical Megabank has been set up in the Tohoku area which has been hit by the big earthquake in March 2011. It has begun collecting samples in 2013 with the aim of establishing two cohorts of 150,000 people. It has also completed whole genome sequencing of 1000 individuals in the fall 2013.

For the ethical, legal and social issues that have to be addressed, projects in Japan tended to work on them relatively independently. Many project have been initiated, but each project struggles to establish their own policy without sufficiently learning from existing projects. In order to maximize the resources and efforts used for the projects and to overcome the shortcomings aforementioned, it has become imperative to exchange experiences beyond individual projects. To this end, we have conducted a cross project comparison of ethical and social issues as well as governance mechanisms of several large scale projects in Japan.

Our analysis has shown several features of the current Japanese projects. One is that many of them are considering data sharing as a necessary activity, while sharing materials is still controversial especially with foreign researchers. Another interesting feature is that healthy participants for cohort studies are often recruited in conjunction with the system of health monitoring in the local community, which is feasible based on the existing trust between participants or community and researchers. Our findings are valuable for creating a common framework of research governance in the global community.

2347T

De-implementation of newborn bloodspot screening: ethical and policy issues. *S.G. Nicholls.* Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada.

Background: Newborn bloodspot screening (NBS) is one of the oldest and largest population-based screening programs in the world and involves testing blood taken via a heel prick for a number of serious and life limiting conditions. Today, NBS is applied across most continents and was recently recognized by the Centers for Disease Control (CDC) as one of ten great public health achievements of the last decade. While there has been increasing debate about criteria for expanded screening, far less attention has been paid to the exclusion or removal of conditions within existing panels. This is highly relevant given recent experiences of screening for Krabbe disease in New York: in five years of testing only a handful of cases have been identified at a cost of millions of dollars. In addition, some parents have received results where it remains uncertain if, how, or when symptoms might appear and what could be done. Other conditions, such as Duchenne Muscular Dystrophy (DMD) have been withdrawn from programs on the basis of problems with the supply of products needed to undertake the test. In part, due to limited number of programs worldwide, there has been a lack of quality control materials. Recent discussion in the field of implementation science has called for greater exploration of issues pertaining to the de-implementation of technologies and specifically "[research] to understand better the other, cognitive or political factors that facilitate or hinder de-implementation" (Prasad and Ioannidis, 2014). Objective: To explore the ethical and policy issues relating to the de-implementation of newborn bloodspot screening programs or program elements. Results: Using case studies a number of ethical and policy issues are identified. Key questions to be answered include: who are the relevant stakeholders in decisions to de-implement screening programs or program elements? What role should these different stakeholders play? What are the relevant priorities of factors such cost-effectiveness or justice? How are harms and benefits to be evaluated? Conclusions: To date there has been little consideration to the de-implementation of newborn screening programs or program elements. Experiences with other population screening programs suggest that de-implementation decisions may be controversial. Due consideration needs to be given to the appropriate roles and weight applied to ethical issues that frame the discussion of empirical evidence.

2348M

Concerns of researchers and physicians regarding the chromosomally integrated HHV-6. *V. Noël[1], R. Drouin[2], L. Flamand[3], C. Bouffard[1].* 1) Université de Sherbrooke, Sherbrooke, Canada; 2) CHR du Grand-Portage, Rivière-du-Loup, Canada; 3) Université Laval, Québec, Canada.

INTRODUCTION: The human herpesvirus 6 (HHV-6) has the unique capacity, among HHVs, to integrate to the telomeres of the chromosomes (ciHHV-6) of nucleated cells after infection. ciHHV-6 can therefore be found in gametes and be transmitted both from one generation to the next, and through organ and tissue donation. This worrisome situation affects more than 70 million people around the world and may be associated with more than 50 illnesses (pneumonitis, encephalitis, infertility, cancers, etc.), of which some are fatal. Despite its consequences on the health of individuals and populations, this phenomenon has attracted little interest in scientific and public health milieus. To highlight the importance of the problem, several researchers and physicians publish warnings in purely scientific journals. Thus, our objectives are to understand the positions of these researchers and the solutions they propose, in order to identify their clinical and socio-ethical concerns regarding the effects of ciHHV-6. METHODOLOGY: 1) Comprehensive narrative synthesis of the literature: a) research by key words in databases; b) thematic inductive analysis; c) interpretative synthesis. RESULTS: Of the 90 articles analysed, 62 contained opinions, recommendations or lines of question bearing on: a) population health, b) grafts and transplants, c) diagnostic tests, d) clinical and therapeutic care and e) pathogenicity. Certain authors also raise the idea of a screening test, as a solution for the negative effects of ciHHV-6 on population health. Finally, the problems identified by researchers and the solutions they propose incite us to revisit the notion of reproductive autonomy, as well as the principle of beneficence. We will conduct interviews based on our results with researchers and physicians who are members of the HHV-6 Foundation. CONCLUSION: This research demonstrates the importance of possessing knowledge making it possible to reflect on: a) the medical and socio-ethical consequences of ciHHV6, b) the relevance of offering a screening test and c) patient care. It could contribute to the establishment of measures that would act early on the consequences of ciHHV6, to support the work of researchers and physicians, to improve and harmonize practices for the wellbeing of affected individuals and populations.

2349T

Regulating Gamete Donation in the US: Ethical, Legal and Social Challenges. *M. Sabatello[1, 2].* 1) Psychiatry, Columbia University, New York, NY., USA; 2) Center for Global Affairs, NYU, New York, NY, USA.

The practice of gamete donation has received growing attention in the past two decades. As an increasing number of individuals and couples resort to this practice, and the children born as a result (donor-conceived-children) are reaching maturity, many are grappling with the question how to address the dilemmas that arise. Do recipient-parents have a duty to tell their donor-conceived child about his/ her genetic origins? Should the identity of the donor be disclosed or remain anonymous? Does the child have a right to know her conception story and to receive information, including identifying, medical, and genetic information, about the donor? And if a donor-conceived-child has a right to know, who has the duty to tell her about it? Focusing on the US context, this paper explores these questions having in mind the larger question of what do we as a society owe children born as a result of assisted reproduction, especially gamete donation. It underscores the ethical, legal and social dilemmas that arise, comparing and contrasting with international developments in this arena. It highlights the medical justifications for regulating this field, explores the emerging so-called right of the child to know his/ her genetic origins ("the right to know"), and considers the challenges such a right evokes to existing principles of medical ethics in the US as well as other broader societal implications of such a right.

2350M

Illuminating the changing landscape from newborn screening to newborn sequencing: Ethical, psychological, and societal implications for research and policy-making in the genomics era. *L. Bush[1], K. Rothenberg[2].* 1) Pediatric Clinical Genetics, Columbia University Medical Center, New York, NY; 2) Genomics and Society, National Human Genome Research Institute, Bethesda, MD.

As the genomics and bioethical communities engage in newly-funded research investigating the application of whole exome/genome sequencing both on healthy newborns (NBSeq) within the context of a population-based newborn screening (NBS) initiative as well as with ill newborns as a diagnostic test, it is critical that the complexity of emerging issues are illuminated to enhance understanding and exploration of the challenges ahead -- the ethical, legal, psychosocial, and policy implications on our multi-cultural society. Historically, both NBS and genetic testing of children have been fraught with controversy - such as informed consent autonomy issues, psychological impact of false positives, wide variability in access, what constitutes the best interest of the child - and concerns raised will accelerate in parallel with the mounting information generated by genomic sequencing. Foresight and thoughtful planning are especially needed in the research domain when infants are involved. Ethical deliberation with diverse voices is essential to set policy, particularly since the potential benefits of genomic findings on a newborn co-exist with a multitude of potential risks. By expanding identification of conditions exponentially through the implementation of new comprehensive genomic technologies, NBS as a social justice leveler may be jeopardized with NBSeq as some states lack resources to provide necessary follow-up, especially as many variants of uncertain significance are discovered. While tolerance to uncertainty varies, the potential for heightened anxiety must be seriously considered. Moreover, the ability to protect the interests of vulnerable newborns by traditionally mandating NBS for some conditions may be eroded with NBSeq due to informed consent and return of results blurring boundaries among a public health screening measure, research, and clinical test paradigms. From an ethical and societal perspective, this is really tough terrain. To facilitate reflection and ethical inquiry among ASHG attendees and enrich the policy process, contextual nuances surrounding this changing genomic landscape are illuminated.

2351T

Whole genome sequencing in newborn screening? A Statement on the continued importance of targeted approaches in newborn screening programmes. *B.M. Knoppers[1], H.C. Howard[2], M. Cornel[3], E. Wright Clayton[4], K. Senecal[1], P. Borry[5].* 1) Centre of Genomics and Policy, Human Gen Dept, McGill University, Montreal, Qc, Canada; 2) Centre for Research Ethics & Bioethics, Uppsala University, Uppsala, Sweden; 3) Department of Clinical Genetics and EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands; 4) Center for Biomedical Ethics and Society, Vanderbilt University, Nashville, USA; 5) Department of Public Health and Primary Care, University of Leuven, Leuven, Belgium.

Background: Whole genome sequencing (WGS) is a powerful technology that which sequences the entire genome of a person. It can be used at any stage of life, including at birth. The rapid development of WGS has led some to predict that the use of this technology will change the current practice of medicine and public health by enabling more accurate, sophisticated and cost-effective genetic testing. Some believe that the earliest applications of WGS will be restricted to settings in which genetic testing is already a routine part of clinical or public health practice, such as state newborn screening (NBS) programs. That being said, a critique of this view could point to the fact that the routine nature of NBS and thus its often implied consent, as well as its public health context, and the particular vulnerability of the population tested, could, in fact, make it an unsuited context into which to first welcome a WGS-based approach. Objectives: The aim of this project was to develop consensus recommendations on the possible introduction of WGS into NBS program. Method: To formulate the Recommendations, we: 1) reviewed the normative texts guiding newborn screening programs and relevant literature addressing the issues raised by WGS in NBS to identify both existing and emerging guidance; 2) developed a background paper and draft recommendations; 3) validated the recommendations through a series of consultations with: members of the Professional and Public Policy Committee of the European Society of Human Genetics; Human Genome Organisation Committee on Ethics, Law and Society; and P3G International Paediatric Platform. Results: Along with the members of the above mentioned Committees, we formulated recommendations with regards to the potential use of WGS in newborn screening programs. Conclusion: The responsible use of WGS within a public health programme should not be dictated by the mere availability technologies, but rather on the basis of its public health potential and the best interests of the target population(s).

2352M

Abundant pleiotropy complicates the return of genetic results and incidental findings. *J.M. Kocarnik[1, 2], C.M. Connolly[2], S.M. Fullerton[2, 3].* 1) Epidemiology, University of Washington, Seattle, WA; 2) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Bioethics and Humanities, University of Washington, Seattle, WA.

Introduction: Current guidelines regarding the return of incidental findings generated in the course of whole genome sequencing recommend only returning information that meets stringent thresholds of clinical validity and utility. The American College of Medical Genetics and Genomics (ACMG), for example, recommends that variants in 56 genes associated with 24 phenotypes are sufficiently important that they should be sought out and reported when any clinical sequencing is performed (Green et al. 2013). However, these and other guidelines do not account for the issue of pleiotropy, where a single gene is associated with multiple phenotypes that may differ in their appropriateness for return. In such pleiotropic genes, one genotype-phenotype relationship may be clinically actionable, while another might be of unproven validity. Since the genetic information is the same, returning information on the former also returns information on the latter. We previously reported that 77% of the genes in the ACMG recommendations have pleiotropic relationships (Kocarnik et al. 2014). To further characterize the potential extent of this issue, we evaluated how many genes are associated with additional phenotypes across the entire Online Mendelian Inheritance in Man catalog, a publicly accessible database of human genes and genetic disorders (OMIM.org). Methods: We recorded how many and what kind of MIM phenotypes were noted as associated with each gene listed in OMIM. Initially, only phenotypes with an assigned MIM number were counted. Results: As of 5/14/2014, 879 of the 4555 genes in the catalog (19%) listed associations with multiple unique MIM phenotypes (range 2-14). Some of these genes, such as PTEN and PIK3CA, were associated with more than ten phenotypes. When not restricted to phenotypes that had been assigned MIM numbers, 1075 (24%) of these genes listed multiple phenotypic relationships (range 2-16). Conclusion: A substantial fraction of the genes in the OMIM catalog are associated with multiple phenotypes, which could be conveyed inadvertently when returning genetic results. Pleiotropy appears to be a pervasive issue that has not been addressed in current incidental findings guidelines, which provide no guidance for how to deal with these extraneous yet easily-discoverable relationships. Future recommendations will need to consider the potential impact of such pleiotropic relationships when determining how best to return individual genetic results.

2353T

Unknown knowns? Challenging concepts for patient regarding whole-exome sequencing. *J.M. O'Daniel[1], D. Skinner[2,3], A.K.M. Foreman[1], C. Rini[4], M. Roche[5], G. Henderson[5].* 1) Genetics, University of North Carolina, Chapel Hill, NC; 2) Frank Porter Graham Child Development Institute, University of North Carolina; 3) Anthropology, University of North Carolina; 4) Gillings School of Public Health, University of North Carolina; 5) Social Medicine, University of North Carolina.

As the technical and diagnostic merits for whole-exome sequencing are assessed, parallel research is underway to elucidate and meet the challenges it may present for clinicians and patients. NCGENES is a NIH-CSER study aiming to investigate these challenges in a cohort of ~500 genetics patients. One of the project goals is to explore patient perspectives and educational needs to inform results return of both diagnostic and incidental findings. To aid this study, six focus groups were held in November and December 2013 with a total of 40 participants from the UNC genetics clinics: 31 adult patients (23 hereditary cancer / 8 general genetic) and 9 parents of minor or intellectually disabled patients. The group discussions utilized a vignette to explore several areas related to incidental findings resulting from whole-exome sequencing: terminology, desire for information, types/categories of information, patient consent/choice, and process for return of results. A number of important findings emerged. First, participants viewed incidental findings to be of possibly equal or even greater, importance than the diagnostic results. As such, a more neutral term for them was desired, such as "additional." In general, participants stated that any information that may be significant to current or future treatment should be a "routine part of the test," however; they also agreed that patients should have a right to refuse any incidental findings. Parents of minor/intellectually disabled patients had more ambivalence than adult patients in desire for incidental findings if they did not pertain to the immediate health of the child. This may be attributed to their more variable diagnoses and testing experiences. Many participants desired clinician involvement to help determine the health importance of results and what information to request; others voiced the need to learn all available information so as to become their own health advocates. Throughout all focus groups, the concept of running a medical test and yet not interpreting all the data was a significant challenge. This failure to differentiate genomic data from test results has important implications for counseling patients, and possibly whether they ask for incidental findings.

2354M

Changing Preferences About Secondary Results From Exome Sequencing: Separating Preference Setting from Informed Consent. *J. Crouch[1], S.M. Jamal[2], A.G. Shankar[2], J. Yu[2], M.J. Bamshad[2,3], H.K. Tabor[1,2].* 1) Seattle Children's Research Institute, Seattle, WA; 2) Department of Pediatrics, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA.

Debate continues over whether, how and when people should be allowed to select preferences for return of secondary results (SR) identified through clinical exome and whole genome sequencing. One unresolved issue is whether soliciting SR preferences at the time of informed consent for sequencing is an effective approach for patient decision-making. This debate is complicated by a lack of empirical data on the SR preference setting process. We recruited 145 people whose exomes had been sequenced previously to select preferences for SR return using My46, a web-based results management tool. Participants selected "yes," "no" or "I am undecided" as their preference for each of 11 SR categories: carrier status, medication response, ancestry, and disease risk (heart & lungs, brain & nervous system, blood, senses, cancer, muscle & bones, metabolism & hormones, and digestive system). Participants could change these preferences at any time. Seven percent (n=10) changed at least one of their SR preferences, either during their first preference setting session (n=6) or 10-34 days after their initial session and after participating in a study interview (n=4). Most of them (n=7) selected additional SR categories for reasons including curiosity, family history, conducting further personal "research," discussions with family members, improved understanding about which results were available, and changing priorities after recovering from a recent major surgery related to a primary diagnosis. However, 3 participants decided not to learn about certain SR information or to receive only a few results before deciding whether to receive more. Preferences for SR about blood conditions were most commonly changed (n=4), followed by ancestry (3) and cancer (2). Among 51 participants interviewed, 8 (16%) anticipated changing preferences in the future. These findings suggest that for some individuals, SR preference setting is a dynamic and iterative process influenced by discussions with others, taking time to reflect on personal interpretations of benefits and risks of specific SR categories, and responding to changing personal contexts such as health status. While more empirical data across different clinical and research contexts are needed, SR return policies should consider disassociating informed consent and preference setting in order to accommodate an informed and deliberative process, and to allow individuals to prioritize SRs and modify preferences over time.

2355T

Next generation DNA sequencing and incidental findings: consultand's opinion about the impact of being informed. *J. Forbes[1], C. Riolfi[1], M. Vianna[2], R. Pavanello[1], S. Igreja[1], T. Genesini[2], M. Zatz[1].* 1) Human Genome Res Ctr, University Sao Paulo (USP) , São Paulo, São Paulo, Brazil; 2) Institute of Lacanian Psychoanalysis - IPLA, Sao Paulo, Brazil.

Incidental findings, will be each time more common after the implantation of Next generation DNA sequencing, which is already a reality in our center (Human Genome and stem-cell research center- HUG-CELL center) in São Paulo University, Brazil. This technical advance has brought ethical and psychological concerns to geneticists and psychoanalysts, the professionals who are either responsible for genetic counseling or for providing support to those with psychic suffering caused by a diagnosis of a genetic disorder. These concerns relate to the potential advantages and disadvantages of communicating an incidental finding, that is, mutations not related to the specific patient' disorder and/or of an uncertain prognosis. The concern with what to do with this kind of finding is relatively new. For instance, in December 2013, the document "Antecipate and communicate - Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts" was produced. This document advises to researchers on when and how to tell people about such findings. However, these recommendations do not include the patient's willingness to know or not the potential risk to develop a disease in the future. From a psychoanalytic point of view, we question whether the knowledge about someone's body should be given to a third party. Is it reasonable for the society to delegate to researchers the ethical responsibility to impose on the individuals seeking a specific diagnosis the knowledge about additional unrelated genetic risks? In order to address this question we are currently studying people's opinion about the potential effects of being informed about incidental findings as a result of next generation DNA sequencing. The results of this survey, which is undergoing at the HUG-CELL center will be presented.

2356M

Compare and Contrast: A cross-national study across UK, USA and Greek Experts toward Return of Incidental Findings from Clinical Sequencing. *E.G. Gourna, N. Armstrong, S.E. Wallace.* Heath Sciences, University of Leicester, Leicester, United Kingdom.

Guidance for the return of incidental findings (IFs) from clinical sequencing is being drafted at national and international levels. However, no studies have as yet been published comparing attitudes of experts across different countries to determine where similarities and differences of opinion and practice lie. We investigated attitudes toward returning IFs from clinical sequencing across UK, USA and Greek experts. Thirty in-depth interviews were conducted with experts including clinical geneticists, lab-based geneticists and experts with legal background. The majority of experts expressed numerous concerns and these were consistent regardless of their country. Major differences were only observed between experts according to their professional background. Clinical geneticists were mainly driven to return only medically actionable findings while the other experts also favored returning IFs that had personal utility for patients. Clinical geneticists were more inclined to order targeted tests instead of Next-Generation Sequencing (NGS) to avoid finding IFs, while lab-geneticists considered NGS as a good source of useful information for the patient's health. All experts saw the need for extensive pre-test counseling where patients could be helped to make informed choices to receive Ifs, but post-test decisions were seen as more complicated and genetic counseling was considered crucial. No consensus was reached regarding the dissemination of information to family members. In all three countries, the final decision to disclose or not to disclose, usually lied with the clinical geneticist, who was also responsible for communicating the findings and was occasionally supported by a multidisciplinary team. Our study showed more similarities than differences across countries. Professional background seemed to play a more important role, while even experts with the same professional background relied on their previous experiences to inform their attitudes. Our experts could only agree that medically actionable results should be communicated to patients. As well, they acknowledged how time and labor intensive the decision process is, and asked for clear guidance to support them. However, the type of guidance seen as most appropriate differed according to the infrastructure and resources of their health system. We conclude that while over-arching principles can be created at the international level, guidance for specific practice needs to be country-specific.

2357T

Participant satisfaction with a preference-setting tool for the return of individual research results in pediatric genomic research. *I. Holm[1,2,3], B. Iles[4], S. Ziniel[2,5,6], P. Bacon[7], S. Savage[1], K. Christensen[8], E. Weitzman[2,5,9], R. Green[8,10], N. Huntington[2,11].* 1) Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA; 2) Department of Pediatrics, Harvard Medical School; 3) Manton Center for Orphan Disease Research, Boston Children's Hospital; 4) McLean Hospital; 5) Division of Adolescent/Young Adult Medicine, Boston Children's Hospital; 6) Center for Patient Safety and Quality Research, Program for Patient Safety and Quality, Boston Children's Hospital; 7) Johns Hopkins University School of Medicine; 8) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School; 9) Children's Hospital Informatics Program at the Harvard-MIT Division of Health Sciences and Technology, Boston Children's Hospital; 10) Partners Personalized Medicine; 11) Division of Developmental Medicine, Boston Children's Hospital.

Although return of individual research results (IRR) to participants in genomic biobank research may be desirable, the feasibility of IRR return and how to incorporate participant preferences are unclear. There is skepticism that participants truly understand the implications of their choice of IRR to receive suggesting that incorporating preferences may be unrealistic. The role of preferences is particularly complex in pediatrics where a parent receives IRR on their child. We developed an online preference-setting tool for the return of IRR based on the preventability and severity of a condition, dimensions that resonated with parents in interviews. The tool also allows parents to opt-out of receiving results from 4 categories: mental illness, developmental disorders, childhood-onset degenerative conditions, and adult-onset conditions not treatable during childhood. We tested this tool in an online survey emailed to parents of patients <18 years of age at Boston Children's Hospital. Parents were randomized to the hypothetical scenario that their child was enrolled in 1 of 4 biobanks with different policies for return of IRR on their child: 1) receive no IRR (None); 2) receive all IRR (All); 3) given a choice to receive all or no IRR (Binary); and 4) use the preference setting tool to choose categories of IRR to receive (Granular). Parents were given a hypothetical research result report for their child. In this analysis we report on participant satisfaction with the process, biobank, and hypothetical results received. There were 11,391 emails delivered with a response rate of 24%. The granular group was the most satisfied with the process, biobank, and hypothetical results received while the None group was least satisfied with each ($p<0.001$). The None group was least likely to agree that the biobank they were randomized to was beneficial ($p<0.001$) and most likely to agree it was a "bad thing" ($p<0.001$). The Binary and All groups rated satisfaction nearly as high as the Granular group. No significant difference was found between the groups regarding agreement/disagreement with the statement that the biobank was harmful. Recognizing that the results returned were hypothetical, our data suggest that return of IRR may increase satisfaction for biobank participation. Our results also suggest that compared to limited or no choice on results to receive, providing participants with the ability to designate their preferences leads to the greatest satisfaction.

2358M

Patients' Perceptions of Whole Genome Sequencing Results and Plans to Use Non-Actionable Findings. *L. Jamal[1,2], J.O. Robinson[1], P. Lupo[1], J. Blumenthal-Barby[1], L.Z. Feuerman[1], J.L. Vassy[3], K.D. Christensen[4], M.J. Slashinski[1], J. Wycliff[1], R.C. Green[4], A.L. McGuire[1] for the MedSeq Project.* 1) Baylor College of Medicine, Houston, TX, United States; 2) Johns Hopkins Berman Institute of Bioethics, Baltimore, MD, United States; 3) Section of General Internal Medicine, VA Boston Healthcare System, Boston, MA, United States; 4) Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States.

Background: There is much debate over best practices for returning results from clinical whole-genome sequencing (WGS). An emerging trend is to categorize WGS results into 'bins' based on their clinical utility. How patients interpret and intend to use WGS results is not well understood. Methods: The MedSeq Project is a randomized clinical trial exploring the use of WGS in cardiology and primary care settings. At baseline (before WGS), 106 patient-participants were surveyed and a subset of 28 completed in-person, semi-structured interviews. Transcripts of recorded interviews were consensus coded and analyzed thematically. Target enrollment of 200 patient-participants is expected by summer 2014. Results: Consistent with previous studies, the majority of patient-participants (87%) wanted all their WGS results. However, qualitative data show that the majority of interviewed patient-participants (n=24) had difficulty distinguishing among result types. When asked, they struggled to provide clear examples of results associated with preventable or treatable conditions as distinct from results with less clinical utility. These difficulties did not reflect poor study understanding or genetic literacy. Many patient-participants (n=17) wanted results associated with non-preventable or treatable conditions to inform their professional or financial plans, supportive care, and quality of life. Conclusion: Patients may not easily distinguish among the types of results described in WGS informed consent documents. They value WGS results for both medical and non-medical reasons. These data challenge the conceptual framework underlying many emerging models of WGS results disclosure, which define the clinical utility of results and base disclosure practices on these definitions.

2359T

What do younger breast cancer patients want to learn about individual results from genome sequencing? *K.A. Kaphingst[1], J. Ivanovich[1], B. Biesecker[2], J. Seo[1], L. Dressler[3], P. Goodfellow[4], R. Dresser[5], M. Goodman[1], M. Ray[1], S. Bell[1], K. Walton[1].* 1) Division of Public Health Sciences, Washington University School of Medicine, St. Louis, MO; 2) Social and Behavioral Research Branch, National Human Genome Research Institute, Bethesda, MD; 3) Mission Health, Fullerton Cancer Center, Asheville, NC; 4) College of Medicine, The Ohio State University, Columbus, OH; 5) Washington University School of Law, St. Louis, MO.

Background: Many communication challenges arise with return of individual results from genome sequencing. While the type of genome sequencing results that should be returned has been hotly debated, few findings are available to inform what patients would like to know about a result. These data are needed to develop communication strategies for the return of results, particularly among populations targeted for early application of genome sequencing such as patients diagnosed with breast cancer at a young age. Methods: We conducted 60 semi-structured individual interviews with women diagnosed with breast cancer at age 40 or younger to investigate what they would like to learn about individual genome sequencing results. We stratified recruitment by family history of breast cancer and BRCA1/2 mutation status to examine differences by these factors. Interviews examined interest in return of individual results for six types of gene variants (e.g., affects risk for a preventable disease, affects treatment response, uncertain clinical significance) as well as what content participants would like to learn about each type of result. Two coders independently coded all interview transcripts; analysis was based on consensus codes. Qualitative thematic analysis was conducted using NVivo. Results: Across different types of genome sequencing results, participants were most interested in learning about the health implications of the variant: "long term, what to expect...I would want to learn as much as I could so I could know how to deal with it." Other common content topics of interest included effect of the variant on disease risk for relatives, prevalence of the variant among different populations, and cause of the variant (e.g., hereditary, environment). We observed some differences in themes across strata. For example, the theme of wanting to learn everything possible about a variant was more common among women who have a known BRCA1/2 mutation. Some participants also highlighted the importance of clear communication in returning results: "I think I'd want to know everything that was in the scientific literature around that, but in layman's terms..." Conclusions: Participants most wanted to learn about the health implications of individual genome sequencing results, but were also interested in specific information about variants. These findings can inform the development of strategies to communicate with patients about genome sequencing results.

2360M

Returning incidental findings to family members of deceased research participants: Perspectives from a cancer biobank. *B.A. Koenig[1], C. Radecki Breitkopf[2], S.M. Wolf[3], M.E. Robinson[4], K.G. Rabe[2], N.M. Lindor[5], G.M. Petersen[2].* 1) Social & Behavioral Sci/Inst Hlth & Aging, Univ California, San Francisco, San Francisco, CA; 2) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 3) Consortium on Law and Values in Health, Environment, and the Life Sciences, University of Minnesota, Minneapolis, MN; 4) Biomedical Ethics Program, Mayo Clinic, Rochester, MN; 5) Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ.

Intense debate surrounds the return of incidental findings to participants in genomics research. Should researchers offer results to family members after the participant's death? Prior law and policy protect individual privacy, choice, and control, preserving each person's right not to learn the results of genetic tests. Given that actionable information, such as cancer risk, may be revealed by mutations shared among blood relatives, it is critical to develop evidence-based policy governing the management of results discovered after a participant's death. There are no rigorous empirical data documenting preferences of research participants. We report the results of the first survey to inform this debate, using a comprehensive tool based on 51 in-depth interviews with participants and family members. Surveys were mailed to 6,137 participants from a registry of pancreatic cancer patients, their family members, and healthy controls, asking preferences about return of findings. 2 repeat mailings to non-responders were sent at 30-day intervals. We addressed protecting "the right not to know" vs. offering results; weighing the wishes of individuals vs. family benefit; and participants' vs. researchers' roles and obligations. 3,645 surveys were returned (464 affected/probands, 399 spouses, 1,040 unaffected blood relatives, 1,727 controls and 15 individuals whose status could not be determined; 59% response rate). 13% endorsed protecting the "right not to know," with the majority choosing to "offer genetic research results to all participants, even at the risk of upsetting those who may not want to be offered results." 57% endorsed "whether blood relatives would benefit" as a justification for offering results vs. 43% who gave priority to "the wishes of the individual participant." 77% thought researchers bore responsibility to offer findings to participants. Participants supported offering results in spite of privacy protections, suggesting that the incidental findings debate must broaden to include consideration of return to family.

2361T

How Research Participants Value Result Confirmation in *CLIA* Compliant Laboratories. *M.Y. Laurino[1,2], A. Truitt[1], L. Tenney[2], D. Fisher[2], G.P. Jarvik[3], P. Newcomb[2,5], D. Veenstra[4], S.M. Fullerton[1,6].* 1) Institute for Public Health Genetics, University of Washington, Seattle, WA; 2) Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Division of Medical Genetics, University of Washington, Seattle, WA; 4) Pharmaceutical Outcomes Research and Policy Program, University of Washington, Seattle, WA; 5) Department of Epidemiology, University of Washington, Seattle, WA; 6) Department of Bioethics and Humanities, University of Washington, Seattle, WA.

The Clinical Laboratory Improvement Amendments of 1988 *(CLIA)* requires that clinical laboratories meet quality standards to ensure that patients receive accurate, reliable, and timely results. Given that clinical utility is one of the main justifications supporting return of results from genetic and genomic research, there is disagreement about the necessity of validation in a *CLIA* compliant laboratory prior to returning research results. Under the *CLIA* research exception, results of tests performed in non-*CLIA* compliant laboratories may be returned with the caveat that such results are used for informational purposes only and not for clinical care. At present, limited data exist regarding the extent to which research participants act to confirm results of tests performed in non-*CLIA* compliant laboratories. To address this gap, this study examined whether participants at the Fred Hutchinson Cancer Research Center Colon Cancer Family Registry site pursued the recommendation to confirm their research results in a *CLIA* compliant laboratory by using a mixed-methods approach. Of registry participants who received Lynch syndrome (LS) related findings, 29/35 (83%) participated in a 2-month and 12-month post-test disclosure survey. In addition, 8/29 (28%) also participated in an in-depth follow-up interview. Of the 29 participants, 14 had disease-causing mutations, 14 tested negative for the familial mutation, and 1 had a variant of uncertain significance. Six of 29 (21%) participants validated results in a *CLIA* compliant laboratory, and the research and clinical results were concordant in all cases. Five additional participants shared their intent to confirm their research results but had not done so within 12 months post-test disclosure. The majority of those who validated their research findings, 4/6 (67%) had disease-causing mutations. For the 8 participants who participated in in-depth interviews, reasons for pursing or not clinical validation were: acting on the recommendation of the research team and informing future clinical care vs. lacking insurance coverage and perceiving limited personal benefits. The study results provide important empirical data on how research participants value result confirmation in *CLIA* compliant laboratories, and suggests the need to better communicate its advantages.

2362M

Ethical implications of incidental findings found by array-CGH in a routine clinical population. *M. Lefebvre[1,2], D. Sanlaville[3], N. Marle[1,2], C. Christel Thauvin-Robinet[1,2], E. Gauthier[1,2], P. Edery[3], V. Malan[4], A. Afenjar[5], S. Chantot-Bastaraud[6], P. Bitoun[7], B. Heron[8], E. Piparas[9], S. Moutton[10], N. Chassaing[11], J. Lespinasse[12], S. Manouvrier-Hanu[13], M. Marti-Dramard[14], A. Goldenberg[15], C. Ferrec[16], S. Odent[17], C. Le Caignec[18], B. Gilbert-Dussardier[19], A. Toutain[20], E. Schaefer[21], N. Philip[22], F. Giuliano[23], L. Lambert[24], P. Kuentz[25], S. Julia[11], L. Olivier-Faivre[1,2].* 1) Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs de l'Est, FHU-TRANSLAD, CHU Dijon, France; 2) GAD :EA4271 « Génétique des Anomalies du Développement » (GAD), FHU-TRANSLAD, Université de Bourgogne, F-21079, Dijon, France; 3) Genetics Service, Hospices Civils de Lyon, Hôpital Femme-Mère-Enfant, and Eastern Biology and Pathology Centre, Bron Cedex, France; 4) Département de Génétique, Hôpital Necker-Enfants Malades, AP-HP, Paris, France; 5) Service de Génétique, Hôpital Pitié Salpêtrière, Paris, France; 6) APHP, Hôpital Armand Trousseau, Service de Génétique et d'Embryologie Médicales Paris, France; 7) Service de Pédiatrie, Hôpital Jean Verdier, Assistance Publique - Hôpitaux de Paris, Bondy 93143, France; 8) Department of Neuropediatrics, Armand Trousseau Hospital, APHP, Paris, France; 9) Cytogenetics Laboratory, Jean Verdier Hospital, Bondy, France; 10) Department of Clinical Genetics, Bordeaux Children's Hospital, CHU de Bordeaux, Bordeaux, France; 11) Service de Génétique Médicale, Hôpital Purpan, CHU Toulouse, Paris, France; 12) Cytogenetics Laboratory, Chambery Hospital, Chambery, France; 13) Service de Génétique Clinique, Hôpital Jeanne de Flandre, CHRU Lille, France; 14) Unité de Génétique clinique, Hôpital Nord, CHU Amiens, France; 15) Unité de Génétique Médicale, CHU Rouen, Rouen, France; 16) Laboratoire de génétique moléculaire, CHU, Brest, France; 17) Service de Génétique Clinique, CLAD-Ouest, Hôpital Sud, Rennes, France; 18) Génétique Clinique, CLAD-Ouest, CHU de Nantes, Nantes, France; 19) Genetics, University Hospital La Miletrie, Poitiers, France; 20) Service de Génétique, Centre Hospitalo-Universitaire Tours, France; 21) Service de génétique médicale, Hôpital de Hautepierre, Strasbourg, France; 22) Département de Génétique Médicale, Hôpital d'Enfants de La Timone, Marseille, France; 23) Service de Génétique Médicale, Hôpital de l'Archet II, CHU de Nice, France; 24) CHU de Nancy Pole Enfant, Centre de Référence Maladies Rares CLADEst, Service de Médecine Infantile III et Génétique Clinique, France; 25) Centre Hospitalier Universitaire de Besançon, Besançon CEDEX 25030, France.

Microarray-based comparative genomic hybridization (aCGH) is commonly used to assist in diagnosing patients presenting with a phenotype compatible with a microdeletional syndrome or with intellectual disability with or without malformation. Since aCGH is interrogating the whole genome of an individual for micro rearrangement, there is a risk to be confronted to incidental findings (IF), defined as a finding that has potential health importance for the propositus or his family and is not related to the indication of the test. In order to anticipate the ethical issues due to the IF with the generalization of new genome wide analysis technologies, we sent a questionnaire to the French clinicians and cytogeneticists about their experience with IF in aCGH within a 7 years period. We collected data from all the French University Hospitals. All the aCGH were prescribed for developmental delay, intellectual disabilities and/or malformations, the majority of IF was found in children. None were diagnosed antenatally. 39 IF were included, 24 corresponding to autosomal dominant disease with incomplete penetrance (including 11 cancer predispositions), 3 to autosomal dominant disease with complete penetrance, and 12 to X-linked disease. None heterozygotes for frequent autosomal recessive disease were reported. Therapeutic/preventive measures or genetic counselling could be argued for all cases except 3 (Charcot-Marie-Tooth type 1A, SCA11 and choroideremia). Only 1/39 was intentionally not returned to the patients, because the disease was incurable with no access to any preventive care. The choice of returning or not the results was discussed collegially in 22/39 cases. Clinicians reported difficulties to return the results in 33% of the cases. In 9/39 cases, they reported their difficulties in predicting the phenotypic expression of the disease in the future. In 6/39 cases, it was due to psychological difficulties with the family to cope with these unexpected results, and in 3/39 to the difficulty of the family in understanding the results since the question of IF had to been anticipated. Indeed, only 51% of the clinicians used consents mentioning the risk of IF. With the emergence of new technologies such as Next Generation Sequencing (NGS) in the diagnosis yield, these national experiences will be crucial to progress in the ethical reflexion in this area. In particular, it shows the importance of informing the family of the possibility of IF.

2363T

Returning Participants Results from Genome Research in National Health and Nutrition Examination Surveys. *J. McLean, S. Lukacs.* Division of Health and Nutrition Examination Surveys, National Center for Health Statistics, Centers for Disease Control and Preventio, Hyattsville, MD.

Historic DNA banks are struggling to balance consents dating back decades that specify no return of results and advances in genomic research that have increased the potential for recognizing incidental findings. The National Center for Health Statistics (NCHS) National Health and Nutrition Examination Surveys (NHANES) Genetic Program has been working to resolve the retrospective and prospective issue of returning results to survey participants. NHANES is a multistage nationally representative probability sample of the multi-ethnic, civilian, non-institutionalized U.S. population conducted for over fifty years. The NHANES DNA Specimen bank opened to internal and external researchers in 1999. In 2011, the potential issue of incidental findings in genomic research and the implication of returning incidental findings to participants prompted NCHS to host a panel of experts for a one day meeting on returning results to survey participants. From that meeting NCHS initially decided to report findings to participants and use the binning method to determine which results to return. However, the NCHS Board of Scientific Counselors recommended that NHANES obtain wider scientific input. While deliberating this issue, NHANES temporarily closed the DNA bank to new proposals in 2012. NCHS commissioned the National Academy of Science Committee on National Statistics to convene a two day workshop. “Guidelines for Returning Individual Results from Genome Research Using Population-Based Banked Specimens” was held in February 2014. The workshop included participants from the Federal government, universities, and the private sector from a variety of specialty areas including research scientists, bioethicists, lawyers, epidemiologists, and clinical geneticists The preliminary outcome of the meeting 1) retrospectively, guidance was not to return results due to consent language and lack of CLIA certification for lab processing of the DNA specimens and 2) prospectively, consents should allow for returning results and to develop a method to return results. Receiving input and guidance from a panel of experts was a valuable approach to understanding and reconciling the evolving issue of returning results for genomic research particularly for historic DNA banks.

2364M

Parental decision-making for children enrolled in exome sequencing research and attitudes toward receiving variants of uncertain significance. *J. Sapp[1], M. Crenshaw[1], D.A. Dong[2], L.G. Biesecker[1], B.B. Biesecker[1].* 1) National Human Genome Res Inst, Bethesda, MD; 2) Genzyme, Boston, MA.

Genome sequencing is increasingly being employed to identify the cause of rare diseases. Yet policies and practices regarding the interrogation, annotation, validation, and return of genomic research data to participants are currently evolving and under debate. Many participants in genomic research are children whose parents consent and make decisions regarding the return of results on their behalf. In an NIH study to find the genetic causes of rare conditions we explored participants' preferences regarding the return of both primary and secondary findings. We conducted semi-structured interviews with 22 parents of 13 pediatric probands with rare disorders. Parents were asked to reflect on the notion of being offered choices in learning their child's variant results and about their attitudes toward receiving variants of uncertain clinical significance (VUCS). All parents appreciated the opportunity to make individual choices about receiving categorized variant results for their child, and seven participants stated that they would not have participated in genomic research if they had not had the opportunity to learn their child's sequence results. Parents had mixed attitudes toward the possibility of receiving their child's VUCS. Most parents with positive attitudes about receiving these variants discussed their hope that this information could prove useful in their child's future and planned to retain and occasionally revisit these results over time. Nine parents held negative or neutral attitudes toward VUCS because of the ambiguity inherent in these findings and perceived them to be of less utility. Parents' own personal experiences with uncertainty informed their attitudes toward learning their child's VUCS; some parents referenced a learned comfort with uncertainty as part of coping with their child's illness while others perceived receipt of these results as an undesired additional burden and worried about where they would find expertise to meaningfully interpret findings of this nature. Our findings suggest that investigators employing genome sequencing where both primary and secondary variant results are offered to parents of pediatric research participants can expect enhanced participant engagement, informed consent, and decision-making. Further understanding of how these attributes affect longitudinal outcomes of receiving results may inform discussion of best practices regarding the return of genomic data to research participants.

2365T

"Bring it on!" Preference setting for secondary results from exome sequencing using My46. *H. Tabor[1,2], J. Crouch[1], S.M. Jamal[2], A.G. Shankar[2], J-H. Yu[2], M.J. Bamshad[2,3].* 1) Treuman Katz Ctr Pediatric Bioethics, Seattle Children's Hospital, Seattle, WA; 2) Dept of Pediatrics, University of Washington, Seattle, WA; 3) Dept of Genome Sciences, University of Washington, Seattle, WA.

Exome sequencing is widely used in clinical and research settings, yet it is unclear whether and how individuals should be allowed to set their preferences about what secondary results (SR) they want to receive. It is also unknown how different patient populations will respond to the opportunity to receive a variety of SR, which SR they will choose, and why or why not. We recruited 145 people whose exomes (58 adults) or child's exomes (87 parents) were previously sequenced to select SR preferences using My46, a web-based results management tool. Participants selected "yes," "no" or "I am undecided" for each of 11 SR categories (SRC): carrier status, medication response, ancestry, and disease risk (e.g., blood, cancer, etc). The majority (79%) selected "yes" for all SRC (83% of parents, 75% of adults). A subset (n=51) were interviewed about SRC selection, including 12 who refused ≥1 SRC. Most interviewees (92%) described optimism and excitement; 31% volunteered that they were not worried or anxious. A subset (n= 11) simultaneously expressed ambiguity and nervousness about possible "bad" results but said that optimism about benefits outweighed concerns. Participants used both family history and their or their child's condition or future health prioritize their SR preferences. The average number of SRC refused was 1.8 for parents and 3.5 for adults. The SRC most frequently refused by parents was ancestry (n=7) and by adults was brain and nervous system (n=10). Two parents refused SRC to defer SR until adulthood (carrier status) or avoid worry in childhood (cancer). These findings suggest that most people are comfortable receiving all or most SR, at least when organized categorically, but a subset want to be able to refuse, defer or prioritize specific SR for reasons that may vary across populations and conditions. This conclusion is tempered by the fact that a limited number of SR examples were offered for each SRC, raising the question as to whether offering more examples with a broader range of impacts would facilitate better informed and/or more selective decision-making. Either way, it is clear that a one-size-fits all opt-out SR policy in which all SR are packaged together will poorly accommodate many people.

2366M

Returning findings within longitudinal cohort studies: the 1958 Birth Cohort as an exemplar. *S. Wallace[1], N.M. Walker[2], J. Elliott[3].* 1) Health Sciences, University of Leicester, Leicester, United Kingdom; 2) Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory Department of Medical Genetics Cambridge Institute for Medical Research University of Cambridge Cambridge, UK; 3) Centre for Longitudinal Studies Institute of Education London, UK.

Population-based, prospective longitudinal cohort studies are considering the issues surrounding returning findings to individuals as a result of genomic and other medical research studies. While guidance is being developed for clinical settings, the process is less clear for those conducting longitudinal research. Research was conducted on behalf of The UK Cohort and Longitudinal Study Enhancement Resource programme (CLOSER) to examine consent requirements, process considerations and specific examples of potential findings in the context of the 1958 British Birth cohort in order to inform future decision making regarding policy. Beyond deciding which results to return, there are questions of whether re-consent is needed and the possible impact on the study, whether there is a need to introduce third parties such as genetic services to assist in the feedback process and how that will be managed, and what resources are needed to support and manage the feedback process. Recommendations are made for actions a cohort study should consider taking when making vital decisions regarding returning findings. Any decisions need to be context-specific, arrived at transparently, communicated clearly, and in the best interests of both the participants and the study.

2367T

Participant preferences regarding the return of mental health related research results from a pediatric biobank and associations with social-demographic factors, comfort and concerns with novel health information. *E. Weitzman[1,2,3], S. Ziniel[1,2,4], S. Savage[2], K. Christensen[1], N. Huntington[1,3], P. Bacon[5], C. Cacioppo[6], R. Green[7,8], I. Holm[1,6,9].* 1) Department of Pediatrics, Harvard Medical School, Boston, MA; 2) Division of Adolescent/Young Adult Medicine, Boston Children's Hospital; 3) Children's Hospital Informatics Program at the Harvard-MIT Division of Health Sciences and Technology, Boston Children's Hospital; 4) Center for Patient Safety and Quality Research, Program for Patient Safety and Quality, Boston Children's Hospital; 5) Johns Hopkins University School of Medicine; 6) Division of Genetics and Genomics Boston Children's Hospital; 7) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School; 8) Partners HealthCare Center for Personalized Genetic Medicine; 9) Manton Center for Orphan Disease Research, Boston Children's Hospital.

Preferences for return of individual research results (IRR) from a pediatric biobank may depend on features of a disease category, which can hold social, historical and medical meanings. Mental illness (MI) is one such category encompassing common problems that may complicate other conditions and carry stigma. To investigate these issues we developed a preference-setting tool for IRR based on the preventability and severity of a condition, allowing exclusion of 4 disease categories including MI. We consented and enrolled parents of Boston Children's Hospital patients <18 years of age to an online study, randomizing them to 1 of 4 hypothetical biobank designs: 1) receive no IRR (None); 2) receive all IRR (All); 3) may choose all or no IRR (Binary); and 4) use the preference setting tool for return of IRR (Granular). Following an educational video, granular arm participants completed a survey about preferences, social and demographic factors and perceptions about new health information. We report on preferences concerning MI IRR. Of 11,394 invited parents 2,718 (24%) participated in the survey, including 1,026 randomized into the granular arm (mean age 44, 91% female, 86% white). Of these 12.8% opted-out of MI IRR. No differences were found among participants who did and did not opt-out of mental health results in sex, race, ethnicity, education, history of participating in medical research or employment in the healthcare setting. Participants who set preferences for return of severe IRR and preventable IRR were less likely to opt out of MI results than individuals who excluded these results (adjusted ORs and 95% CIs for MI IRR as a function of severe and preventable IRR: 0.42, 0.32, 0.57, p<.001; and, 0.73, 0.57, 0.93, p=.01). Participants who reported a family history of developmental delay were less likely to opt out of MI IRR (8.3%) than were participants who did not (15.3%) (adjusted OR 0.62, 95% CI 0.4, 0.96, p<.05). Participants who reported a family history of MI were less likely to opt out of MI IRR (10.6%) than participants who did not (15.2%) (p<.05) but results were not significant in multivariate models. Most participants set preferences to return MI IRR concerning their child. Doing so was associated with preferences for IRR about severe and preventable conditions. Few other factors distinguished this group. Findings suggest high willingness to be informed of a common and complex problem.

2368M

Uptake and motivations to learn incidental genome sequencing results among cancer patients. *E. Glogowski[*1], Y. Bombard[*1, 2], K. Schrader[*1], J.G. Hamilton[3], M. Salerno[1], M. Corines[1], S. Patil[4], M.J. Massie[3], V. Joseph[1], Z. Stadler[1], S. Lipkin[5], K. Offit[1], M. Robson[1], *equal contributions.*
1) Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Service, New York, NY; 2) University of Toronto, Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, ON, Canada; 3) Department of Psychiatry and Behavioral Sciences, Memorial Sloan-Kettering Cancer Center, New York, NY; 4) Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY; 5) Weill Cornell Medical College, New York, NY.

Background: There are increasing efforts to disclose incidental results of whole genome or exome sequencing (WG/ES) to patients and research participants; however, limited data exist as to individuals' interest and motivations to learn their incidental data in the cancer setting. Our aims were to: (1) assess the uptake and (2) explore motivations to learn incidental WG/ES disease risk information. **Methods:** We are offering the return of incidental results to patients whose genomes/exomes are sequenced as part of other cancer predisposition discovery research, in an effort to assess the safety and utility of returning incidental results. Our protocol includes: pretest counseling and results return in person, a follow-up call to confirm result category selections prior to CLIA-validation, an in-depth interview, and distress and behavioral measures. We descriptively analyzed survey data and used content analysis of counseling notes. **Results:** Of 31 eligible patients invited thus far, 8 have enrolled in the study, resulting in a 29% uptake rate (average age: 55, 8 college-educated, 4 married and 3 employed). Thirteen (42%) were lost to follow-up and 10 (32%) declined via reply postcard, follow-up call or in person. Motivations to participate included: curiosity (6/8); interest in incidental disease risks for themselves (8/8) and/or relatives (6/8); and a desire to improve health (7/8) and help research (7/8). Reasons for declining study participation included: worry about the results (5/10), difficulty traveling into the city/clinic (6/10), concern about getting or keeping insurance (4/10) and/or a job (1/10); and lack of desire to learn genomic results (4/10) and/or disease risks (4/10). Additional reasons for declining the study offered were: a recent cancer recurrence, time constraints, and a preference for targeted clinical testing. **Conclusion:** A majority of individuals ascertained through cancer discovery research decline to learn their incidental genomic findings due to a lack of desire, worry about learning their results, or perceived barriers to participation. Disinterest in incidental findings among cancer patients appears to arise from a preference for targeted, cancer-specific information. These preliminary results provide insights into cancer patients' interest levels, motivations and perceived barriers, which may inform study designs aimed at returning incidental WG/ES results in the oncology setting.

2369T

Penetrance of Actionable Incidental Genomic Findings in Exomes from the Framingham Heart Study. *N.B. Gold[1], A.G. Bick[2], H.M. McLaughlin[3,4], P. Kraft[5], H.L. Rehm[3,4], J.G. Seidman[2], C.E. Seidman[2,6], R.C. Green[1,4].*
1) Division of Genetics, Department of Medicine, Brigham and Women's, Harvard Medical School, Boston, MA; 2) Department of Genetics, Harvard Medical School, Boston, MA; 3) Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 4) Partners HealthCare Personalized Medicine, Boston, MA; 5) Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard School of Public Health, Boston, MA; 6) Division of Cardiology, Department of Medicine, Brigham and Women's Hospital and the Howard Hughes Medical Institute, Harvard Medical School, Boston, MA.

Background: The American College of Medical Genetics and Genomics (ACMG) has recommended that pathogenic variants (PVs) in 56 genes, corresponding to 24 diseases, should be returned as incidental findings following clinical sequencing. Penetrance of these variants is largely understood based upon familial disease. We hypothesized that PVs in these genes would, as a group, be associated with suggestive clinical features (SCFs) of the corresponding diseases in a population unselected for family history. Methods: We classified 2068 variants in 56 genes from 462 Framingham Heart Study (FHS) subjects as pathogenic, likely pathogenic (LPVs), uncertain significance, or benign. Longitudinal clinical history (mean follow-up 20.4 years) and the most recent echocardiograms, electrocardiograms and lipid measurements were reviewed. Results: The fraction of subjects with PVs or LPVs who exhibited SCFs was compared to the expected fraction based on those without PVs or LPVs. PVs and LPVs were observed in 1.5% and 0.6%, respectively, of individuals. SCFs were observed in 5 subjects with PVs (71%), while only 0.9 (13%) were expected (p = 1.9 x 10-4). SCFs were only seen for cardiomyopathy, dyslipidemia and BRCA2-associated breast and prostate cancer. No individuals with LPVs had SCFs. Conclusions: These data indicate that in a population unselected for family history, subjects harboring PVs for at least a subset of the ACMG gene list are at high risk for the development of disease. Furthermore, the identification of 5 unaffected FHS subjects harboring PVs or LPVs raises issues around the debate of whether to return this actionable information to subjects unaware of their risk.

2370M

Attitudes of Genetics Professionals Toward Return of Incidental Results from Exome and Whole Genome Sequencing. *J. Yu[1], T.M. Harrell[1], S.M. Jamal[1], H.K. Tabor[2], M.J. Bamshad[1,3].* 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA.

Professional recommendations for return of results from exome and whole genome sequencing (ES/WGS) have been controversial. The lack of clear guidance about whether, and if so, how to return ES/WGS incidental results limits the extent to which individuals and families may benefit from ES/WGS. Genetics professionals' perspectives, particularly of those at the forefront of using ES/WGS in clinics, are largely unknown. Data on stakeholder perspectives could help clarify how to weigh expert positions and recommendations. We conducted an online survey of 9857 genetic professionals to learn their attitudes on return of incidental results from ES/WGS and recent American College of Medical Genetic and Genomics Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. Of 847 respondents, 760 completed the survey. The overwhelming majority of respondents think that incidental ES/WGS results should be offered to adult patients (85%), healthy adults (75%), and parents of a child with a medical condition (74%). The majority think that incidental results about adult onset conditions (62%) and carrier status (62%) should be offered to parents of a child with a medical condition who has undergone ES/WGS. About half think that offered results should not be limited to those deemed clinically actionable. The vast majority (81%) think that patient / family preferences should guide return. Clinician and researcher attitudes differed significantly on a number of questions. Multivariate analysis showed that researchers were nearly twice as likely as clinicians to be open to offering non-actionable results (p=0.008) and 85% more likely to agree with returning adult onset results to parents of children who have undergone sequencing (p=0.027). Genetic professionals' perspectives on return of ES/WGS results differ substantially from current recommendations and vary between professions. This underscores the need to establish clear purpose for recommendations on return of incidental ES/WGS results as professional societies grapple with developing and updating recommendations. Further, professional differences in attitudes suggest the need for greater discussion about return in clinical versus research settings.

2371T

Participant Experiences Receiving Incidental Findings Through Genome Sequencing. *K.L. Lewis[1], P. Connors[2], M.F. Wright[1], T. Hyams[1], L.G. Biesecker[1], B.B. Biesecker[2].* 1) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Social and Behavioral Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

As an increasing number of people have genomic sequencing, there is still debate about whether and how to return individual results. Several studies have examined patients' hypothetical preferences for the return of results; however, little is known about participants' actual experiences. We report on a qualitative interview study of research participants' perspectives on the medical and psychological impacts of receiving results. Participants were recruited from the ClinSeq® study, which enrolls mostly health participants and offers genome sequencing results to individual participants. Subjects are offered a choice about receiving results unless they are of urgent medical significance, and given results in person by a physician and a genetic counselor in most cases. Twenty-nine participants (67% of those eligible) agreed to participate and were interviewed. Participants received results for a wide range of conditions, the majority of which were medically actionable and pertained to cancer, cardiomyopathy or arrhythmia susceptibility. The interviews were recorded, transcribed, and coded. Half of the interviews were reviewed by an independent, second coder; 80% consensus was achieved and discrepant codes were reconciled. Most participants reported sharing their result with at least one physician, but had no changes to their medical care. Participants stated that they valued the result regardless of its medical utility because it encouraged them to do their own research on the condition, prompted them to be vigilant in watching for symptoms or seeking out screening, helped them feel prepared for what might lie ahead, and satisfied their curiosity. Roughly half of the participants received a result that was unexpected, but only three felt distressed. These findings provide insight into participants' perspectives of the value of their results and whether they are likely to share their results. The data provide early evidence regarding the clinical and emotional reactions of participants to results and suggest that, among early adopters, adults are resilient even when results are unexpected and have serious health implications.

2372M

Does prior comfort with new health information influence parents' preferences for receiving genetic research results about their children? *N. Huntington[1,2], S. Savage[3], S. Ziniel[4,5], P. Bacon[6], E. Weitzman[2,4,7], K. Christensen[8], R. Green[8,9], I. Holm[2,3,10].* 1) Division of Developmental Medicine, Boston Children's Hospital; 2) Department of Pediatrics, Harvard Medical School; 3) Division of Genetics and Genomics, Boston Children's Hospital; 4) Division of Adolescent and Young Adult Medicine, Boston Children's Hospital; 5) Center for Patient Safety and Quality Research, Boston Children's Hospital; 6) Johns Hopkins University School of Medicine; 7) Children's Hospital Informatics Program, Boston Children's Hospital; 8) Division of Genetics, Department of Medicine, Brigham and Women's Hospital; 9) Partners Personalized Medicine; 10) Manton Center for Orphan Disease Research, Boston Children's Hospital.

Background: In genomic biobank research, questions remain about how to maximize benefit and minimize harm when returning individual research results (IRRs) to parents about their children. Result- and parent-specific factors can influence a parent's reaction to receiving IRRs. We examined the impact of comfort with health information on parents' preferences for receiving IRRs about their children. **Methods**: An online survey was completed by 2,718 parents of recent Boston Children's Hospital patients (response rate 24%). Respondents were randomized to one of three survey groups: 1)"No choice"-half receive All IRRs, half receive None); 2)"Binary choice"-parent chooses to receive All or None; 3)"Granular choice"-parent sets preferences for IRRs based on preventability and severity of conditions, plus options to opt-out of specific condition categories (mental health, developmental disorders, degenerative conditions, adult-onset conditions). Parents then viewed a hypothetical IRR report reflecting their survey condition/preferences and rated their satisfaction with the report (scale 1-10). Before setting preferences, parents were asked 9 questions about their comfort with getting new health information about their child. A cluster analysis on the responses identified three types of Health Information Comfort (HIC): 1)Positive (HIC-P: sees benefits of knowing with little harm); 2)Ambivalent (HIC-A: sees both benefits and harms); and 3)Negative (HIC-N: sees harms with little benefit). **Results**: IRRs received were categorized as All, None or Some (Granular group only). HIC was related to results preferences for the Binary ($p<0.001$) and Granular ($p<0.001$) groups: HIC-P were most likely to want IRRs, HIC-N least likely and HIC-A were in-between. In the Binary group, most HIC-A chose All results, in a proportion similar HIC-P. In the Granular group, HIC-A preferences shifted towards Some results, with proportions similar to HIC-N. As expected, there was no relationship between HIC and results received in the No Choice condition. HIC was related to satisfaction with results in the Binary ($p=0.02$), Granular ($p<0.001$) and No Choice ($p=0.008$) conditions: HIC-P>HIC-A>HIC-N. For all HIC types, satisfaction was highest in the Granular group and lowest in No Choice. **Conclusion**: Finding the right alignment between comfort with knowing health information and preferences for which IRRs to receive may serve to maximize the benefits of returning IRRs to biobank participants.

2373T

Navigating Contested Ideas about DNA Ancestry Testing in American Indian and Freedman Populations in Oklahoma. *J.W. Blanchard[1], G. Tallbull[1], C.M. Wolpert[2], M. Foster[1], C.D. Royal[2, 3].* 1) Department of Anthroplogy, University of Oklahoma, Norman, OK., Select a Country; 2) Institute for Genome Sciences & Policy Duke University Medical Center Durham, NC; 3) Department of African and African American Studies Duke University.

Commercial genetic and genomic ancestry testing has been available for sale on the markets for over ten years. Today there are nearly 200 DNA ancestry products, packages, and services being offered by about 40 different companies. Direct-to-consumer DNA Ancestry tests incite critical skepticism among various publics and scholars alike, especially with regard to the potential for such tests to perpetuate racial essentialism, to challenge personal identities and community histories, and to promote unrealistic ideas about the usefulness of test results. The long-term implications of these relatively new and increasingly accessible genomics technologies are largely unknown for indigenous and other historically exploited communities that have long been suspicious of the goals and local value of genetic exploration. This presentation explores the processes whereby researchers in an ongoing, long-term study engage Oklahoma tribal and Freedmen populations about the risks and benefits of community participation in genomic research and in genetic ancestry testing. These communities, in particular, have especially timely ideas about the use of genetic testing for ancestry purposes. At the center of ongoing enrollment disputes between some Oklahoma tribes and Freedmen individuals are historically contested ideas about "blood," "ancestry," and "citizenship". This suggests that the availability and use of ancestry testing for members of these communities has potentially significant implications for both those seeking to authenticate specific ancestral identities to secure tribal membership, as well as for tribal nations whose sovereign rights to determine membership may be challenged by the introduction of such testing.

2374M

Direct-to-Consumer Ancestry Testing: Psychological and Behavioral Impacts. *D.L. Boeldt[1,2], A.G. Lucy[5], L. Ariniello[1,2], O. Libiger[1,2,4], V. Bansal[1,2,4], E.J. Topol[1,2,3,4], L. Kessler[5], N.J. Schork[1,2,4], C.S. Bloss[1,2].* 1) Scripps Translational Science Institute, La Jolla, CA; 2) Scripps Health, La Jolla, CA; 3) Scripps Clinic Medical Group, La Jolla, CA; 4) Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA; 5) Arcadia University, Glenside, PA.

Introduction: There is a growing population of consumers who are purchasing direct-to-consumer (DTC) genetic ancestry tests. Ancestry.com, a DTC ancestry testing company, recently reported that more than 400,000 individuals have purchased their AncestryDNA test. Despite consumer interest, however, few studies have examined the impact of ancestry testing at an individual level. The current study aimed to evaluate attitudes towards ancestry testing and characteristics associated with response to testing in a large cohort. **Methods:** The Scripps Genomic Health Initiative (SGHI) is a longitudinal cohort study originally designed to assess response to DTC genomic testing for complex disease risk. SGHI participants who completed a baseline health assessment were also provided with personal genetic ancestry testing performed by a team of scientists at the Scripps Translational Science Institute. Individuals who viewed their ancestry results were invited to complete an additional optional survey (http://www.stsiweb.org//sghi/) to assess response to the information provided. Participants also self-reported race/ethnicity prior to receiving their ancestry results. **Results:** 3,465 individuals were provided with personal genetic ancestry testing, 1,317 logged into the website to view their results, and a subset of 322 (24% response) completed the ancestry response survey. Although 46% of respondents reported that their ancestry results were surprising or unexpected, <1% reported being distressed by the results. In addition, 39% reported a change or possible change in their perception of their cultural roots, 24% a change in the likelihood that they would travel to certain parts of the world, and 21% that the results reshaped their personal identity. A majority (81%) of participants stated that they planned to share the test results with their family members, and 12% reported the intention to share results with a healthcare provider. **Conclusion:** Data from this selected sample of ancestry test recipients suggests that although genetic ancestry testing does not induce stress, it does impact a notable proportion of individuals with respect to cultural perceptions, personal identity, and propensity to share genetic information with family and healthcare providers. Our findings suggest that ancestry information may be particularly useful in promoting sharing and thus peer-peer education in genomics.

2375T

Motivations for Participation in Genomic Research in Hispanics vs. Non-Hispanics. *M.L. Cuccaro[1, 2], C.P. Manrique[1], M.A. Quintero[1], R. Martinez[1,2], J.L. McCauley[1,2].* 1) Dr. John P. Hussman Institute for Human Genomics , University of Miami Miller School of Medicine, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL.

Racial and ethnic minorities are underrepresented in genomic research. This underrepresentation has the potential to exacerbate existing health disparities. Increasing the engagement of minorities in genomic research is one strategy to reduce health disparities and to identify population specific genetic risk for different diseases. As part of ascertainment for a study of genetic risk for multiple sclerosis (MS), we surveyed individuals who had either agreed or refused to participate about their reasons for this decision. Our dataset (N=101) was predominantly female (80%) and mainly white (96%). The most frequent age ranges were 46-55 (27%) and 56-65 (20%). Hispanics constituted (80%) of the sample. Among the 101 individuals surveyed 95% had agreed to participate in the MS genetics study. Individuals were asked to identify, from 11 possible choices, the reasons which most influenced their decision. The most frequently cited reasons for participating in the MS genetics study were helping to find a cure (56%), having the disease (46%), and helping future generations (37%). We tested whether these three reasons differed by ethnic group using Pearson chi-square. Hispanics and non-Hispanics did not differ in their rates of endorsement of participating in genomic research to help find a cure (HI 56%, NHI 56%, $p=0.97$) or helping future generations (HI 34%, NHI 50%, $p=0.26$). However, the groups differed significantly on endorsement of having the disease as a reason to participate in the study. Specifically, 52% of the Hispanics endorsed having MS as a reason to participate vs. 19% of non-Hispanics ($p=0.015$). We also tested whether these reasons differed by age group and found no differences. However, among the youngest individuals (18-35 years) 74% endorsed finding a cure as a reason for participating vs. 39% of the oldest group (>56 years). Our findings offer a preliminary perspective about the motivations for participating in genomic research among Hispanics and non-Hispanics. The results suggest that motivations to participate in genomic research may differ by ethnic group and that the impact of having a particular disease (e.g., MS), may be more likely to motivate involvement in Hispanics. Larger samples across different diseases can shed light on group motivations to participate in genomic studies. Such findings can help tailor recruitment strategies and engagement to aid in closing the gap in health disparities in genomic research.

2376M

Reasons for declining preconception carrier testing using genome sequencing: Implications for research and practice. *M. Gilmore[1], C. Young[2], M. Leo[2], J. Davis[2], T. Kauffman[2], E. Esterberg[2], F. Lynch[2], B. Wilfond[3], K.A.B. Goddard[2].* 1) Department of Medical Genetics, Kaiser Permanente Northwest, Portland, OR; 2) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 3) Truman Katz Center for Pediatric Bioethics, Seattle Children's Hospital, Seattle, WA.

Genome sequencing is an emerging technology for use in clinical practice, and its utility is not well understood for applications such as preconception carrier testing. An important facet of understanding the implementation of this new technology is to investigate reasons why potential participants decline expanded carrier testing using genome sequencing. We are investigating preconception expanded carrier testing in generally healthy couples. Potential participants are identified because their provider ordered at least one preconception carrier test, typically cystic fibrosis screening. They are contacted by phone to see if they are interested in the study. If they decline participation, they are invited to complete a survey on their reasons for declining as well as demographic information. We collect similar information for those who consent to join the study. We are comparing participants and decliners on factors such as race, education level, income and characteristics related to knowledge of genetics, such as knowing a family with a child with a genetic condition, to determine whether there may be any participation bias. In addition to the survey, we have implemented qualitative approaches including journaling by study staff and focus groups to enhance our understanding of reasons why potential participants decline the use of expanded preconception carrier testing using genome sequencing. Of the women who declined participation at any time after initial contact (N=45), seventy-two percent have completed the decliner survey. The most common reasons cited were a lack of time (fifty-five percent), not interested in participating (twenty-four percent), didn't want to know about their risk (twenty-one percent), and worried about discrimination (nine percent). Synthesis of qualitative data support the reasons captured in the survey. The high response rate for the decliner survey may suggest a high level of interest in preconception expanded carrier testing for this population. Concerns about discrimination are less common than we anticipated as a reason to decline to participate. Further analysis of decliner survey results will enhance our understanding of genomic sequencing technology implementation in the clinical setting.

2377T

Personal Genomics Online in Australia: A mixed methods study of Australian consumers' knowledge, attitudes and experiences of direct-to-consumer personal genome testing. *J. Savard[1], J. Mooney-Somers[1], A. Newson[1], I. Kerridge[1,2].* 1) Centre for Values, Ethics and the Law in Medicine (VELiM), School of Public Health, The University of Sydney, New South Wales, Australia; 2) Hematology Department, Royal North Shore Hospital, New South Wales, Australia.

Health, including both one's current and future health, is increasingly being understood in genetic terms, with personal genomic information marketed as a means by which individuals can "know" and "control" their health and reduce their likelihood of illness. Unfortunately, there is often a discrepancy between the expectations that consumers have of what personal genome tests may provide and what explanatory power they actually have. In part, this is because we still lack scientific and epidemiological knowledge needed to usefully interpret the information these tests divulge. In this paper we present findings from the Australian Genetics Online Study - a mixed methods exploration of the beliefs, knowledge, expectations and experience of Australian consumers regarding direct-to-consumer personal genome testing (DTC-PGT). The first phase of the study, which involved an online public survey, explored lay knowledge, expectations and experiences of DTC-PGT. The results of this study suggest that Australian consumers have an interest in genetic testing - both for health and ancestry purposes, and many value this information as much as they value other sources of personal information - but appear disinclined to pursue it. The second phase of the study, semi-structured interviews with potential and current consumers of a personal genome testing service, revealed a wide range of motivations for pursuing testing along with varying degrees of understanding with respect to what the test could (and did) tell the consumer about themselves, their future health and/or their ancestry. The final phase of the project, an autoethnography, documented the researcher's (Jacqueline Savard's) journey through the process of becoming a consumer of DTC-PGT. The results of the Genetics Online Project provide the first quantitative and qualitative account of DTC-PGT in Australia. These findings will be discussed in light of ongoing debates about the impact that publicly accessible genetic testing is having on the design and delivery of health care, information about health care and about the impact of genetics on the social construction of health, illness and identity.

2378M

A novel scale to assess perceptions of uncertainty in genome sequencing information. *B. Biesecker[1], W. Klein[2], K. Lewis[3], L. Biesecker[3], P. Han[4].* 1) Soc Behavioral Res Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892; 2) Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892; 3) 3Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892; 4) Maine Medical Center Research Institute, Center for Outcomes Research and Evaluation, 81 Research Dr Scarborough, ME 04074.

The scope of uncertainty in genomic sequence information is vast. Yet our understanding of how adults perceive medical uncertainty is poorly understood. Such perceptions are likely to play a key role in deciding to undergo genomic sequencing and in understanding and acting on the findings. Guided by a taxonomy of medical uncertainty developed by Han and colleagues, we aim to further understanding of perceptions of uncertainties. The NHGRI ClinSeq® cohort study represents a strategic opportunity to assess predictors of decisions to learn and use genomic sequencing information. We present analytic data on baseline perceptions of uncertainty originating from a novel scale. Methods: The Perception of Uncertainties of Genome Sequence Results (PUGSR) scale assesses perceptions of uncertainty specific to genomic sequencing. Development of items was informed by focus group findings. Ten items assessed three sub-domains of uncertainty: medical, affective and trustiworthiness. Results: Five hundred seventeen ClinSeq® participants completed the scale prior to making a decision about whether to learn their sequence results. There was a normal distribution in responses with an overall mean uncertainty score of 3.5/5 (SD 0.58) with high internal consistency (α=0.835). Results of exploratory factor analysis revealed three factors; medical (four items), affective (four items) and trustworthiness (two items) uncertainty about sequence information. Perceptions of medical uncertainty were highest and contributed most to the variance (2.85). These data can be used to guide informed consent for genome sequencing, to clarify the role of uncertainties in decision-making, and to anticipate uncertainties of greatest concern to recipients of genome sequencing information.

2379T

The influence of false positive results from newborn screening for cystic fibrosis on cascade testing and family planning. *Y. Bombard[1], F.A. Miller[2], R.Z. Hayeems[2,3], C.J. Barg[2], S.J. Patton[2], P. Durie[4,5], J.C. Carroll[6], P. Chakraborty[7,8], B.K. Potter[9], K. Tam[10], L. Taylor[11], E. Kerr[12], C. Davies[7], J. Milburn[7], F. Ratjen[4,5,11], A. Guttmann[2,4,5,13].* 1) Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada; 2) Institute of Health Policy, Management and Evaluation, University of Toronto, ON, Canada; 3) Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada; 4) Division of Paediatric Medicine, Department of Paediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 5) Department of Pediatrics, Faculty of Medicine, University of Toronto, ON, Canada; 6) Department of Family and Community Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada; 7) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 8) Department of Pediatrics, Faculty of Medicine, University of Ottawa, ON, Canada; 9) Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, ON, Canada; 10) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 11) Division of Respiratory Medicine, Department of Pediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 12) Department of Psychology, The Hospital for Sick Children, Toronto, ON, Canada; 13) The Institute for Clinical Evaluative Sciences, Toronto, ON, Canada.

Objectives: The majority of infants with false positive (FP) newborn screening (NBS) results for cystic fibrosis (CF) are eventually identified as carriers of one CF mutation. Carrier detection can enable reproductive planning, provided results are appropriately shared. We examined carrier communication, carrier testing and family planning following NBS detection of carrier infants. **Methods**: We surveyed mothers of infants confirmed to be FP after confirmatory testing (Time-1) and one year later (Time-2) to ascertain self-reported communication, intended and actual carrier testing, and subsequent reproductive choices. Mothers were invited for an interview. We report cross-sectional analyses for Time-1, and longitudinal analyses for carrier testing uptake by Time-2. **Results**: We received completed surveys from 134 of 245 mothers (55%) at Time-1 and 96 of 214 (45%) at Time-2, and interviewed 66 mothers. 77 of 127 mothers (61%) reported one mutation in their infants (i.e., carrier-mothers); 30 (24%) reported no mutations and 20 (16%) were unsure (presumed non-carriers). Carrier testing uptake was 55% (n=23/42) among carrier-mothers. 13 intended to have carrier testing at time-1 and of those 6 were tested by time-2. 91% of carrier-mothers told relatives they may be carriers (n=70/77), as did 48% (n=24/50) of non-carrier mothers. Results were expected to influence family planning for 35% of carrier-mothers (n=26/74) and 24% of non-carrier mothers (n=11/46). Interviewees valued carrier information whether or not they intended to share it with relevant relatives. **Conclusion**: While mothers describe reproductive value in carrier results, survey data are equivocal given moderate intended and actual CT uptake and limited influence on family planning.

2380M

Personalized medicine in the context of prenatal diagnosis and medically assisted procreation. *G. Lapointe[1], R. Drouin[2], C. Bouffard[1].* 1) Université de Sherbrooke, Sherbrooke, Québec, Canada; 2) CHR du Grand-Portage, Rivière-du-Loup, Québec, Canada.

Personalized medicine (PM) is based on the analysis of an individual's cumulated genomic, behavioral and environmental data. It aims to evaluate, on a personalized basis, a person's risk of suffering from illnesses and to determine the most efficient and safe preventative and therapeutic approaches. The 4P medical approach (predictive, preventative, personalized and participative) adds a participative dimension in which citizens are more responsible for and in control of their own health. Prenatal diagnosis (PND) and medically assisted procreation (MAP) are also affected by the new role of individuals in questions of health. Yet, as the consequences of decisions made in the context of PM are not the same for the fetus and embryos as they are for adults and children, this situation raises particular clinical and socio-ethical concerns. However, while a vast literature exists on PM, it does not appear to take into account PND and MAP. Yet the information that genomic tests reveal about a fetus and embryo raise major clinical and socio-ethical issues that deserve to have their place in the PM movement. Based on the PM literature, our objectives were the following: 1) create a repertoire of the major PM issues, 2) determine whether PND and MAP have been discussed in terms of PM, 3) identify the clinical and socio-ethical issues that emerge. METHOD: Scoping review of academic and grey literature (perspectives, legal documents, government reports, etc.): 1) publication research by key words in databases (Pubmed, EBSCO, etc.), 2) comprehensive and comparative thematic analysis. RESULTS: Of the 98 articles analyzed, 5 major themes emerged: the level of responsibility, equity, aptitude to understand information, genetic counseling and the importance of the opinion of citizens. Yet, few article discuss PND and MAP in terms of PM (10/98). When the subjects are broached, it is in the context of non-invasive prenatal testing (4/10) and the instances involve concerns such as social control of reproduction, eugenics and the risk of coercion. No study mentioned participation and the opinions of citizens. CONCLUSION: From a clinical as well as a socio-cultural level, the participative dimension of PM implicates a societal responsibilization that will be problematic in the context of PND and MAP. It is imperative to develop knowledge based on convincing data concerning the positions of stakeholders.

2381T

ACE, ACTN3 and ADRB3 polymorphisms in elite soccer players according to the team position. *L. Ruzic, V. Vucetic.* Faculty of Kinesiology, University of Zagreb, ZAGREB, Croatia.

The aim of the this paper was to investigate the relationships between three polymorphisms in ACTN3, ACE and ADRB3 genes and playing positions in elite soccer players of the Croatian national selection and the national championship winner. We examined 44 top-level Croatian soccer players (4 goalkeeper, 13 defense -fullback, 15 middle field and 12 forward attackers). Depending on the polymorphisms on the above mentioned genes the subjects obtained the indices between 0 and 6; 0 being "no strength type" and 5 being the highest "strength/explosivness type" . We looked into the relationship between the aerobic and anaerobic components of the position at which the player plays and the obtained indices. As the sample comprised only 4 goalkeeper their genetic strength indices were as follows: 2 goalkeepers type 6, 1 goalkeeper type 5 and 1 goalkeeper type 4 which strongly indicated the genetic tendency of the goalkeepers towards the explosive/strength activities. Among other players we should emphasize that only 1 player (2%) in middle field was type 2 while 28% of the others were type 3, 40% type 4 and 30% type 5 which lead us to believe that even though the soccer is highly demanding aerobic game it still requires a big amount of explosive strength and speed so the endurance type genotypes are not largely present among top level soccer. The Spearman rank correlation showed a significant but weak correlation between the playing position and the strength genetic index ($p<0.05$) meaning that fullback and attack positions had higher indices, being more explosive whilst the midfielders showed lower indices, e.g. tendency toward endurance. An interaction of three polymorphisms (ACE and ACTN-3 and ADRB3) might be related to the position at which the soccer player plays but it probably does not have a high usefulness selection of the younger soccer players as the positions are attributed later in career. Larger scale studies and more genes analyzed would give better insight in the significance of the genetic predispositions for soccer players. Pimenta, E.M., Coelho, D.B., Veneroso, C.E., Barros Coelho, E.J., Cruz, I.R., et al. Effect of ACTN3 geneon strength and endurance in soccer players.(2013). J. Strength. Cond. Res. 27, 3286-3292. Massidda, M., Corrias, L., Ibba, G., Scorcu M., Vona, G., Calò, C.M. Genetic markers and explosive leg-muscle strength in elite Italian soccer players. (2012). J Sports Med Phys Fitness. 52, 328-334.

2382M

Influence of the social environment on the development of breast cancer through epigenetic modifications: A comprehensive scoping review. *O. Martin[1], R. Drouin[1,2], C. Bouffard[1].* 1) Division of Genetics, Department of Pediatrics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, Québec, Canada; 2) CHR du Grand-Portage, Department of Obstetrics-Gynecology, Rivière-du-loup, Québec, Canada.

For 20 years now, several research groups have attempted to establish a link between stressful events and susceptibility to developing breast cancer, but results have been contradictory. While some show evidence supporting the direct link between acute and chronic stress at a young age, traumatic events and breast cancer development, others suggest that stressful events cannot be considered risk factors.Nonetheless, it has been established that epigenetic modifications may play a predominant role from the onset of breast cancer and throughout its development. We can't dismiss the idea that social environments that submit an individual to an acute and chronic stress might modify methylation patterns, thereby increasing susceptibility to certain chronic illnesses including cancer. The results presented here stem from a comprehensive and comparative analysis of academic literature bearing on the potential of the social environment, mainly acute and chronic stress, to induce epigenetic modifications linked to breast cancer among individuals with or without a known mutation in terms of a breast cancer susceptibility gene. Methodology: Comprehensive scoping review of academic literature: 1) research by keywords in: databases (Pubmed, EBSCO, etc.), reference lists of key papers, hand searching of journals not indexed on databases. Rigorousness criteria: English and French articles, from 1980-2014, epigenetics, breast cancer; 2) comparative thematic analysis. Results: Despite several publications on epigenetic modifications and the environment as factors of susceptibility of breast cancer development, the social environment remains a largely unconsidered subject. While it is ever more accepted that acute and chronic stress might be a susceptibility factor, no study has yet been able to draw a real correlation between epigenetic modifications induced by stress and the development of breast cancer. Important knowledge gaps currently exist regarding the links between social environments, stress, epigenetic modifications and breast cancer. Among other things, the effect of these relationships in a heterozygous patient for a BRCA1 or BRCA2 gene remains unknown. Conclusion: There is still a great deal to be done to obtain sufficient data to establish that social environments and human behaviour may induce epigenetic modifications that may increase the risk of developing breast cancer. However, the latest discoveries in the field tend to lean towards this possibility.

2383T

Genetic Ancestry Testing and Identity: Exploring the Relationship. *C.M. Wolpert[1], J.D. Powell[1], K. Haynie[3], R. Kittles[4], C.D. Royal[1,2].* 1) IGSP, Duke University Medical Center, Durham, NC; 2) AAAS, Duke University, Durham, NC; 3) Dept. of Political Science, Duke University, Durham, NC; 4) University of Illinois Chicago, Illinois.

Genetic ancestry testing (GAT) uses individuals' DNA to estimate their genealogical and geographic ancestry. Scholars have hypothesized that GAT provides information that may cause individuals to reshape their identities. Yet, there are limited empirical data to support this view. Furthermore, there is very little discussion in the research literature about the types of identity that are likely to be affected by GAT. In this study we sought to investigate whether and how GAT test takers (N = 455), the majority of whom self-identified as non-white, spontaneously referenced identity in relation to GAT. Three time points were investigated: Time 1 (pre-test), Time 2 (post-test - immediately after receiving test results), and Time 3 (8-10 year follow-up). Test takers completed self-report instruments at all three time points and a semi-structured interview at Time 3; both quantitative and qualitative data were analyzed. At Time 1, 48% (n = 218) of the 455 GAT test takers spontaneously referenced identity. Further analysis of these data revealed that different dimensions and types of identity were referenced: identity exploration, including ethnic identity exploration; personal identity; essential identity; and group identity (in decreasing order of frequency). At Times 2 and 3, individuals started referencing identity commitment. At Time 3, test takers had high levels of ethnic identity commitment as measured with the Multigroup Ethnic Identity Measure-Revised (MEIM-R) and ethnic identity achievement as measured with the Ethnic Identity Scale (EIS), both of which are associated with well-being. Hence, in this data set, identity was a salient issue for about half the participants, with identity exploration being most frequently discussed prior to testing. Subsequent to GAT, some test takers referenced identity commitment. Thus, it appears that before testing potential test takers use GAT as a tool for exploring their identity and after GAT some individuals commit to a new or existing identity.

2384S

Psychosocial factors associated with the uptake of contralateral prophylactic mastectomy among *BRCA1/2* noncarriers with newly-diagnosed breast cancer. *J.G. Hamilton[1], M. Salerno[2], K. Amoroso[2], M. Sheehan[2], M. Harlan Fleischut[2], E. Glogowski[2], B. Siegel[2], A.G. Arnold[2], E.E. Salo-Mullen[2], J. Hay[1], K. Offit[2], M.E. Robson[2]*. 1) Behavioral Sciences Service, Department of Psychiatry & Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, NY; 2) Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY.

Women who are newly diagnosed with breast cancer can consider whether to have a contralateral prophylactic mastectomy (PM) to reduce their future risk of cancer in their unaffected breast. Pre-surgical *BRCA1/2* genetic testing may provide valuable risk information to guide this choice, given that *BRCA1/2* mutation carriers have a 27-37% ten-year risk of developing a new cancer in their unaffected breast. Despite *BRCA1/2* mutation noncarriers' much lower risk of contralateral breast cancer (which can be as low as 3-10%), some still choose to undergo a PM; the factors that motivate *BRCA1/2* noncarriers to select PM are not well understood and warrant investigation. Thus, as part of a larger study of the impact of pre-surgical *BRCA1/2* testing, we examined the frequency of uptake of contralateral PM among newly-diagnosed breast cancer patients who were identified as *BRCA1/2* mutation noncarriers. In addition, we examined the association between various sociodemographic and psychosocial factors and *BRCA1/2* noncarriers' subsequent decision to undergo PM. Self-report questionnaire data from 75 *BRCA1/2* noncarriers (mean age=44.1 years, range=29-59) were analyzed. A sizeable minority of the *BRCA1/2* noncarriers (29.3%) elected to undergo a PM after learning their mutation status (as compared to 88% of the 8 *BRCA1/2* carriers in the sample). Both bivariate and multivariable analyses indicated that perceiving that one's physician had recommended PM (OR=7.79, p=0.02), perceiving greater risk for breast cancer (OR=1.86, p=0.04), and perceiving greater pros of PM (OR=1.34, p=0.006) were all significantly associated with noncarriers' decision to undergo PM. However, factors including age, Ashkenazi Jewish ethnicity, breast cancer-related distress, perceived cons of PM, and decisional conflict regarding PM were not significantly related to the decision to undergo PM (all ps>0.05). Results demonstrate that although noncarriers' decision-making regarding PM was unrelated to sociodemographic and emotional factors, their cognitive perceptions of contralateral disease risk, surgical benefits, and physician recommendations were particularly important. Future studies should examine the content of patient-physician communication regarding PM and hereditary risk in greater detail, and explore how these conversations shape and interact with women's past experiences, emotions, and beliefs to influence their cancer risk management decision-making.

2385S

The Effect of Disclosing Coronary Heart Disease Genetic Risk on Shared-Decision Making. *T.M. Kruisselbrink[1,2], H. Jouni[1], R.A. Haddad[1], I.N. Isseh[1], V.M. Montori[3], I.J. Kullo[1]*. 1) Division of Cardiovascular Diseases, Department of Medicine, Mayo Clinic, Rochester, MN; 2) Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 3) Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Mayo Clinic, Rochester, MN.

Background: Whether disclosure of genetic risk influences shared-decision making for reducing coronary heart disease (CHD) risk is not known. We investigated the effect of disclosing genetic risk for CHD on shared decision making during a patient-physician encounter. **Methods:** The Myocardial Infarction Genes (MI-GENES) study randomized participants 40-65 y old, at intermediate CHD risk, and not on statins, to receive the 10-y risk of CHD based either on conventional risk factors alone (CRF) or CRF plus a 27 SNP genetic risk score (GRS) (CRF*GRS). CHD risk was disclosed by a genetic counselor during a 30 min scripted session. Each participant then met with a physician to engage in shared-decision making regarding the need for statin therapy. This encounter was facilitated by the Statin Choice decision aid (Weymiller et al, Arch Intern Med, 2007) modified to display both conventional and genetic CHD risk, with and without statin therapy. Afterwards, study participants were asked to complete a validated 11-question shared decision-making questionnaire (SDM-Q). Responses to each question were transformed to agree (1) and disagree (0) yielding a shared decision making score ranging between 0-11. **Results:** To date, 122 participants have completed the second study visit (mean age 59.1±5y, 48.3% male). CHD risk (risk =8.4±3.4%) estimated from CRF was communicated to 62 study participants (59±5.3y, 50% male) and 60 participants (59.1±4.7y, 53.3% male) received CHD risk estimated from CRF*GRS (risk = 9.2±7.4%). There was no difference in shared-decision making perception in the two groups (10.7 ±1 vs. 10.4±1.9, P=0.34). Specifically, there was no difference in participants' perception of adequate consultation time between the two groups: CRF vs. CRF*GRS [62 (50.8%) vs. 59 (48.3%), P=0.23, respectively]. Incorporating the GRS in the decision-making process did not affect participant's perception of sufficient discussion of different treatment options including use of statins vs. other medications or lifestyle modifications, selection of the treatment, advantages of each treatment, and feeling of inclusion in the treatment decision (P=NS for all). **Conclusion:** We demonstrate that disclosure of genetic risk for CHD does not affect participants' perception regarding shared-decision making. Participants in both study groups perceived that they had sufficient consultation time to discuss their CHD risk, different treatment options, and selection of treatment.

2386S

Adopting Genomes - Motivations of Adopted Persons when seeking Personal Genomic Services. *N.M. Baptista[1, 2], D.A. Carere[3], J.R. Duggan[1], T.A. Moreno[4], J.L. Mountain[5], S.A. Broadley[2], J.S. Roberts[6], R.C. Green[1], the PGen Study Group.* 1) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115; 2) School of Medicine, Gold Coast Campus, Griffith University, QLD 4222, Australia; 3) Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115; 4) Pathway Genomics Corporation, 4045 Sorrento Valley Blvd., San Diego, CA 92121; 5) 23andMe, Inc. 1390 Shorebird Way, Mountain View, CA 94043; 6) Department of Health Behavior and Health Education, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109.

Background: The use of personal genomic testing (PGT) by adopted individuals is largely unstudied. Using data from the Impact of Personal Genomics (PGen) Study, a longitudinal study of new PGT customers of 23andMe and Pathway Genomics, we conducted an exploratory analysis to characterize adoptees in the PGen population, and examined their motivations for seeking testing and their satisfaction with PGT results. **Methods:** PGen Study participants were surveyed before receiving results (baseline), and 2 weeks and 6 months after receipt of results. Of the 1,464 baseline survey respondents eligible for follow up, 1,042 completed the 6 month survey. At baseline, participants rated the importance, on a 3-point scale ("very important," "somewhat important," "not at all important"), of 12 possible motivations for seeking PGT. At 6 months, satisfaction with PGT results was measured by endorsement of 6 statements on a 5-point scale from "strongly disagree" to "strongly agree." Associations between adoption status and each motivation or satisfaction statement were analyzed using chi-square tests and logistic regression, adjusting for age, gender, race, education, company, and prior PGT. In all analyses, responses were dichotomized to "very important" and "somewhat/not at all important" for motivations, and to "strongly/somewhat agree" and "neither agree nor disagree/strongly/somewhat disagree" for satisfaction statements. **Results:** Of 1,042 6 month survey respondents, 57 (5%) were adoptees. In comparison to non-adoptees, adoptees were younger (42±13 years vs. 47±16 years, p=.01), and less highly educated (63% vs. 81% college degree or higher, p<.01). No significant difference was found between adoptees and non-adoptees' gender, race, income, or prior PGT. Adoptees were more strongly motivated to seek PGT because of their limited family health history (86% vs. 32% considered it "very important", p<.0001), and to learn their personal risk of specific diseases (OR=2.2, 95%CI 1.1-4.4, p=.03). Adoptees were no more satisfied than non-adoptees with the health information they gained, but were more likely to believe that seeking PGT was a wise decision (OR= 4.7, 95%CI 1.1-19.9, p=.03). **Conclusions:** Adoptees were younger, less highly educated, and more motivated than non-adoptees to seek genetic testing because of limited family health history and an interest in their disease risk. Belief that PGT was a wise decision may reflect a knowledge gap that adoptees wanted to fill.

2387S

Expert knowledge shapes decision-making for couples receiving uncertain prenatal chromosomal microarray testing results. *M.A. Rubel, K. Kellom, F.K. Barg, B.A. Bernhardt.* University of Pennsylvania, Philadelphia, PA.

Chromosomal microarrays (CMA) are increasingly utilized in prenatal settings to detect copy-number variants (CNVs) of probable or clinical significance not detectable by conventional cytogenetics. Up to 20% of microarray testing results in some kind of positive finding, including both pathogenic and uncertain results, which can complicate decision-making about pregnancy. When faced with ambiguous results, patients and their partners often turn to various forms of "expert knowledge" to mitigate uncertainty, including clinicians, internet resources, medical literature, and personal support networks. This study aims to investigate how couples utilize expert knowledge in the face of uncertain test results to navigate the decision-making process to terminate or continue a pregnancy. This qualitative study documents experiences of couples receiving uncertain prenatal CMA results from diagnosis centers in the US via phone interviews. Men and women separately discuss their experiences with CMA testing, understanding of and emotional response to the results, factors affecting decision making, and needs throughout the testing process. Interview transcripts were coded using NVivo10. Of the 18 male partner/female patient transcripts analyzed to date, most participants receiving uncertain results sought information from multiple sources. Female patients show a propensity to emphasize the impact of information in making their decision, cite various information sources, and seek expert advice from networks of family and friends. After receiving results from clinicians, women are predominantly information-seekers, with many reporting independently locating relevant clinical literature through internet searches. While most women found internet sources of expert knowledge helpful, some noted a fine-line between feeling informed by versus anxious from expert knowledge. In cases where uncertain results were not bolstered by additional understanding, expert knowledge could become "toxic knowledge," meaning that the information produced greater anxiety and emotional hardship. Male partners refer to the results and decisions as uncertain with greater frequency and tend to primarily rely on clinicians and their female partner to find and disseminate results, often seeing their role to be the primary emotional support. This study highlights both partner's varying utilization of expert knowledge as critical factors in the decision-making in a pregnancy with uncertain outcomes.

2388S

A genomic decision aid linked to the electronic health record to disclose coronary heart disease risk and enable shared decision-making. *K. Shameer[1], H. Jouni[1], R. Chaudhry[2], A.K. Dalenberg[1], V.M. Montori[3], I.J. Kullo[1].* 1) Division of Cardiovascular Diseases, Department of Medicine, Mayo Clinic, Rochester, MN; 2) Division of Primary Care Internal Medicine, Department of Medicine, Mayo Clinic, Rochester, MN; 3) Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Mayo Clinic, Rochester, MN.

Background: The myocardial infarction genes (MI-GENES) study is a randomized controlled trial of disclosing genetic risk of coronary heart disease (CHD) leveraging the electronic health record (EHR). We created a genomic decision aid integrated within the EHR to disclose CHD risk to study participants and enable shared decision-making regarding statin therapy to reduce risk. **Methods:** The Generic Disease Management System (GDMS), a web-based clinical decision support system used at the point of care and integrated within the EHR, was configured to estimate the 10-y CHD risk based on conventional risk factors extracted from the EHR. A link within GDMS provides access to the Statin Choice decision aid which has been proven effective for CHD risk communication (Weymiller et al, Arch Intern Med, 2007). This tool aids with initiation of statin therapy depicting not only the 10-y risk for CHD but also the potential benefit of using statins and the related costs/side effects. A feature was added to the tool enabling the genetic counselor, physician, and patient to visualize the change in 10-y CHD risk after incorporating the genetic risk score (GRS) based on 27 genetic susceptibility loci. CHD risk was disclosed by a genetic counselor in a 30 min scripted session using a pictogram that displays how many among 100 patients with the same risk profile will suffer a myocardial infarction in the ensuing 10 years. Afterwards, the participant meets with a physician for shared decision-making regarding the need for statin therapy. The tool is also equipped with a report generating function and a 'frequently asked questions' page for additional information about GRS. **Results:** The tool is being used for disclosure of genetic risk of CHD in the MI-GENES study. So far, we have randomized 212 participants (47% males, mean age 58.8±5.1y) to receive the 10-y probability of CHD estimated by conventional risk factors vs that estimated from conventional risk factors plus a GRS. Patients are surveyed for satisfaction with genetic counseling and physician encounters, understanding of genetic risk, and shared decision-making. **Conclusion:** In summary, we have developed and successfully implemented a genomic decision aid integrated within the EHR to disclose CHD risk and enable shared decision-making in the clinical setting. The results of the study will facilitate development of best practices for incorporating probabilistic genetic risk for common diseases in the EHR.

2389S

The predictive power of breast cancer family history in the clinic. *H.C. Cox, E. Rosenthal, R. Wenstrup, B.B. Roa, K.R. Bowles.* Myriad Genetic Laboratories, Inc., Salt Lake City, WA.

Background: Women with an estimated >20% lifetime risk of breast cancer are candidates for more aggressive clinical management, including screening at younger ages, at more frequent intervals, and with more sensitive technologies, i.e. breast MRI. Family history is a key component of most models for estimating breast cancer risk, but family history analysis may be insufficient to identify at-risk individuals carrying moderately penetrant pathogenic mutations due to limited sibship sizes in contemporary families. We utilized pedigree simulation to estimate the probability that a female proband, who is a carrier of a pathogenic mutation conveying a moderate increase is breast cancer risk, will be identified as having at least a 20% lifetime breast cancer risk as determined by the Claus model. Methods: The SIMLA and SLINK pedigree programs were used to simulate 200 three-generation pedigrees each for 2, 3, 4, or 5-member sibships. The proband was assumed to be a 40-year old female carrying one copy of an autosomal dominant pathogenic mutation conveying a ~25% risk of breast cancer to age 80. Simulated pedigrees were one-sided and limited to either the maternal or paternal side segregating the disease allele. Phenotypes were simulated according to age-dependent liability classes modeled from the Surveillance, Epidemiology, and Ends Results (SEER) breast cancer incidence data. Resulting pedigrees were assessed by the Claus model to determine the proband's eligibility for modified medical management. Results and Discussion: Analysis of simulated pedigrees indicate that <9% of female probands, carrying a pathogenic mutation conveying a ~25% risk of breast cancer, would receive modified clinical risk management based only on Claus model risk assessment. Thus, genetic testing may be critical for identifying individuals carrying pathogenic mutations in moderate penetrance breast cancer susceptibility genes who would benefit from increased surveillance, as outlined in current professional society guidelines. Although population screening for moderately penetrant gene mutations may not yet be economically feasible, clinicians may wish to consider broader pan-cancer testing using a panel composed of both high and moderate penetrance genes when screening the patient for other cancer risks, such as colon, endometrial, and ovarian cancers.

2390S

Colorectal cancer screening for people with a family history: should recommendations vary by age? *I. Lansdorp-Vogelaar[1], S.K. Naber[1], K.M. Kuntz[2], N.B. Henrikson[3], M.S. Williams[4,5], N. Calonge[6], K.A.B. Goddard[7], D.T. Zallen[8], E.P. Whitlock[7], T.G. Ganiats[9], C.M. Rutter[3], E.M. Webber[7], A.C.J.W. Janssens[10], M. van Ballegooijen[1], A.G. Zauber[11].* 1) Department of Public Health, Erasmus Medical Center Rotterdam, Rotterdam, Netherlands; 2) Department of Health Policy & Management, University of Minnesota, Minneapolis, MN, United States; 3) Group Health Research Institute, Seattle, WA, United States; 4) Intermountain Healthcare, Salt Lake City, UT, United States; 5) Genomic Medicine Institute, Geisinger Health System, Danville, PA, United States; 6) The Colorado Trust, Denver, CO, United States; 7) Center for Health Research, Kaiser Permanente Northwest, Portland, OR, United States; 8) Department of Science and Technology in Society, VirginiaTech, Blacksburg, VA, United States; 9) Department of Family and Preventive Medicine, University of California, San Diego, CA, United States; 10) Department of Epidemiology, Emory University, Atlanta, GA, United States; 11) Memorial Sloan-Kettering Cancer Center, New York, NY, United States.

Objective. People with first degree relatives (FDRs) diagnosed with colorectal cancer (CRC) are at an increased risk for developing CRC. It is therefore recommended to screen them more frequently than the general population. Although the relative risk for colorectal cancer decreases with the age of the person at risk, none of the screening recommendations take this age into account. The aim of this study is to determine the potential benefits of age-specific screening guidelines for people with a family history of CRC. **Methods.** For the age groups 30-44, 45-49, 50-54, 55-59, 60-64, 65-69 and 70+, we based the relative risk of having 1, 2, 3 and 4 or more FDRs diagnosed with CRC as compared to the general population on a literature review. These relative risks were incorporated into the MISCAN model to estimate costs and effects of colonoscopy screening strategies varying in age range and interval for people with 1, 2, 3, or ≥4 FDRs diagnosed with CRC. For each age and level of FDR, we determined the most cost-effective screening schedule with an incremental cost-effectiveness ratio (ICER) comparable to that of screening in the general population. Because past screening could influence the cost-effectiveness of future screening, we assessed the influence of screening history on the optimal screening interval in a sensitivity analysis. **Results.** While effectiveness of colonoscopy screening increased with the number of FDRs people have, total costs of care decreased. Colonoscopy screening was even cost-saving in people with a family history. Consequently, screening should start earlier at age 40 (35 for those with ≥4 FDRs) in people with a family history. The optimal screening interval ranged from 5 years for people with 1 FDR, to 1-3 years for people having 4 FDRs or more. Despite the decreasing relative risk with age, the optimal screening interval was similar for all age groups. **Conclusions.** Although the relative risk of people with FDRs diagnosed with CRC decreases with age, optimal screening intervals do not vary with age but do vary with number of affected FDRs. People with a family history should be offered screening from age 40 (35 if ≥4 FDRs) with an interval of 5 years in case of 1 FDR, 3 years in case of 2-3 FDRs and 2 years in case of ≥4 FDRs.

2391S

Improving pedigree capture: Development and validation of an interview-optimized iPad app to eliminate the need for paper. *J.M. Miller[1], M.J Italia[1], E.T Dechene[2], A. Wilkens[2], C.E Gaynor[1], N.B Spinner[3,4], I.D Krantz[5], P.S White[6,7].* 1) Center for Biomedical Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 4) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 5) Department of Pediatrics, University of Pennsylvania, Philadelphia, PA; 6) Department of Pediatrics, Cincinnati Children's Hospital, Cincinnati, OH; 7) Department of Biomedical Informatics, University of Cincinnati, Cincinnati, OH.

Large gene panel, exome and whole genome sequencing-based clinical tests produce data at a scale that is time intensive and difficult to interpret. Trio and inheritance analyses are increasingly utilized as a means for reducing the volume of variants that must be reviewed. In addition to the sequence data, these analysis methods greatly benefit from well-documented pedigree and family history data. Electronic capture of pedigree and family member data provides a way to use this data for subsequent analysis, review, and use. However, while a number of electronic solutions have been developed, these systems typically do not perform well in a busy clinical setting. Consequently, these systems have not generally supplanted the use of paper-based manual abstraction, especially for pedigree representation, resulting in decreased efficiency and increased risk of transcription error. We hypothesized that a multi-touch iPad application could compete with paper on clinical usability, producing both suitable clinical documentation and structured data useful for inheritance analysis in the laboratory. As part of the NHGRI's Clinical Sequencing and Exploratory Research (CSER) Program, we developed Proband, an iPad app for capturing pedigrees, available in the Apple App Store. The app enables the user to create complex family pedigrees by implementing nomenclature outlined by the Pedigree Standardization Work Group (PSWG). The primary method of user input is a series of intuitive touch gestures. To streamline data entry, the app relies on the user's context to make the appropriate functionality available at the right time. Proband runs on a local data store and generates pedigree results as PDF or XML. We performed formal usability testing with ten genetic counselors. Each counselor was asked to recreate a pedigree after viewing a training video. The counselors then completed the System Usability Scale (SUS), a standard ten-question survey used to measure user satisfaction. The mean score was 90, well above the average SUS score of approximately 70. Qualitative responses indicated a functional convenience equivalent to, and for some tasks, exceeding that of paper. Initial user feedback at our hospital as well as other institutions has been very positive. In one instance, a counselor in a busy clinic completely replaced paper with Proband, creating over 100 pedigrees to date. We continue to evaluate and improve Proband for use in clinical diagnostic environments.

2392S

Non-specific psychological distress in women undergoing noninvasive prenatal testing because of advanced maternal age. *N. Suzumori[1], K. Kumagai[1], S. Goto[1], T. Ebara[2], Y. Yamada[2], M. Kamijima[2], M. Sugiura-Ogasawara[1].* 1) Division of Clinical and Molecular Genetics, Department of Obstetrics and Gynecology, Nagoya City University, Nagoya, Japan; 2) Department of Occupational and Environmental Health, Nagoya City University, Nagoya, Japan.

Objective The objective of our study was to describe our clinical experience in providing noninvasive prenatal testing (NIPT) for fetal aneuploidy to pregnant women, highlighting the degree of non-specific psychological distress. Methods Data were collected from Japanese women who were offered and underwent NIPT after genetic counseling and control pregnant women who did not undergo NIPT as part of the Japan Environment and Children's Study (JECS) Control A planning. The degree of mental distress was assessed using the Kessler 6 (K6) screening scale with a score ≥10 indicating depression or anxiety disorder. Results Among the 505 women who underwent NIPT because of advanced maternal age, 9.1 %, had a K6 score ≥10. Compared with matched Controls (n=1,010) adjusted for maternal age and gestational age, the NIPT group showed a trend toward higher K6 scores (OR 1.44, 95% CI 0.97-2.13, P = 0.07). Higher K6 scores were associated with women whose husbands were the primary decision makers during NIPT counseling (P = 0.06). Conclusions Women electing NIPT tend to have higher scores of depression/anxiety, and those with higher depression scores tended to defer the decision to their husbands.

2393S

DIAGNOSTIC CHALLENGES AND BEHAVIOR PROBLEMS IN RWANDAN PATIENTS WITH DISORDERS OF SEX DEVELOPMENT: ETHICAL ISSUES IN AFRICAN CONTEXT. *L. Mutesa[1], J. Hitayezu[1], A. Uwineza[1,2], S. Murorunkwere[1], J. Ndinkabandi[1], A. Ndatinya[3], F. Rutagarama[3], O.R. Karangwa[3], A. Gasana[4], J.H. Caberg[2].* 1) Medical Genetics, National University of Rwanda, KIGALI, Kigali, Rwanda; 2) Center for Human Genetics, University of Liege, Belgium; 3) Department of Pediatrics, Rwanda Military Hospital, Rwanda; 4) Department of Urology, Rwanda Military Hospital, Rwanda.

Background Disorders of sex development (DSD) comprise a variety of congenital diseases with anomalies of the sex chromosome, the gonads, the reproductive ducts and genitalia. The most common DSDs result from disruption of androgen levels and activity that affect later embryonal development, such as congenital adrenal hyperplasia and androgen insensitivity syndrome (AIS). DSDs are always challenging and very difficult to manage. Socio-economic and cultural aspects have a great impact on decision making regarding the management of these conditions. The situation is more complicated in resource-poor settings like in Africa, where access to education and medical care is limited in both quantity and quality of infrastructure, diagnostic tools and medical professionals. In addition, traditional values and beliefs are also very strong in various cultures and sexual issues are taboo subjects. Methods The present study is a 5-year prospective descriptive cohort of patients with suspicion of DSDs referred to our genetic clinic between January 2009 and January 2014 for genetic investigations and counseling. All patients underwent abdominal ultrasound or MRI and hormonal analysis before genetic testing including karyotype and molecular tests. Results In total, 49 patients aged between 1 and 39 years were clinicaly and geneticaly diagnosed with DSDs associated with behavior problems in most of cases. The majority were diagnosed either with sexual ambiguity and hypospadias, or micropenis, or primary amenorrhea, or poor development of secondary sexual characteristics, or primary infertility. In most of female cases, the ultrasound and MRI revealed absence of uterus and ovaries. The FSH, LH or testosterone hormons revealed major abnormal values in more than 60% of patients. The AIS and Rokitansky syndrome were observed in the majority of these patients. The choice of gender identity after karyotyping raised several psychological and ethical issues in the majority of adults patients. The outcome of surgery was successful for the social integration of some of these patients. Conclusion The present study showed that patients with DSDs have major behavior problems in African context. They suffer from extreme anxiety and very high psychological behavior disorders related to their gender identity. Our data suggest that gender assignment has to be avoided before expert and multidisciplinary evaluation especially in young patients.

2394S

Making sense of diagnostic uncertainty after newborn screening for cystic fibrosis. *C.J. Barg[1], R.Z. Hayeems[1,2], F.A. Miller[1], Y. Bombard[1,3], P. Durie[4,5], J.C. Carroll[6], P. Chakraborty[7,8], B.K. Potter[9], K. Tam[10], L. Taylor[11], E. Kerr[12], C. Davies[7], J. Milburn[7], K. Keenan[4], A. Price[13], F. Ratjen[4,5,11], A. Guttmann[1,4,5,14].* 1) Institute of Health Policy, Management & Evaluation, University of Toronto, Toronto, ON, Canada; 2) Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada; 3) Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada; 4) Division of Paediatric Medicine, Department of Paediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 5) Department of Pediatrics, Faculty of Medicine, University of Toronto, ON, Canada; 6) Department of Family & Community Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada; 7) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 8) Department of Pediatrics, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; 9) Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, ON, Canada; 10) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 11) Division of Respiratory Medicine, Department of Pediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 12) Department of Psychology, The Hospital for Sick Children, Toronto, ON, Canada; 13) Department of Pediatrics, London Health Sciences Centre, London, ON, Canada; 14) The Institute for Clinical Evaluative Sciences, Toronto, ON, Canada.

Objectives: Diagnostic uncertainty related to cystic fibrosis (CF) has long been challenging to understand and manage. Newborn screening (NBS) for CF extends these challenges to the early days of life. We explored the parental experience of diagnostic uncertainty arising from NBS. **Methods**: Drawing from a mixed methods prospective cohort study of screen positive and screen negative control infants, we report on qualitative interviews with parents of children who received inconclusive results for CF after NBS and follow-up testing. Through the tertiary care centre that manages the majority of screen positive care in Ontario, we recruited parents of children identified since the inception of CF NBS in Ontario (2008). We used qualitative description to analyze transcript data. **Results**: We conducted 19 interviews with 14 parents of infants with uncertain diagnoses, ranging from 3 months - 4 years in age. Five parents completed interviews at two time-points, separated by one year. We learned that parents gain support through research involvement, but struggle to understand the meaning of an uncertain diagnosis in the face of an apparently healthy newborn, worry about their infant's health-related vulnerability, and fear labeling and over-medicalization from continued medical surveillance. Time appears to mitigate concerns. **Conclusion**: The experience of diagnostic uncertainty is deeply challenging for some families, particularly in the early newborn period. These results should inform decisions by NBS programs and clinical teams about protocols for testing, clinical follow-up, and support for families of those who screen positive, but warrant triangulation with survey data and longitudinal follow-up.

2395S

Reconciling the downsides of screening: Mothers' experiences with false positive CF NBS results. *F.A. Miller[1], R.Z. Hayeems[1,2], Y. Bombard[1,3], S.J. Patton[1], C.J. Barg[1], P. Durie[4,5], J.C. Carroll[6], P. Chakraborty[7,8], B.K. Potter[9], K. Tam[10], L. Taylor[11], E. Kerr[12], C. Davies[7], J. Milburn[7], F. Ratjen[4,5,11], A. Guttmann[1,4,5,13].* 1) Institute of Health Policy, Management & Evolution, University of Toronto, Toronto, ON, Canada; 2) Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada; 3) Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, ON, Canada; 4) Division of Paediatric Medicine, Department of Pediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 5) Department of Paediatrics, Faculty of Medicine, University of Toronto, Toronto, ON, Canada; 6) Department of Family and Community Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada; 7) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 8) Department of Pediatrics, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; 9) Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; 10) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 11) Division of Respiratory Medicine, Department of Pediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 12) Department of Psychology, The Hospital for Sick Children, Toronto, ON, Canada; 13) Institute for Clinical Evaluative Sciences, Toronto, ON, Canada.

Objectives: One of the downsides of population screening is the generation of false positive screening results. In newborn screening (NBS) for cystic fibrosis (CF), however, many of these results identify carriers of one mutation, which is often seen as a benefit. We sought to understand how mothers interpret this false positive experience. **Methods**: As part of a longitudinal mixed-methods cohort study of NBS for CF, we surveyed mothers of infants with false positive results within 3 months of birth and one year later. Participants who agreed were also interviewed. We present descriptive and qualitative thematic analyses. **Results**: We received completed surveys from 134 of 245 mothers (55%) at time-1 (T1) and 96 of 214 (45%) at time-2 (T2). We report on 31 T1 interviews. Quantitatively, majorities agreed to feeling informed (T1:95%, T2:86%) and relieved (T1:90%; T2:83%) about their baby's result; and minorities agreed that learning about a false positive (T1:32%; T2:40%) or carrier result (T1:4%; T2:3%) was a disadvantage of NBS. Qualitatively, mothers identified their NBS experience as highly burdensome. Tolerance for this burden was expressed in 4 ways - by downplaying the burden in retrospect, appreciating the system of care, identifying benefit for self (reproductive and otherwise), and identifying benefit for others. **Conclusion**: These findings suggest significant tolerance for the downsides of NBS. They also identify ways in which burdens are made tolerable. Notably, reproductive benefit is only one way; systems of care that help to mitigate burdens, and an ethic of care for others, are also important.

2396S

Numeracy, Genetic Knowledge, and Perceived Risk for Coronary Heart Disease in the MI-GENES Study. *H. Jouni[1], I.N. Isseh[1], R.A. Haddad[1], K.D. Christensen[2], R.R. Sharp[3], R.C. Green[2], I.J. Kullo[1].* 1) Division of Cardiovascular Diseases, Department of Medicine, Mayo Clinic, Rochester, MN; 2) Division of Genetics, Department of Medicine, Brigham & Women's Hospital & Harvard Medical School, Boston, MA; 3) Biomedical Ethics Program, Mayo Clinic, Rochester, MN.

Background: Genetic knowledge, numeracy, and risk perception likely influence how patients interpret and respond to genetic test results. We assessed these characteristics in participants of the Myocardial Infarction Genes (MI-GENES) study that is exploring integration of genomic results into the electronic health records. **Methods:** Participants were residents of Olmsted County, aged 40-65 y, at intermediate risk for coronary heart disease (CHD), and not on statins. They were randomized to receive the 10-y risk of CHD based either on conventional risk factors alone or conventional risk factors plus a 27 SNP genetic risk score. We assessed numeracy, genetic knowledge, and self-perceived risk for CHD at baseline before disclosing CHD risk. Numeracy and genetic knowledge were assessed using previously validated surveys. We adapted the Health Information National Trends Survey (HINTS) to assess CHD risk perception. **Results:** The mean age of 212 participants was 58.8±5y, 47% were male, all were non-Hispanic white, 21% had completed high school and 62% had graduated from college or graduate school. At baseline, participants scored highly on numeracy; 42.9% answered all of the 8 questions correctly and 31.1% answered 7 questions correctly. For genetic knowledge, 22% answered ≥14 of 16 questions correctly while most participants (62%) answered between 9-13 questions correctly. The mean perceived 10-y risk was significantly higher than the estimated 10-y risk for CHD (25.6%±20.3% vs 8.5%±4.1%, respectively, P<0.01). Although patients overestimated their risk for CHD, the majority (80%) described their risk as the same or lower than that of healthy peers. Participants with family history of CHD were more likely to perceive increased genetic susceptibility to CHD than participants without such history (63% vs. 26%, respectively, P<0.01). Higher numeracy was associated with lower self-perceived risk for CHD (ρ -0.2, P=0.004). Genetic knowledge scores were not associated with perceived risk regardless of family history (P=NS). **Conclusion:** Early adopters of genetic testing for CHD in this study were facile with numbers and knowledgeable about genetics, but greatly overestimated their risk for CHD. Interestingly, numeracy but not genetic knowledge, was inversely correlated with perceived CHD risk. Our results indicate that education and counseling protocols for genetic testing should also address numeracy misperceptions about disease risk.

2397S

Genomic Sequencing in the Infant Population: Exploring Parental Motivations, Expectations and Utilization of Sequencing Results in the *Tell Me More* Study. *F. Facio[1], B. Solomon[1], S. Haga[2], E. Klein[1], K. Huddleston[1], A. Khromykh[1], S. Hull[3], B. Berkman[3], E. Sutton[4], K. Hurley[5], J. Evans[6], J. Vockley[1], J. Niederhuber[1].* 1) Inova Translational Medicine Institute, Falls Church, VA; 2) Institute for Genome Sciences & Policy, Duke University, Durham, NC; 3) Bioethics Core, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) Center for Bioethics and Social Sciences in Medicine, University of Michigan, Ann Arbor, MI; 5) Teachers College, Columbia University, New York, NY; 6) Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC.

Advances in genome sequencing (GS) technologies have increased their use in research and clinical practice. However, many barriers exist to translating GS to routine clinical care, including a lack of data documenting motivations, expectations, and behavioral outcomes as well as parental perspectives and behaviors. We describe findings from a large, prospective trio-based whole GS study within a multi-hospital health system (~20,000 annual deliveries, half of which are at the main study hospital). Approximately 10% of these deliveries enroll in the longitudinal GS study, largely limited by full trio availability. We have enrolled ~1500 trios to date (we accrue ~80-100 trios/month). Reasons for participation include potential benefits related to finding medically relevant information, altruism, lack of cost, enthusiasm regarding science and genomics, and the fact that the study is conducted directly through the same health care system. Reasons for study decline include unwillingness to undergo phlebotomy (especially fathers due to the need for an additional blood draw), privacy concerns and questions related to government access to genetic/genomic and other health information, concerns about potential effects on insurance coverage, and stigma associated with research, especially related to genetics/genomics. Focus groups of study participants have shown an overwhelming desire to know both specific and general genetic/genomic information. As an initial step, we have started returning CLIA-validated pharmacogenetic results within the focus groups to determine the impact of return of these results. To explore the sociobehavorial aspects of GS further, we have also designed the *Tell Me More* study, which aims to investigate parental motivations for enrolling or not enrolling their infants, expectations about GS results and the use and impact of such information. This nested study is recruiting a total of 40 participants - 20 parents who opted to pursue GS for their newborn (intervention group) and 20 who declined (control group), and uses phenomenological methods to explore parental experiences and decision-making processes. Parents in both groups complete interviews post informed consent and parents in the intervention group complete interviews post disclosure of results. Ultimately, this study will generate results that can be used as building blocks for larger projects and inform best practices in clinical applications of GS in the infant population.

2398S

Supporting the International Rare Diseases Research Consortium: achievements and challenges. *P. Lasko[1,2], B. Cagniard[3], S. Höhn[3], L. Lau[3], S. Peixoto[3], S. Aymé[3], IRDiRC Consortium.* 1) International Rare Diseases Research Consortium, Montreal, Quebec, Canada; 2) Institute of Genetics, Canadian Institutes of Health Research, McGill University, Montreal, Canada; 3) SUPPORT-IRDiRC, INSERM US 14, Paris, France.

The International Rare Diseases Research Consortium (IRDiRC) brings together members that have agreed to common goals and principles and to work in a coordinated and collaborative manner within a multinational consortium to advance research in this critical area. Its over 35 members are public and private research funding organizations and companies that have each dedicated over 10 million US$ to research into rare diseases. Present members are from Europe, North America, Asia and Australia. IRDiRC's two main objectives are to deliver 200 new therapies for rare diseases and the means to diagnose most rare diseases by the year 2020. To accomplish this, IRDiRC engages in a range of initiatives. The first one is to support financially strategic research projects. The funding agencies remain independent in their calls for proposal but discuss jointly their strategy and agree on main areas to be supported. The second instrument is to identify and highlight strategic infrastructures which could be used by the researchers to speed up their projects. The third is to identify and promote standards to be used to make data interoperable and as easy to access and share as possible. Many working groups have been put in place to advice the consortium in these and other areas, and their suggestions are reviewed by three scientific committees who directly advise the executive board. A label 'Recommended by IRDiRC' has recently been created to identify preferred tools, protocols, and guidelines that directly contribute to IRDiRC goals. Consortium achievements are posted in a timely manner on its website (www.irdirc.org), enabling all the activities of its many committees and working groups to be as transparent as possible. As part of its dissemination and communication plan, IRDiRC publishes feature articles on strategic research projects and infrastructures and offers on its website a space for its members to communicate. It organized an International conference in Dublin in 2013 which was the real public launch of this initiative. The next IRDiRC conference will be focused on fostering international collaboration with Asia and will be held in Shenzhen (China) on 7-9 November 2014 thanks largely to support from BGI.

2399S

Assessment of factors that should be addressed in prenatal counseling for non-invasive prenatal test. *L. Wang[1], Y. Gu[1], Q. Meng[1], X. Tang[1], H. Wang[1], S. Yang[1], H. Mao[1], F. Liu[1], J. Zhang[1], H. Wu[1], Q. Shi[1], N. Zhong[2].* 1) Department of Prenatal Diagnosis, Lianyungang Maternal and Child Health Hospital, Lianyungang, China; 2) Department of Human Genetics, New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY.

Objectives: To investigate the potential factors affecting the women's attitude toward using a non-invasive prenatal test (NIPT) as well as knowledge of prenatal diagnosis in pregnant women, and the clinical efficiency of identifying fetal chromosomal aneuploidies in pregnant women using an NIPT. Method: A survey of the potential factors (age, education, and residential area) affecting pregnant women's choice of prenatal diagnosis and their understanding of prenatal diagnosis that was conducted in our hospital. Moreover, 917 women with high-risk pregnancies were invited to participate in an NIPT trial. Complete karyotyping of the amniotic fluid was performed to validate each abnormal case identified in NIPT. Results: Women older than 35, rather than 18 to 35 years, were statistically significantly inclined to choose NIPT as their prenatal diagnosis procedure ($P < 0.05$). The percentages of women knowing of the purpose, advantage, limitation and costs of NIPT were less than that of amniocentesis. Moreover, an average of 24.46% women understood the indications for NITP. The NIPT trial with validation by full karyotyping of the amniotic fluid identified 3.60% high-risk pregnancies with fetal aneuploidies, including 2.73% trisomy 21 (Tri21), 0.33% trisomy 18 (Tri18), 0.44%Turner syndrome (45, X), and 0.11% Klinefelter's syndrome (47, XXY). In addition, in the NIPT tiral, one woman with maternal mosaicism (45, X and 46, XX) and one case of false-positive Tri18 caused by confined placental mosaicism (CPM) were also identified after karyotype validation. Conclusions: Pregnant women's age was the primary factor affecting women's choice for NIPT, and more education on NIPT is necessary for promoting clinical application of NIPT. Meanwhile, discordant positive results found with NIPT can be caused by confined placental mosaicism and maternal mosaicism, which underlines the need to confirm all positive NIPT results via conventional invasive testing and karyotyping at NIPT's current stage of development.

2400S

Genetic Counselling for Psychiatric Disorders: Exploring the Psychiatric Health Professionals' Perspective in the United Kingdom. *S.E. Jenkins[1,2], M. Arribas-Ayllon[2].* 1) Wessex Clinical Genetics Service, Southampton, Hampshire, United Kingdom; 2) Institute of Medical Genetics, Cardiff University, Cardiff, Wales, United Kingdom.

Background Historically genetic counselling is not offered for psychiatric disorders in the United Kingdom through regional clinical genetics departments provided by the NHS. The purpose of genetic counselling through these specialist centres is to provide patients and their families with information about the nature and consequences of a disorder, in addition to the probability of developing symptoms of the disorder and passing it on to future generations. The options available to the patient and their families in management and treatment are discussed, and in some cases genetic testing can then be explored. Recent genomic advances identifying a greater genetic contribution to mental illness is anticipated to increase the demand for psychiatric genetic counselling and is expected to cause significant changes in the management and treatment of these disorders. **Methodology** The study employed both quantitative and qualitative methods of research to explore the attitudes and experiences of psychiatric health professionals with respect to identifying a potential for genetic counselling services within clinical psychiatry in the UK. **Results** Data analysis revealed the following emerging themes: -Demand for genetic counselling; -Responsibility for genetic counselling provision; and -Barriers for the service. Evaluation of the data led to the conclusions that although demand for psychiatric genetic counselling has not been voiced in the UK at present, psychiatric health professionals believe that such a service would be a useful and desirable. Genomic advances which identify susceptibility loci for psychiatric disorders may have significant implications for genetic counselling in clinical psychiatry if these discoveries lead to genetic testing. Psychiatric health professionals describe clinical genetics as a skilled profession capable of combining complex risk communication with much needed psychosocial support. However, the possibility of such a service is confronted with a range of barriers and challenges including, but not limited to, the complexities of uncertainty in psychiatric diagnoses, patient engagement and ethical concerns regarding limited capacity and increased suicidality.

2401S

Autism Spectrum Disorder in Taiwan: Parents' report. *J. Ye[1], T.Y Huang[2], S.X Zhao[1], L. Xu[1], E. Jung[3], Y.Y Wu[4], D. Tsai[5], D. Talwar[1], L.S. Chen[1].* 1) Department of Health and Kinesiology, Texas AM University, College Station, TX; 2) Department of Special Education, National HsinChu University of Education, HsinChu, Taiwan; 3) Department of Educational Psychology, Texas AM University, College Station, TX, USA; 4) Department of Psychiatry, Chang Gung Memorial Hospital- Linkou Medical Center, Chang Gung University College of Medicine, Tao-Yuan, Taiwan; 5) Department of Social Medicine, National Taiwan University College of Medicine, Taipei, Taiwan.

Background: In the United States, one in 50 school-age children are affected by Autism Spectrum Disorder (ASD). As a global challenge, limited research has been conducted on ASD outside the U.S. This study reports preliminary findings and characteristics of children with ASD at pre-schools and elementary schools in Taiwan. Methods: A total of 443 parents of children with ASD, recruited from pre-schools and elementary schools of Hsinchu city and county, Taoyuan and Miaoli counties, completed the paper-and-pencil survey. Results: The average maternal and paternal ages at the time of ASD child's birth were 29.47 years (SD=4.6) and 33.28 years (SD= 5.33), respectively. The majority of the affected children were predominantly boys (6.94:1). The age of confirmatory diagnosis for ASD was 4 years (SD=2.2). About one-fifth (18.5%) of the parents were unaware about the classification of ASD that affected their child. Most of the parents (73.5%) reported that they received either very low or low support from the society for their children. Conclusions: Our study indicated the average age when children were diagnosed with ASD was beyond the recommended age of two. Some parents were ignorant about the type of ASD associated with their child's ASD and most parents reported low support from the society. Relevant agencies in Taiwan should address this urgent issue to facilitate better care and support for the affected family.

2402S

Carrier screening of recessive disorders in vitro fertilization couples and gamete donors and recipients by targeted next generation sequencing. *A. Abulí[1], B. Rodríguez-Santiago[2], M. Garcia-Aragonés[2], M. Boada[3], B. Coroleu[3], E. Clua[3], M. Parriego[3], A. Veiga[3], M. Del Campo[1], P. Barri[3], L. Pérez-Jurado[4], L. Armengol[2], X. Estivill[1].* 1) Genomics and Personalized Medicine Unit, Dexeus Woman's Health, Quirón-Dexeus University Hospital, Barcelona, Spain; 2) Research and Development Department, qGenomics, Barcelona, Spaing; 3) Reproduction Medicine Service, Obstetrics, Gynecology and Reproduction Department, Quirón-Dexeus University Hospital, Barcelona, Spain; 4) Genetics Group, Pompeu Fabra University, Barcelona, Spain.

Screening tests for carriers of recessive disorders interrogating specific mutations in selected genes using SNP-genotyping approaches have been developed. By contrast, next generation sequencing (NGS) technologies enable deep gene sequencing in a cost-efficient manner. We developed an NGS-based approach targeting 215 genes causing prevalent and severe recessive diseases for testing in vitro fertilization (IVF) couples and donors/recipients in donation programs. We measured overall sensitivity and specificity of different known genetic mutations present in a blinded training set. Bioinformatics analysis is keystone in the process, as it combines several algorithms optimized for the identification and annotation of different types of mutations (point mutations, indels, copy-number and rearrangements). Our experience with NGS gene panels showed extremely high sensitivities (>99%) for all kinds of mutations. For carrier test validation, we obtained DNA from 57 unrelated individuals: 39 patients and 18 previously genotyped controls. The validation set was composed of 49 different known mutations (67 in the patients), including 29 SNVs, 13 indels and 25 CNVs causing different diseases: cystic fibrosis, phenylketonuria, spinal muscular atrophy, hypothyroidism, thalassemia, factor V deficiency and Duchenne muscular dystrophy. All but one (48/49) different mutations were correctly scored in the blinded study and only one deletion-type mutation remained undetected. This information allowed us to finely tune the algorithm to reach maximum sensitivity. All single nucleotide changes were validated and no known recessive mutations were called in the control samples. To date, we evaluated the spectrum of mutations found in 113 patients referred for carrier testing by IVF clinics. We have identified 62 carriers (54%) and a total of 56 different mutations. We detected multiple carriers for common mutations in the European population but we also identified rare mutations and novel deleterious variants that are predicted to be pathogenic (nonsense, frameshift mutations and mutations at splice site positions), as well as collected evidences for discrepancies between disease prevalence and population frequency of apparently pathogenic mutations. A significant proportion of the identified mutations, specially the novel mutations, would have not been detected by traditional methods. NGS has higher detection rates resulting in lower residual risks.

2403S

Inherited cancer predisposition in children: challenging issues faced by a genetic clinic in a pediatric oncology hospital. *F.T. de Lima[1,2], V.F.A. Meloni[3], C.R.P.D. Macedo[2], N.S. da Silva[2], A.M. Cappellano[2], R.V Gouveia[2], E.M.M. Caran[2].* 1) Gynecology and Obstetrics Department, UNIFESP-EPM, São Paulo , São Paulo, Brazil; 2) Pediatric Oncology Department - GRAACC (Support Group for Children and Adolescents with Cancer), UNIFESP-EPM, São Paulo, São Paulo, Brazil; 3) Genetics and Morphology Department, UNIFESP-EPM, São Paulo, São Paulo, Brazil.

INTRODUCTION: Genetic predisposition in childhood cancer is variable, and the access of children with neoplasms to genetic services skilled in hereditary cancer and dysmorphology is limited, even in well-developed countries. OBJECTIVES: We described the structure, the patients' clinical characterization and challenges faced by a new genetic clinic, in a public pediatric oncology hospital in Brazil. METHODS: We reviewed the initial proposal for the clinic's structure, the protocols and charts of all patients seen from August 2012 to December 2013, as well as the patient's database. RESULTS: The initial structure was centered in patients' care, with standard initial clinical evaluation protocol, and follow-up protocols for specific conditions, e.g., neurofibromatosis (NF) and tuberous sclerosis (TS). Appointments occur once a week. The main reason for referral was a diagnosis with known genetic implications, followed by association of cancer and malformation. Other reasons included cancer family history and concerns about recurrence in another child. Almost half of the 172 patients (152 families) seen had NF (46/26,7%) or retinoblastoma (41/23,8%); 37 (21,5%) patients had other diagnosis. The exact diagnosis was not obtained in 48 patients (27,9%), classified as suspected malformation syndromes associated to cancer (25), or suspected familial/hereditary cancer (23). Most were seen just once (113/65,7%). Issues faced included an appreciable number of patients needing evaluation, lack of knowledge about the goals of genetic counseling and about family history. To overcome some, a folder with orientations was created for easy referral. The need for a stronger and closer interaction with the clinical staff was also observed, and periodic grand rounds were proposed. Due to the hospital structure, focused on childhood cancer care, the identified at-risk family members were referred to external health care. In Brazil, there were no public provision of genetic testing, and since patients cannot afford their costs, DNA and tumor samples were stocked on a biobank, with all legal and ethical policies being observed, hoping for a change in this situation. CONCLUSIONS: A genetic cancer clinic in a pediatric oncology hospital, focused on care, research and education, helps the identification of a specific tumor predisposition or a congenital malformation syndrome in children with cancer, adding value on patient's management and counseling of at-risk relatives.

2404S

Implementing a universal public health policy for the care of rare diseases in Brazil. *D.DG. Horovitz[1], M.J.B Aguiar[2], V.E.F. Ferraz[3], M.F. Galera[4].* 1) Centro de Genetica Medica, Instituto Fernandes Figueira / Fiocruz, Rio de Janeiro, RJ., Brazil; 2) Faculdade de Medicina da Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 3) Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil; 4) Departamento de Pediatria - Faculdade de Medicina da Universidade Federal de Mato Grosso, MT, Brazil.

Brazil is an 8 million km^2 developing country, with 190 million population and 3 million births/yr. Since 1988 the Unified Health System (SUS) became universal and accessible to all. Better access to health care and partial control of poverty-related diseases led to progressive increase in burden of birth defects, shifting from fifth to second leading cause of infant mortality, highlighting the need for specific health policy strategies. Scattered governmental actions related to birth defects/genetic diseases comprise birth defects notification, newborn screening, rubella immunization, folic acid flour fortification and few treatment programs for rare diseases. Despite close relationship between clinical genetics and management of birth defects/genetic diseases, less than 30% of the total demand is currently met, due to difficult access to specialized services, insufficient laboratory support and lack of funding. Most genetic services are not fully integrated to the SUS, and non-recognition of medical genetics in the health system impeded their growth. A national policy for clinical genetics was published in 2009 but never implemented. In 2012, after significant pressure from users, a working group to elaborate a national policy for rare diseases in the SUS, composed of specialists, users and technicians from the Ministry of Health was instituted. After public consultation in 2013, the policy was published early 2014 and its implementation is underway. It was designed to cover rare diseases as lines of care, classified as genetic (congenital anomalies and later onset diseases, intellectual disability, inborn errors of metabolism) and non-genetic (inflammatory, infectious, autoimmune and other) rare diseases. The policy will be centered in clinical services, for which minimal requirements must be met (access to laboratory investigation, developmental services, genetic counseling); it will also regulate treatments for rare diseases. An official policy for rare diseases means recognition of the importance of medical genetics in the care and prevention of birth defects, reimbursement for consultations and exams, rational use of resources and enhanced coverage. The organization of a functional integrated network will not only use the already existing services as the backbone, but also stimulate the creation of new clinics throughout the country. Rare disease prevention, education, and epidemiological data collection are recommended as complementary measures.

2405S

Effect of Co-Payment on Behavioral Response to Consumer Genetic Testing. *J. Outlaw[1], L. Ariniello[2], W. Liu[1], D. Boeldt[1], N. Wineinger[1], N.J. Schork[1], E.J. Topol[1], C.S. Bloss[1].* 1) Rady School of Management, UCSD, San Diego, CA; 2) Scripps Translational Science Institute, La Jolla, CA.

Existing research in consumer behavior suggests that a person's perception and usage of a product post purchase might depend on how the product was marketed, including the price paid. As improvements in technology make genetic testing cheaper and more accessible to increasing numbers of consumers, it is important to consider the effect marketing practices may have on the adoption of and response to testing. In particular, it is unclear how the price paid for a genetic test may influence the consumer's subsequent behavior. In the current study, we examined the effect of price paid for a direct-to-consumer (DTC) genetic test on consumer post purchase behavior. Participants were enrolled in the Scripps Genomic Health Initiative (SGHI), a longitudinal cohort study originally designed to determine the impact of DTC genomic testing on consumers. Co-payment amounts for the test varied from $150 to $470, representing a price increase over time to encourage early enrollment. A small proportion of participants received fully subsidized testing from their employer. Of the subjects recruited for the study, 2,037 completed behavioral and psychological assessments before and approximately 5.6 months after genomic testing. Results indicated that with respect to clinical screening test completion, consumers that paid for the genetic test were significantly more likely to both obtain follow-up health screenings ($p<0.05$) and to share results with their physician ($p<0.05$), compared to those whose test was fully subsidized. Prior to receipt of their genetic test results, neither group differed in their propensity to visit a physician. In addition, neither group showed any significant changes in diet, exercise, or anxiety after receiving genetic risk results. The out-of-pocket price paid for DTC genetic tests may influence how results are interpreted and acted upon by consumers. Cognitive dissonance theory may help explain the increase in screening propensity for paying consumers. Such individuals may obtain follow-up screenings as a way of confirming the value of the test and validation of their initial decision to expend personal resources to obtain it. Understanding of co-payment effect on behavioral response to genetic testing may help shape public policy in this rapidly evolving area of health care.

2406S

Costs and burden of pediatric hospital admissions with a single gene or chromosomal diagnosis, United States, 2009. *R. Fisk Green[1], S. Grosse[2], R.S. Olney[2].* 1) Carter Consulting, National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA; 2) National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA.

To update and better understand the impact of single gene and chromosomal disorders on pediatric hospitalizations and inpatient costs, we analyzed a sample of pediatric inpatient hospital admissions with an ICD-9 code indicative of a single gene or chromosomal disorder diagnosis using 2009 national data from the Kids' Inpatient Database, Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. About 3% of admissions included a single gene or chromosomal disorder diagnosis, with an estimated annual cost >$5 billion (about 12% of estimated total pediatric inpatient hospital costs). Diagnoses observed most often and with highest estimated total admission costs included sickle cell disease, cystic fibrosis, muscular dystrophy, hemophilia and other bleeding disorders, metabolic newborn screening disorders, hereditary immunodeficiencies, velocardiofacial (VCF)/DiGeorge syndrome, neurologic disorders, skeletal dysplasias, and trisomy 21. Admissions with severe combined immunodeficiency (SCID) or VCF/DiGeorge syndrome diagnoses had the highest estimated mean cost per admission, while estimated mean costs per day were highest for admissions with diagnoses of Fragile X syndrome, MEN syndrome, hemophilia and other bleeding disorders, Apert syndrome, or VCF syndrome. For those admissions with a single gene or chromosomal code, the estimated mean length of stay was more than twice as long as those without these codes. Among admissions with a single gene or chromosomal code, those with combined immune deficiency/SCID or galactose disorders had the longest estimated mean lengths of stay. About 20% of admissions in which the child died during hospitalization included a single gene or chromosomal code, with the highest estimated number of deaths for organic acid disorders. The highest estimated number of admissions for most single gene or chromosomal codes occurred among infants <1 year of age, and the estimated number of admissions decreased with age. However, the estimated number of admissions was higher in late adolescence, compared with younger ages, for admissions with some single gene or chromosomal diagnoses, including sickle cell disease, cystic fibrosis, neurofibromatosis, von Willebrand disease, Ehlers Danlos syndrome, and hereditary peripheral neuropathy/Charcot Marie Tooth. Our results confirm that single gene and chromosomal disorders, while individually rare, substantially contribute to hospital inpatient costs and utilization.

2407S

Utilizing telemedicine to support informed decision making and expand access to cancer genetic services in community clinics. *A.R. Bradbury[1,2], L. Patrick-Miller[3], D. Harris[1], E. Stevens[1], B. Egleston[4], L. Fleisher[5], R. Mueller[1], A. Brandt[1], J. Stopfer[1], P. Higgins[1], S. Domchek[1].* 1) Medicine, Division of Hematology-Oncology, University of Pennsylvania, Philadelphia, PA; 2) Medical Ethics and Health Policy, University of Pennsylvania, PA; 3) Medicine, Division of Hematology-Oncology, University of Chicago, Chicago, IL; 4) Biostatistics, Fox Chase Cancer Center, Philadelphia, PA; 5) Children's Hospital of Philadelphia, Philadelphia, PA.

Background: Given high demand and a limited workforce, many patients do not have access to genetic providers. Telemedicine (TM) has been used to expand services to low access populations and could increase the use of pre-test counseling, informed decision-making and the appropriate use of genetic services. Methods: Patients at 3 community sites received pre-test (V1) and post-test (V2) counseling for cancer susceptibility with a genetic counselor (GC) at an expertise center through secure 2-way audio and visual communication. Surveys assessed knowledge, satisfaction, psychosocial responses and experiences. We used paired T-tests for changes between time points and linear regressions. Results: Of 100 patients approached, 83% consented to telegenetic services. 57 have completed V1 and 70% proceeded with testing. 31 patients have received results, including 3 carriers (BRCA2, MSH2, PMS2). 4% of sessions were aborted due to technology failure, 30% had disconnections but were completed by TM. Although 34% of patients reported technical difficulties, 94% were satisfied with their telegenetics visit. Knowledge (BL 20.9, SD 2.8; postV1 22.0, SD 3.0; $p<0.01$) and satisfaction with genetic services (SGS) significantly increased (postV1 39.5, SD 3.9; postV2 42.2, SD3.6; $p<0.01$), general anxiety (BL 7.4, SD 4.1; postV1 6.6, SD4.1; $p=0.02$; postV2 5.7, SD 3.5, $p=0.06$) and depression (BL 3.9, SD 3.9; postV1 3.5, SD 3.4; $p=0.05$; postV2 2.9, SD 3.5, $p=0.07$) significantly decreased, event anxiety did not change significantly (Table 1). Patients reported many advantages to telegenetics (e.g. less travel burden, informative) and few disadvantages (e.g. audio and technical challenges, less personal). Conclusions: Ensuring informed decision making across diverse clinical populations is increasingly important given the range of testing options (targeted v. multiplex) with variable utility and risk for uncertainty. Telemedicine delivery of cancer genetic services is feasible, increases knowledge, identifies mutation carriers and family history related cancer risk, and is associated with high satisfaction, suggesting an innovative model for delivery in community practices without institutional access to genetic providers.

2408S

Cree Leukoencephalopathy and Cree Encephalitis Carrier Screening Program: Evaluation of Knowledge and Satisfaction of High School Students. *J. Le Clerc-Blain[1], V. Gosselin[2], A. Bearskin[3], J.E. Torrie[2], G.A. Mitchell[1,4,5], B.J. Wilson[6], A. Richter[1,4,5], A.M. Laberge[1,4,5].* 1) CHU Sainte-Justine, Montreal, Quebec, Canada; 2) Cree Board of Health and Social Services, Chisasibi, Quebec, Canada; 3) Eeyou Awaash Foundation, Chisasibi, Quebec, Canada; 4) Medical Genetics Division, CHU Sainte-Justine, Montreal, Quebec, Canada; 5) Department of Pediatrics, University of Montreal, Montreal, Quebec, Canada; 6) Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada.

BACKGROUND: A population-based carrier screening program (CSP) started in 2006 for Cree encephalitis (CE) and Cree leukoencephalopathy (CLE), two neurodegenerative autosomal recessive conditions with high carrier rates in the James Bay Cree communities (Northern Quebec, Canada). Developed by local health authorities in collaboration with the Eeyou Awaash Foundation (community family support group), the CE-CLE CSP offers education/counseling sessions and carrier screening, targeting high school students (≥ 14 years), and women of reproductive age and their partners, mostly in prenatal settings. OBJECTIVE: To describe high school students' knowledge and satisfaction with the CE-CLE CSP. METHODOLOGY: Surveys were handed out before (survey A) and after (survey B) the CSP education sessions to grade 9-11 students in six high schools (n=267). Students who had already been tested were excluded (n=43). Data was collected on demographics, knowledge of CE-CLE (clinical features, myths, inheritance, carrier risk) and satisfaction with the program. We report results from our analysis of both surveys. Chi-square tests were used to compare results between surveys. RESULTS: Of 224 eligible students, 90 (40%) answered survey A, and of these 42 answered survey B. Ages ranged from 14-20 years. Seventeen (19%) were either pregnant or already had had at least one child. Eleven (12%) reported a positive family history for CE or CLE. When comparing groups who answered both surveys, improvement in knowledge following the education session was significant for 11/14 questions ($p<0.05$). Of the 42 respondents, 34 (81%) thought they had enough or more than enough information to help them decide about carrier testing, and 30 (71%) thought it is a good or very good idea to offer carrier testing in high school. Immediately following the education session, 59/224 (26%) of eligible students had CE-CLE carrier testing. FURTHER STUDIES: We will increase participant numbers by administering surveys A and B in three more high schools, and assess screening outcomes (e.g. impact of carrier status on reproductive decisions, use of cascade screening), knowledge retention and satisfaction prospectively in a follow-up survey (survey C), one year after participation in initial surveys. CONCLUSION: Education sessions seem to have a positive impact on high school students' knowledge of the CE-CLE CSP. High uptake and positive image of carrier screening confirms its acceptability in high schools.

2409S

Determinants of the value of genetic testing in clinical decision-making. *B. Lerner[1], N. Marshall[2,5], S. Oishi[2], A. Lanto[2], A.B. Hamilton[2,4], E. Yano[2,5], M.T. Scheuner[2,3,5].* 1) Veterans Hospital Administration, Boston, MA; 2) VA Greater Los Angeles Healthcare System, Sepulveda, CA; 3) VA Greater Los Angeles Healthcare System, Los Angeles, CA; 4) David Geffen School of Medicine, University of California, Los Angeles, CA,; 5) UCLA Fielding School of Public Health, Los Angeles, CA.

Objective: The value of genetic testing in clinical decision-making depends on factors within the environmental, organizational, provider, patient and encounter domains. We sought to identify determinants within those domains that are associated with favorable value ratings for the use of genetic tests for six clinical indications. Methods: We administered a key-informant Web-based survey to clinical chiefs of neurology, medical oncology, gastroenterology, cardiology and primary care at Veterans Health Administration facilities. The value of genetic tests was rated on a 5-point scale, then dichotomized into “not at all, minimally and moderately valuable” and “very to extremely valuable.” Bivariate and logistic regression analyses identified associations between determinants and the value ratings for each indication. Results: The response rate was 63%; (n=353). Genetic testing was rated as very-extremely valuable for the indications: inform clinical management (58.6%;), inform disease prevention (56.4%), assist with reproductive options (50.1%), assist with life planning (43.9%), confirm a suspected diagnosis, (39.9%), and confirm an established diagnosis (32.3%.) The most frequently identified determinants associated with favorable value vs. unfavorable ratings regardless of indication were a culture that fosters genetic testing (ranging from OR 4.8, CI 2.3-9.9, p<0.001 to OR 2.3, CI 1.2-4.6, p<0.01) and the availability of genetic testing guidelines (ranging from OR 2.8, CI 1.3-5.9, p<0.001 to OR 1.9, CI 0.99-3.8, p<0.1). Other significant associations include: evidence demonstrating utility of genetic tests (OR 3.4, CI 1.7-6.9, p<0.001); provider specialty using primary care as the reference (ranging from OR 0.3, CI 0.2-0.8, p<0.01 to ability to OR 2.8, CI 1.1-7.2, p<0.5), obtain genetic consults (OR 2.4, CI 1.4-4.2, p<0.001); experience ordering genetic tests (OR 2.6, CI 1.3-5.2, p<0.01); and patient demand for testing (ranging from OR 1.8, CI 1.0-3.2, p<0.05 and OR 1.6, CI 0.94-2.9, p<0.1). Conclusion: We identified several determinants associated with the value of genetic tests in patient care. These findings can guide the design of multi-level implementation efforts within healthcare organizations adopting genetic testing services. Such programs should focus on strengthening a culture that facilitates testing, promoting guidelines for genetic testing, educating providers and patients about genetic testing, and ensuring access to genetic consults.

2410S

Lean - production management rules applied to a genomics core facility. *J. Altmüller[1,2], C. Becker[1], P. Nürnberg[1].* 1) Cologne Center for Genomics, Cologne, Germany; 2) Institute for Human Genetics, University of Cologne, Germany.

Lean management is a production management systems approach, originally coming from the automobile company Toyota. It was reviewed to have an enormous value creation per employee, superior to anything known before. Starting 25 years ago, this success model was investigated in the western world, standardized, adopted, and expanded in many fields of production. The Next Generation Sequencing platform with its large variety of applications, fast changing protocols, and comparably small projects intensified the need for professional workflow and capacity planning and quality control implementation. Adopting lean principles, our core facility steadily improves service orientation, data quality and cost effectiveness. With this presentation we will show how lean management tools can help both research and diagnostic labs to maximize flexibility, minimize waste, and reduce mistakes.

2411S

Newborn screening for cystic fibrosis: role of primary care providers in caring for screen positive children. *J.C. Carroll[1], R.Z. Hayeems[2,3], F.A. Miller[2], C.J. Barg[2], Y. Bombard[2,4], P. Durie[5,6], P. Chakraborty[7,8], B.K. Potter[9], J.P. Bytautas[10], K. Tam[11], L. Taylor[12], E. Kerr[13], C. Davies[7], J. Milburn[7], F. Ratjen[5,6,12], A. Guttmann[2,5,6,14].* 1) Family & Community Med, Mount Sinai Hospital University of Toronto, Toronto, ON, Canada; 2) Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, Canada; 3) Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, Canada; 4) Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada; 5) Division of Paediatric Medicine, Department of Paediatrics, The Hospital for Sick Children, Toronto, Canada; 6) Department of Pediatrics, Faculty of Medicine, University of Toronto, Canada; 7) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, Canada; 8) Department of Pediatrics, Faculty of Medicine, University of Ottawa, Ottawa, Canada; 9) Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Canada; 10) Division of Health Care and Outcomes Research, Toronto Western Research Institute, Toronto, Canada; 11) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada; 12) Division of Respiratory Medicine, Department of Pediatrics, The Hospital for Sick Children, Toronto, Canada; 13) Department of Psychology, The Hospital for Sick Children, Toronto, Canada; 14) The Institute for Clinical Evaluative Sciences, Toronto, Canada.

Context: Expanded newborn screening (NBS) has increased positive screening results, prompting attention to the role of primary care providers (PCP) in providing care to children who screen positive for genetic disorders. Objective: To explore PCPs’ role in caring for children who receive a positive cystic fibrosis (CF) newborn screen and are confirmed to have CF, an uncertain CF diagnosis or are CF carriers. Participants: PCPs in Ontario, Canada, identified by Newborn Screening Ontario as having a positive CF NBS result in their practice in the previous 6 months. Intervention: Mailed questionnaire inquiring about their role in caring for children with CF, uncertain CF diagnosis and CF carriers, confidence in care, and management of CF family planning issues. Outcome measures: The primary outcome measure was the role PCPs identified in providing well-baby care for children with CF, uncertain CF diagnosis and CF carriers. Results: 329/653 PCPs (50%) completed surveys. Survey respondents were: 65% family physicians/nurse practitioners, 21% pediatricians, 14% midwives. Most PCPs (81%) agreed they have an important role to play in NBS. For infants confirmed to have CF, for routine well-baby care: 22% PCPs would provide total care, 69% share with a specialist, 9% refer to a specialist completely; for infants with uncertain CF diagnosis: 49% would provide total care, 46% share with a specialist, 5% refer to a specialist completely and for CF carriers: 89% would provide total care, 9% share with a specialist, 2% refer to a specialist completely. Most PCPs were extremely/very confident (54%) or moderately confident (35%) in providing reassurance about the health of CF carriers. Only 24% knew how to order CF carrier testing for parents, 66% knew how to refer for prenatal diagnosis. Conclusion: The majority of PCPs are willing to treat infants with a range of CF NBS results in some capacity. It is concerning that 11% indicated that carriers should have specialist care involvement and only 54% were very confident reassuring about carrier status. This raises issues about the possible medicalization of carrier status, prompting the need for specific education for PCPs about genetic disorders and the meaning of genetic test results.

2412S

An assessment of clinician and researcher needs for support in the era of genomic medicine. *C.A. Brownstein[1,2,5], S.K. Savage[1], S.I. Ziniel[2,3,4], I.A. Holm[1,2,5], J. Stoler[1,2], D.M. Margulies[1,2,6,7].* 1) Genetics-Research Connection, Boston Children's Hospital, Boston, MA; 2) Department of Pediatrics, Harvard Medical School; 3) Division of Adolescent/Young Adult Medicine, Boston Children's Hospital; 4) enter for Patient Safety and Quality Research, Program for Patient Safety and Quality, Boston Children's Hospital; 5) Manton Center for Orphan Disease Research, Boston Children's Hospital; 6) Division of Developmental Medicine, Boston Children's Hospital; 7) Center for Biomedical Informatics, Harvard Medical School.

Introduction: To assess and characterize the need for a centralized service to support the utilization of genomic testing at Boston Children's Hospital (BCH), we conducted a web-based survey of clinicians and researchers regarding their past, current, and anticipated future use of next generation sequencing and their anticipated needs for support. Results: We received 283 survey responses from 36 departments and divisions (5.5% response rate). A large proportion of clinician respondents anticipate that they will use exome/genome sequencing (44.8%) and/or candidate gene panels (50% clinician) within the next year. About an equal percentage of researcher respondents anticipate the need for exome/genome sequencing (48.0%), while a smaller percentage anticipates the use of candidate gene panels (31.8%). Estimated volume data show an increase in the use of these tests in the future compared to past estimates. However, few respondents (13.6%) said that they felt "Completely Ready" or "Pretty Much Ready" to incorporate NGS into their clinical practice or research. Respondents indicated varying degrees of need for a diverse list of support services, with interpretation and clinical correlation assistance ranked as the most needed services. Conclusion: The results of this study highlight the current and growing need for various types of education, support, and assistance (e.g., data and result interpretation) and make a convincing argument for hospitals and medical centers to invest in the development of institutional genomic services. Some of the key challenges to the success of such genomic service models will involve the need for more precise, powerful, and efficient interpretation tools; expanded data sources; and expert resource networks to inform the interpretation process, not to mention issues of cost, billing, and reimbursement around these components. Data from this study will guide BCH's establishment of a new institutional service, Interpretive Genomic Services (IGS), which will support clinicians and researchers in their use of NGS testing and data. BCH will leverage existing institutional expertise in genetics and genomics, genetic counseling, bioinformatics, and many subspecialties to create a true multi-disciplinary service.

2413S

Physician Response to EMR-Based Genetic and Non-Genetic Risk Results for Actionable Complex Disease and Pharmacogenomics in a Randomized Controlled Trial of Genomic Counseling. *K. Sweet[1], J.F. Garcia-Espana[2], A.C. Sturm[1], T. Schmidlen[2], K. Manickam[1], J. McElroy[1], A.E. Toland[1], J.Scott Roberts[3], M. Christman[2].* 1) Division of Human Genetics, Ohio State University, Columbus, OH; 2) Coriell Institute for Medical Research, Camden, NJ; 3) School of Public Health, University of Michigan, Ann Arbor, MI.

Background: The correct interpretation of genomic variant and family history information for actionable complex disease and pharmacogenomics (PGx) may motivate physicians and patients to take preventive and clinical actions. The impact of genomic counseling performed by genetic counselors on this process has yet to be established. The OSU-Coriell Personalized Medicine Collaborative (CPMC®) is a physician-blind randomized controlled trial of in-person genomic counseling (GC) for patients with chronic disease in a hospital setting. **Method:** 208 patients (mean age: 58 years; 45% female; 56% with a bachelor degree or higher) were enrolled by 8 Cardiovascular Medicine (CM), 12 Internal Medicine (IM), and 11 Family Medicine (FM) physicians over a two year period; 5 were recruited by Media/Research Match. Participants received 8 actionable complex disease reports (e.g., type 2 diabetes; macular degeneration; coronary artery disease) and one PGx result (*CYP2C19*: clopidogrel) via the CPMC® web portal. Patients seen for GC received a summary letter post-visit. All CM/IM physicians received a pre-study educational module on genetics, genomics, PGx, study processes, the availability of test reports and GC summary letters in the EMR; FM physicians had no educational intervention. The primary outcome analyzed was clinical relevant actions taken (yes; no) by the physician team. The EMR was searched for key terms: genetic, genomic, variant, risk, family history, genetic study, CPMC. Relevant actions taken (e.g. ophthalmology referral) were recorded. A secondary outcome was the difference in physician actions by educational intervention (CM/IM; FM). **Results:** Analyses were conducted based on the intention-to-treat principle. One quarter of patients had actions taken by a physician. A multivariable logistic regression model estimated the independent effect of genomic counseling on physician action, adjusted for gender, education and significant covariates. Patients in the GC arm were more likely to have actions taken as compared to the no-GC arm (OR: 2.17; 95% CI: 1.44-3.27; p=0.0002). There was a trend for a positive association between higher level of patient education and actions taken (OR=1.67; 95% CI: 0.97-2.85; p=0.063). No significant difference in physician actions by educational intervention group was found (CM/IM, 26%; FM, 19%; p=0.32). **Conclusion:** Genomic counseling of patients was associated with increased preventive and clinical actions taken by physicians.

2414S

A rapid evidence review for the inclusion of genetic data in clinical care for a common, complex disease. *J. Malinowski[1], E.W. Clayton[2], D.C. Crawford[1].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN.

Personalized medicine (PM), defined as differentiated clinical care based, in part, on the genetic background of an individual, has been implemented for cancer treatments and to guide pharmaceutical interventions for various clinical traits. Despite a recent history of direct-to-consumer (DTC) companies offering risk prediction for dozens of diseases and traits based on the genotyping of hundreds of thousands of single nucleotide polymorphisms (SNPs), the clinical validity and utility of these SNPs for common, complex diseases have not been adequately investigated. We performed a rapid evidence review in an academic setting to identify analytic evidence that the inclusion of genetics in routine clinical care improves health outcomes for complex disorders. We identified a common, complex disease (hypothyroidism) and a related quantitative trait (serum thyroid hormone (TSH) levels) with several known genetic associations, including *FOXE1* rs7850285, *PTPN22* rs24766012, *SH2B3* rs3184504, *VAV3* rs4915077, and *HLA* region rs2517532. These variants were previously used by a DTC company to report risk of developing hypothyroidism. We developed key questions to identify analytic validity, clinical validity, clinical utility, and the ethical, legal, or social issues (ELSI) pertaining to the use of these genetic data in the clinical setting. Using medical subject headings (MeSH) terminology, we performed a PubMed search for articles, completed abstract and full text reviews, and analysis of the results. All data collection and review were maintained in REDCap databases. Six hundred and thirty abstracts were evaluated, from which 346 articles were full text reviewed (54.9%). Fifteen (4.3%) of the full text reviewed articles included odds ratios or effect sizes and p-values for the genetic association. None (0) of the articles provided evidence of clinical validity or utility for the genetic associations. This rapid review was completed in seven months over two academic semesters with two individuals. Our study supports others that have found a general lack of published data to indicate genetic research findings have demonstrable clinical utility. Despite these negative results, our rapid review methodology serves as an example of deploying an evidence review in an academic setting to systematically identify the essential data required to accurately assess the utility of including genetic data to improve health outcomes for common, complex diseases.

2415S

Genetic Test Recipient as a Source of Data: A Novel Approach to Obtaining Genotypic and Phenotypic Data for Input into Genomic Databases. *B.E. Kirkpatrick[1], E. Riggs[1], M. Giovanni[1], R. Green[2], A. Janze[3], P. Krautscheid[4], J. Krier[2], C.L. Martin[1], D. Metterville[5], D. Miller[6], M. Murray[1], H. Rehm[5,7], D. Riethmaier[8], W. Rubinstein[9], B. Smith-Packard[1], C. Tan[10], K. Wain[11], W.A. Faucett[1] on behalf of the Clinical Genome (ClinGen) Resource.* 1) Geisinger Health System, Danville, PA; 2) Division of Genetics and Department of Medicine, Brigham and Women's Hospital, Boston, MA; 3) GeneDx, Gaithersburg, MD; 4) ARUP Laboratories, Salt Lake City, UT; 5) Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, MA; 6) Department of Laboratory Medicine and the Division of Genetics, Boston Children's Hospital, Boston, MA; 7) Department of Pathology, Harvard Medical School, Boston, MA; 8) InVitae, San Francisco, CA; 9) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD; 10) Department of Human Genetics, University of Chicago, Chicago, IL; 11) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

Large datasets of genomic variants and observed phenotypic consequences have been identified as important resources of information that will help clinicians and researchers understand the extensive variability in the human genome and its impact on disease. The Clinical Genome (ClinGen) Resource, together with the National Center for Biotechnology Information (NCBI), is developing ClinVar, an open-access database of genomic variants, associated phenotypic information, and clinical classifications with supporting evidence. Genomic databases including ClinVar are incomplete, and various avenues are being investigated to increase the amount of information available. As genotypic information in isolation has limited value, the availability of phenotypic information is crucial to variant interpretation. Previous efforts have focused on obtaining linked genotypic and phenotypic information from laboratories, researchers, and clinicians. These approaches often yield limited information due to lack of time, resources, and/or incentive. In contrast, patients themselves are valuable sources of information regarding their own medical histories, and they may be highly motivated to share this information for the benefit of advancing scientific understanding. By developing an online patient registry portal, ClinGen seeks to enhance the detailed genotypic information available from laboratories with detailed phenotypic information provided by patients. Registry participants will complete an online consent process explaining the project and implications of genomic data sharing. Each participant will be asked to complete a general, review-of-systems-type online health questionnaire; follow-up questionnaires will obtain detailed information on health issues identified on the original survey. Participants also will be asked to share a copy of their genetic test results, which will be curated by the registry coordinator. This approach will allow for the accurate representation of genotype information. Strategies for allowing the linkage of data from different sources on the same individual into ClinVar are in development. As far as the researchers are aware, this is the first attempt to obtain information directly from genetic test recipients for the purposes of building genomic databases.

2416S

A comparison of anthropometric status, blood pressure and oxygen saturation in athletic and non-athletic females. *H. Kaur, B. Badaruddoza.* Human Genetics, Guru Nanak Dev University, Amritsar, India.

Obesity and overweight for all age groups are major concern in both economically developing and developed countries. Physical and mental health is directly related to obesity and hypertension. This study aimed to compare anthropometric status, blood pressure and oxygen saturation between athletic and non-athletic females. The athletic and non-athletic female subjects aged 18-25 years were selected from Guru Nanak Dev University, Amritsar, India. The anthropometric, physiometric and socio-economic variables along with oxygen saturation was recorded for all the study samples. The data was analyzed on SPSS v. 18.0 using correlation, ANOVA, univariate and multivariate regression analysis.In this study weight gain, BMI, SBP and DBP were found higher in athletic female group as compared to non-athletic group. Whereas, waist circumference, hip circumference, waist-hip ratio and skinfolds were higher in non-athletic female group. The percentage of saturated oxygen content has been found maximum in non-athletic group.

2417S

The NINDS Repository: A Public Collection of Biomaterials for Disease Modeling, Gene and Biomarkers Discovery in Neurological Research. *C. Tarn[1], G. Balaburski[1], S. Heil[1], K. Reeves[1], J. Santana[1], J. Gilroy[1], M. Self[1], K. Gwinn[2], M. Sutherland[2], R. Corriveau[2], C. Pérez[1].* 1) Coriell Institute for Medical Research, Camden, NJ; 2) National Institute for Neurological Disorders and Stroke-NIH, Bethesda, MD, USA.

Neurological diseases are a major public health care concern but the pathological mechanisms in neurodegenerative disorders remain largely not understood. The major challenges in disease mitigation reside on lack of reliable genetic and molecular biomarkers for caustic, diagnosis, and progression monitoring; and limited reproducible cellular models for reearch. The National Institute of Neurological Disorders and Stroke (NINDS) Repository, funded by NINDS, was established with the mission of providing high quality biospecimens as a strategy to facilitate and accelerate research in neurological diseases. The NINDS Repository collects biosamples and de-identified clinical data from diverse patients diagnosed with various neurological disorders as well as neurologically normal controls. In addition, the NINDS Repository features collections of patients-derived fibroblasts and induced pluripotent stem cells (iPSC) with well-defined mutations as essential research tools for understanding the pathological mechanisms and establishes cellular models for neurological diseases. Recently, the NINDS Repository has broadened its collections to include whole blood RNA, plasma, serum, cerebrospinal fluid, and urine, to facilitate biomarker research utilizing longitudinal samples from both affected and neurologically healthy individuals. Since its establishment, biomaterials from more than 44,000 individuals with cerebrovascular diseases, Parkinsonism, motor neuron diseases, epilepsy, Tourette syndrome, Dystonia, Huntington's disease and neurologically-normal controls have been banked in the NINDS Repository. The NINDS Repository has established validated standard operating procedures and rigorous quality control assessments that span the life cycle of all biospecimens collected to provide premium samples. The NINDS Repository aims to ensure and implement standardization for collecting and processing across all samples. In addition, the NINDS Repository utilizes secure and integrated laboratory information management systems to monitor inventory, processing, storage, and distribution of biospecimens, and facilitates sample-data association by cross-referring with other databases. By developing such a centralized collection of human biospecimens and their associated de-identified clinical data, the NINDS Repository thus provides a vital resource for research designed to discover and validate genetic and proteomic biomarkers of neurological disorders.

2418S

Patients and caregivers as sources of innovative ideas and solutions: A multiple case study approach. *V. Francisco[1,2], R. Oliveira[3], P. Oliveira[3], L. Zejnilović[3], S.A. Oliveira[1,2].* 1) Instituto Medicina Molecular, Portugal, Lisboa, Portugal; 2) Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal; 3) Católica-Lisbon School of Business and Economics, 1649-023 Lisboa, Portugal.

Recent academic literature shows that patients and caregivers are a significant source of innovative solutions related to their medical condition. To date, little is known about the process by which these innovations emerge, how they diffuse, and how they impact the lives of patients and caregivers. In this work we follow a multiple-case study approach to map a set of patient innovations and adoptions of patient innovation cases, and systematically explore how and how far patients and caregivers innovate in the health care field. In addition, we propose some explanations for why patients and caregivers stop at a certain stage of progression of the innovation process. We conducted 15 extensive semi-structured interviews with patients and caregivers of the following group of diseases: spinal cord injuries, Angelman syndrome, epidermolysis bullosa, cerebral palsy, and hemiparesis. These individuals shared with us their experiences with their disease and their efforts, or the lack of them, to overcome specific problems related to their health condition. This includes 4 “holistic” case studies and 26 “embedded” case studies. With the information from these cases, we analyze patient innovation paths and present them in the fall-offs conceptual framework. Through a cross-case analysis, we find that duration of the experience with the disease, complexity and pressure of a certain situation, belonging to a group or a community, and perceived value of a solution are among the most important reasons that impact how far patients and caregivers take their innovations. As a result of our multiple-case analysis we present a set of propositions from which future research in the field is warranted.

2419S

Diffusion as a validation process: Learning from patient innovators. *S.A. Oliveira[1,2], T. Fidelis[3], L. Zejnilovic[3,4,5], P. Oliveira[3].* 1) Instituto Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal; 2) Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal; 3) Católica-Lisbon School of Business and Economics, 1649-023 Lisboa, Portugal; 4) Carnegie Mellon University, Pittsburgh, PA 15213, USA; 5) Instituto Superior Técnico, 1049-001 Lisboa, Portugal.

There is growing evidence that patients and their family members innovate in therapies and medical devices. These patient innovators are increasingly recognized as an important source of innovation. However, little is known about the paths they take to validate their solutions before involving health regulators. We attempt to understand these patterns of informal validation and how their comparison with current validation methods of medical innovations can improve such processes. To address this, we study cases of patient innovators who attempted to share their innovations with other patients. More specifically, we structure the observed processes of patient innovation diffusion as a process model for an iterative observational trial process in which incremental validation of the innovation is performed by diffusing it to other patients. We argue that learning from patient innovators’ practices is important for discussing viable frameworks and structures for improving these processes, and discuss related implications to innovation research, management and policy.

2420S

Multiplex panel testing improves the yield of mutation detection in cancer genetics clinics. *J.O. Culver, C.N. Ricker, K. Lowstuter, D.Y. Sturgeon, C.R. Chanock, W.J. Gauderman, K. McDonnell, G.E. Idos, S.B. Gruber.* USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA.

Multiplex hereditary cancer panel testing for individualized cancer risk assessment is commercially available, but the diagnostic yield and clinical utility of panels are not yet known. We conducted a retrospective IRB-approved chart review of 317 cancer genetics clinic patients tested with a multiplex cancer gene panel. We analyzed the yield of panels versus the traditional approach of testing for specific hereditary syndromes. Panels were performed by Myriad (n=242), Ambry (n=64), Fulgent (n=9), University of Washington (n=1), and City of Hope (n=1). Our cohort included 142 patients from the USC Norris Comprehensive Cancer Center and 175 patients from the Los Angeles County + USC Medical Center. The mean age was 48.8 years (SD=12.5) and 85.5% were female; 80.1% had a history of cancer (n=254); race/ethnicity reported was 48.9% Hispanic, 34.4% Caucasian, 4.4% Black, 10.0% Asian, and 2.2% other. Among 317 panel test results, 17.0% contained at least one deleterious mutation (n=54) and 47.3% contained at least one variant of uncertain significance (VUS) (n=150). Of 242 Myriad myRisk panel results, 16.9% had at least one deleterious mutation (n=41) and 48.4% had at least one VUS (n=117). Of the 64 Ambry panel results (CancerNext, OvaNext, BreastNext, ColoNext and others), 12.5% had at least one deleterious mutation (n=8) and 42.2% had at least one VUS (n=27). Among 129 patients with breast cancer, 6.2% carried a mutation in *BRCA1* or *BRCA2* (n=8) and 9.3% carried a mutation in *BARD1, CDH1, CHEK2, MUTYH, PALB2, RAD51D*, or *TP53* (n=12). Among 69 patients with colorectal cancer, 14.5% had a mutation in *APC, MLH1, MSH2, MUTYH*, or *PMS2* (n=10) and 2.9% had a mutation in *RAD50* or *RAD51C* (n=2). Thus, multiplex panel testing increased mutation detection over a traditional approach in 9.3% of breast cancer patients and 2.9% of colon cancer patients. Similar analyses for other cancer sites found the supplemental yield of panels was 7.4% for ovarian cancer (2/27), 11.7% for endometrial cancer (2/17) and 22.2% for gastric cancer (2/9). In the overall cohort of 317 patients, 7.6% (n=24) had a mutation identified on a panel that would not have been tested in a traditional approach. In most cases, identifying these mutations led to enhanced surveillance recommendations for the patient and family members. We conclude that multiplex panel testing increases the yield of mutation detection and adds to the capability of providing individualized cancer risk assessment.

2421M

Cost-effective NGS based BRCA - TP53 screening panel for hereditary breast cancer in India. *M. Sen[1,2], P. Agrawal[1,2], N.S.N. Swetha[1], V. Vittal[1], V. Pathak[1], G. Deshpande[1], D. Vishwanath[1,2], S. Sankaran[1], R. Hariharan[1], P. Ramamoorthy[1], K. Subramanian[1], V. Gupta[1,2].* 1) Strand Centers for Genomics and Personalized Medicine, Strand Life Sciences Pvt. Limited, Bangalore, India; 2) Strand at Mazumdar-Shaw Center for Translational Research, Bangalore, India.

Introduction: Hereditary BRCA1/2 mutations in breast cancer show high correlation with Triple Negative Breast Cancers (TNBC). In India, the incidence rate for TNBCs is approximately 25-30%, significantly higher than the rest of the world. This creates a need for a in-parallel, high throughput, cost-effective screening method for hereditary breast cancer mutations in India. Several studies in Indian cohorts also implicate mutations in TP53 gene (Li-Fraumeni syndrome) associated with multiple hereditary soft-tissue tumors including breast cancer. Cost effective hereditary mutation screening assays can have tremendous utility in the identification and management of women at high risk of hereditary breast cancer. We have designed and standardized a NGS based test for screening BRCA1/2 and TP53 genes. This assay is about one tenth the price of Sanger based sequencing, the current 'gold standard' for diagnostics in detecting hereditary breast cancer. Materials and Methods: We designed a comprehensive BRCA-TP53 NGS panel covering all HGMD/ClinVar mutations with the potential to identify any novel coding mutations. We demonstrate the use of 120mer IDT Lock down probes to perform in-solution capture of genomic DNA libraries using Agilent Sure Select kits. We have used 15 known samples (Hapmap, cell lines and pre-validated clinical samples) and evaluated the performance characteristics of the panel. Results: Our enrichment based assay panel covers 100%of the targets with greater than 20 reads (0 low coverage regions), with an average coverage of 260X per target per sample. The protocol is faster, automation friendly and potentially scalable to 96 parallel sequencing. The overall sensitivity, specificity and reproducibility characteristics of the panel were greater than 95%.Compared to commercially available research use only panels, our laboratory developed assay found additional mutations in the clinical validation samples Our results demonstrate higher design coverage and superior analytical performance of the Strand developed BRCA-TP53 NGS assay compared to known commercial NGS based panels for research use. Our results reflect that the assay performs uniformly even across high GC bias or low complexity regions, in these genes. Results will be presented from both technical and clinical validation of this panel vis-a-vis other commercial panels with these genes.

2422T

A Two-Step Clustering Methodology for Grouping Clinical Genetic Testing Panels. *G. Hooker, J. Staples, T. Murphy, M. Cottrell, M. Schenfield, G. Gross, B. Blackshear, J. Schneider.* Bioinformatics, NextGxDx, Franklin, TN.

Introduction: As the genetic testing landscape expands in both the complexity and quantity of clinically available testing products, the task of finding, comparing, and selecting appropriate genetic tests becomes increasingly burdensome. The genetic testing landscape can be divided into two categories: single-gene tests and panels. While tests, which interrogate a single genetic target are relatively simple to group, panels, which interrogate multiple genetic targets, prove considerably more difficult. We set out to develop an organizational schema that allows for the side-by-side comparison of similar panel testing products. Methods: Leveraging our data resource of all currently available genetic tests from more than 200 CLIA-certified US testing laboratories, we tested the efficacy of four unsupervised machine learning algorithms in clustering multi-target genetic tests (panels) based solely on their diagnostic targets (genes). Using hyper-parameter optimization and domain expertise, we identified a high performance algorithm, DBSCAN, and generated a novel set of genetic test clusters. Next, we manually curated the algorithmically-generated clusters to provide greater clinical relevance. Results: Our dataset initially contained 1875 panels manually grouped into 709 clusters (Avg. 2.7 tests/cluster, SD=2.5). Algorithmic sorting decreased the number to 358, while subsequent expert curation, taking into account intended use of the test, increased the number to 478, (Avg. 4.1 tests/cluster, SD=5.85). To assess clustering efficacy, we calculated the mean intra-cluster vertex connectivity, observing a 33% increase in relatedness within clusters following the two-step process (Manual: 1.45, Two-Step: 2.14). Conclusion: Our two-step clustering process effectively groups similar clinically available panels. As the genetic testing market expands, completely manual curation of genetic testing products becomes increasingly impractical and error-prone. Semi-automated methods of organizing data, such as the one presented here, increase accessibility of the wide range of available testing products, thus improving clinical workflows and, potentially, diagnostic outcomes.

2423S

Recurrent hydrocephalus by POMT2 mutation unraveled by exome sequencing of DNA from preserved Umbilicus. *K. Kosaki[1], T. Takenouchi[2], A. Wataanbe[3], H. Miyake[4], T. Abe[4], C. Torii[1], K. Nakabayashi[5], K. Hata[5], R. Kosaki[6].* 1) Center for Medical Genetics, Keio University, Tokyo, Japan; 2) Department of Pediatrics, Keio University, Tokyo, Japan; 3) Division of Clinical Genetics, Nippon Medical School Hospital, Tokyo, Japan; 4) Department of OBGYN, Nippon Medical School Hospital, Tokyo, Japan; 5) Department of Maternal-Fetal Biol, National Center for Child Health and Development, Tokyo, Japan; 6) Division of Medical Genetis, National Center for Child Health and Development, Tokyo, Japan.

The etiologies of congenital hydrocephalus include infections, vascular abnormalities, mechanical obstruction and chromosomal abnormalities. In the majority of patients, reaching the exact genetic diagnosis of congenital hydrocephalus is often challenging, mainly because of the limitations in the non-specific nature of ultrasonography findings. Here we report a 36-year-old father and a 37-year-old G4P1 mother who were referred to us for genetic counseling for recurrent congenital hydrocephalus. The mother's first pregnancy with a female fetus was complicated by meningocele first detected at 15 weeks of gestation. This pregnancy resulted in the spontaneous abortion. The second pregnancy with a female fetus complicated by hydrocephalus detected at 21 weeks of gestation, resulting in the termination of the pregnancy. The third pregnancy lead to a birth of normal male. In the fourth pregnancy, an ultrasonography examination performed at 21 weeks of gestation revealed severe congenital hydrocephalus, leading to termination of the pregnancy. Since none of the three affected fetuses had undergone a molecular diagnosis of their condition, upon the termination of the fourth pregnancy, the parents agreed to preserve a part of the umbilicus for future diagnostic procedures. The recurrent severe congenital hydrocephalus in both male and female fetuses conceived by healthy parents strongly suggested a autosomal recessive Mendelian disorder, and the parents were counseled as such. The parents were referred to us for further genetic evaluation. DNA samples were extracted from the preserved umbilicus obtained during the fourth pregnancy according to the standard protocol. The parents' DNA samples were extracted from their peripheral blood. Exome sequencing identified de novo compound heterozygous mutations in exon11 and exon19 of POMT2 (NM_013382), i.e., c.1248C>G p.His416Gln and c.1912C>T p.Arg638*, respectively. The present observation illustrates the utility of exome sequencing using preserved umbilicus in undiagnosed fetal structural abnormalities. The correct molecular diagnosis enabled us to provide the parents with the recurrence risk of 25% together with the possibility of prenatal genetic diagnosis for a future pregnancy. In light of the rapid advances in molecular diagnostic technology, the preservation of umbilical cord should be considered in fetal cases with undiagnosed structural defects, particularly when a Mendelian disorder is highly suspected.

2424M

Targeted Single-Molecule Oligonucleotide-Selective Sequencing for Genetic Diagnostics. *S. Bruce, P. Salmenperä, M. Gentile, S. Myllykangas.* Blueprint Genetics, Helsinki, Finland.

Diagnostic applications of next-generation sequencing have been impeded by complex experimental workflows associated with target DNA enrichment methods and errors and biases caused by PCR during sample preparation. Here we demonstrate single-molecule Oligonucleotide-Selective Sequencing (OS-Seq) for simplified and clinical-grade targeted sequencing. OS-Seq consolidates capture and sequencing of DNA in situ flow cell of a next-generation sequencing system. We developed a programmable and automated OS-Seq technique using the MiSeq sequencing system and implemented a PCR-free sample preparation method for single-molecule OS-sequencing. In addition, we introduced batch indexing, which utilizes several index sequences per sample and improves the uniformity of the sequencing yields between pooled samples. We developed two OS-Seq assays targeting coding exons, exon-intron boundaries and known pathogenic variants in introns of 136 and 49 genes (495,567 and 159,198 bases, respectively) implicated in cardiovascular diseases. The median sequencing depths were 420 and 733 and 99.35% and 99.8% of the bases on target regions were covered >15x. We validated SNPs, INDELs and CNVs using whole genome sequenced reference samples. High-quality genotypes for NA12878 were obtained from the Genome-in-a-bottle consortium and INDELs were confirmed with Sanger sequencing. Sensitivity to detect SNPs was >99% and 100%, respectively, and specificity was >99.99% for both assays. Sensitivity and specificity to detect short INDELs was 100% and sensitivity to detect CNVs of >1kb was >99%. We showed that single-molecule OS-Seq omits sequence duplicates, reduces sequencing errors by two-fold and improves uniformity of the coverage by 30%. Our results demonstrate the usability of OS-Seq for clinical sequencing applications.

2425T

Implementing next generation sequencing in clinical practice of children's hospital. *Y. Enomoto, I. Ohashi, Y. Kuroda, C. Hatano, T. Yokoi, K. Ida, T. Naruto, K. Kurosawa.* Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan.

Genetic disorders and birth defects account for a high percentage of the admissions in children's hospitals. Congenital malformations and chromosomal abnormalities are the most common causes of infant mortality. However, highly heterogeneous causes of the disorders in the cases of children's hospital provide low yield of molecular diagnostic procedures. Next generation sequencing (NGS) has a basically broad impact on many facets of biological and clinical research. To date the NGS is not routinely used in genetic diagnosis. Advent of NGS in the clinical laboratory setting requires the adoption of many processes and procedures. We developed diagnostic flows of molecular diagnosis using NGS to identify mutations in several malformation syndrome and related disorders in Kanagawa Children's Medical Center. This panel approach was successfully established in the cases with unknown sequence alterations. Depending on the clinical diagnosis, we identified causative mutations in the patients. NGS-based panel analysis is reliable and cost-effective approaches in clinical setting of heterogeneous diseases. However, the panel analysis is so limited in the case of application for the extremely high heterogeneous disorders, such as autism and unknown multiple anomalies. We present clinical utility of exon sequencing of Mendelian disorders (TruSight One) in the clinical context in our children's hospital. These results indicated the clinical exome sequencing is useful for pediatric practice, and bioinformatic process will become essential for the clinical practice.

2426S

Active organic solvent-free paraffin removal is the key to efficient extraction of NGS-quality DNA from FFPE tissues. *H. Khoja, E. Rudd, J. Han, A. Purdy, S. Kakumanu, A. Palmer, G. Durin, J. Laugharn.* Covaris Inc,, Woburn MA.

Formalin Fixation and Paraffin Embedding (FFPE) of tissues, a mainstay of clinical histological analysis for the past century, is rapidly being adopted for targeted and whole genome sequencing. The extreme formaldehyde fixation and tissue dehydration of FFPE preserved tissue presents not only a technical challenge to reproducible DNA and RNA extraction and molecular analysis, but also a workflow challenge in a clinical setting. In this poster we present truXTRAC™, a novel and highly reproducible method of extraction and purification of DNA and RNA from FFPE tissue utilizing Covaris Adaptive Focused Acoustics™ (AFA). Our novel method utilizes highly controlled acoustic energy for effective removal of paraffin from FFPE cores, sections, and slides enabling efficient and rapid tissue rehydration, tissue digestion, crosslink reversal, and nucleic acid release. This critical step is carried out without the use of dangerous organic solvents or messy mineral oils. We will present quantitative fluorescent microscopy data illustrating the efficiency of active paraffin removal from tissues utilizing this novel approach. We will also illustrate the significant effect of efficient active paraffin removal on downstream applications and analyses of DNA and RNA from FFPE tissues. Our results, utilizing replicates of different tissue types indicate high dsDNA yields, and improved DNA and RNA quality as indicated by commercial qPCR based kits designed to assess DNA quality. To quantitate the quality of the extracted DNA, as a function of sequence coverage across the entire genome, we compared whole genome sequencing data from FFPE samples to sequencing data obtained from matched frozen tissue DNA. Our data clearly indicate significant improvement of sequence coverage and uniformity in FFPE samples processed using Covaris AFA. Our data also indicate that the high quality controllable DNA extraction allows for preparation of large insert (1-6kb) libraries which has previously been considered very difficult to achieve with FFPE stabilized tissues. The simplified single-tube method and rapid workflow allows for the parallel processing of 8 to 96 FFPE samples in a batch format easily adaptable to the throughput requirement of medium to high volume clinical labs. We present data showing highly reproducible extraction and purification results regardless of the throughput of FFPE samples being processed.

2427M

CONNECT1: Next-Generation Sequencing Chip for Dominant Connective Tissue Disorders. *J. Milunsky, A. Nikiforov, K. Wu, X.Y. Zhang, C.T. Baldwin.* Center for Human Genetics, Inc., Cambridge, MA.

Hereditary disorders of connective tissue are fairly common in the general population with Ehlers-Danlos syndrome (EDS), Marfan syndrome, Stickler syndrome, and Loeys-Dietz syndrome being the most well recognized. Significant clinical overlap exists between these disorders making individual sequential Sanger sequencing less cost-effective than simultaneous testing via a next-generation sequencing platform. Aneurysms involving the aorta and other vasculature are typically dominantly inherited and genetic testing is offered for several of the more common genes involved. We have developed and validated a next-generation (Ion Torrent) panel that analyzes 23 dominant genes that cause EDS (types I, II, IV), Marfan syndrome, Loeys-Dietz syndrome (types I-IV), Stickler syndrome (types I-III), and hereditary aneurysms (7 genes). The Ion AmpliSeq software was used to design a custom primer pool to amplify exonic and 25 bases of flanking intronic regions from 23 genes (1274 amplicons). Two Ion Torrent 318 chips were utilized for 15 blinded samples (4 with known mutations in different genes). The average depth across the 23 gene target regions (>40 reads) was 474. All positive samples were accurately identified. Reproducibility was assessed to be 99.2% for 257 SNP's (2 duplicate samples on each chip). 117 amplicons were Sanger sequenced to improve overall coverage to greater than 99% of the targeted regions. Separate MLPA assays were performed for deletion/duplication analyses of those genes reported to have these pathogenic variants. CONNECT1 is the optimal cost-effective first line test for those with suspected dominant undiagnosed hereditary connective tissue disorders or those with a dominant family history of aneurysms.

2428T

Xeroderma Pigmentosum (XP): Single nucleotide variants (SNV) that probably affect function in two Indian families identified using next generation sequencing (NGS) of 8 XP genes. *K. Reddy[1], M. Reddy[2].* 1) Institute of Bioinformatics, Whitefield, Bangalore 560066, India; 2) Ophthalmology, St Johns Medical College, Koramangala, Bangalore 560034, India.

Two Indian families with Xeroderma pigmentosum (XP) were tested using next generation sequencing (NGS) of 8 autosomal recessive XP genes; XPA, XPC, ERCC1, ERCC2, ERCC3, ERCC4, ERCC5 and DDB2. Family 1: had a novel substitution variation in the ERCC5 gene, c.1A>G and a loss of the start site. Each parent had one copy of the gene with this variant and the two affected sons carried variants on both copies of ERCC5. The siblings presented with XP with no apparent neurological abnormality (IQ85 and IQ70 is probably due to impaired vision). In Nature there is also a read through fusion transcript BIVM-ERCC5 [Homo sapiens], an important paralog of ERCC5. The expression of the conjoined genes may be another mechanism for gene regulation in eukaryotes and should be considered in correlating the phenotype with the genotype. Gene sequence and protein profile of ERCC5 and BIVM-ERCC5 may be necessary to assess the functional impact of SNVs in ERCC5 gene Family 2: had a c.1677C>G variation in XPC gene that has been reported in literature and creates a premature stop codon. One copy of the gene in the father, one copy in the mother and both copies in the proband carried this variant. These are the first Indian SNVs described in XP families and they are different from the frequent mutations from other parts of the world. Hence, variants that affect gene function in XP families from all parts of India will be assessed and haplotyped to identify a founder mutation. A focus on the mutations in areas of high consanguinity could potentially simplify the test. Also, understanding the distribution of SNVs in XP gene would aid in counseling families about consanguineous marriage and provide prenatal diagnosis option for carriers. Early eye sight corrections in patients would enhance performance in school. In the long run we hope to decrease the incidence of XP.

2429S

Steroid resistant nephrotic syndrome (SRNS): NGS panel testing to direct therapy and intervention. *L.J. Yarram-Smith[1], A. Bierzynska[2], D. Smith[1], M. Saleem[2,3], M. Williams[1].* 1) Bristol Genetics Laboratory, North Bristol NHS Trust, Southmead Hospital, Bristol.Bristol Genetics Laboratory, Bristol, United Kingdom; 2) Academic Renal Unit, Dorothy Hodgkin Building, Whitson Street, Bristol, BS1 3NY; 3) University of Bristol, Children's Renal Unit, Bristol Royal Hospital for Children, Bristol, BS2 3BJ, UK.

Steroid Resistant Nephrotic syndrome (SRNS) is a disorder of the glomerular filtration barrier. It is characterised by massive proteinuria, hypoalbuminaemia and oedema, and is managed by non-specific heavy immunosuppression. An invasive biopsy is often required and the majority of patients progress to end stage renal failure. SRNS is associated with over thirty genes expressed at the glomerular filtration barrier. Rapid genetic diagnosis is important for therapy and intervention as genetic SRNS is non-responsive to immunosuppression, and has a lower rate of post transplant recurrence. A genetic result enables appropriate family counselling including choice of transplant donor and has the potential to avoid invasive renal biopsy.

Bristol Genetics Laboratory has received >120 worldwide diagnostic referrals (80% paediatric and 20% adult) for NGS clinical panel testing of 37 SRNS genes (Haloplex; MiSeq; open-source pipeline, including CNV analysis using CONTRA). The panel was validated using 24 patients and 440 variants identified in the University of Bristol Academic Renal Unit RADAR study (Nimblegen/exome). Likely pathogenic variants have been identified in 30% of cases, most commonly occurring in NPHS1, NPHS2, WT1, COL4A3 and COL4A4 with a different mutation spectrum in adult and paediatric patients. A further 30% of cases have candidate variants. Variant stratification is challenging with a global service due to the presence of rare variants in under-sequenced populations, this is assisted by detailed phenotyping and new genotype/phenotype correlations are emerging. A pathogenic LMX1B mutation, c.737G>A, was associated with nail-patella-like renal disease, and a patient with a c.287C>T variant presented with thrombocytopaenia and SRNS associated with MYH9-related disease. WT1 variants causing nephrotic syndrome occur in exons 8 and 9, are usually de-novo and associated with early onset disease. A novel exon 9 WT1 variant, c.1169G>A, was identified in both an affected adult and their affected father. Variants in COL4A3, COL4A4 and COL4A5 can cause Alport syndrome and also thin basement membrane nephropathy. Two unrelated patients have variants in more than one collagen gene. Prenatal diagnosis was performed for a Pierson syndrome family with two pathogenic LAMB2 variants, c.[4198_4199delCT];[928T>C]. We present our results illustrated by cases highlighting the clinical benefit of panel testing.

2430M

Validation of a Robust PCR-only Assay for Quantifying Fragile X CGG Repeats. *K.W. Choy[1], C.S.W. Lin[1], K.M. Wong[1], F.M. Lo[2], Y.K.Y. Kwok[1], J.K. Moore[3], S. Wu[3], S.T.S. Lam[2], M. Schermer[3], T.Y. Leung[1].* 1) Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong; 2) Clinical Genetic Service, Department of Health, Hong Kong; 3) PerkinElmer Health Sciences, Inc., 940 Winter Street, Waltham MA 02451, USA.

Sizing of FMR1 trinucleotide repeats requires the use of capillary sequencer instruments, or by a labor Southern blot. To validate a simple, robust PCR assay for quantification of CGG repeats. We validated a new PCR-only method for quantifying Fragile X CGG repeat that utilizes a low-cost capillary electrophoresis instrument and the FragilEase™ reagent kit. Analytical performance was first demonstrated on 12 Coriell reference samples comprising normal through full mutations. Subsequently, a set of 112 archived clinical DNA samples, enriched for premutation and full mutations was analyzed. All samples were amplified successfully. Quantification of repeat numbers was interpreted by the use of standards with known repeats. The repeat numbers from the new assay were concordant with those obtained with the reference method. The intra-assay (CV < 2.5%) and inter-assay imprecision was within 1 CGG repeat. Our result demonstrated this new method is fast, robust that facilitates Fragile X testing in a clinical laboratory.

2431T

Bridging the gap between sequencing gene panels and whole exomes for clinical diagnosis. *S. Abbs[1], H. Martin[1], K. Brugger[1], F. Rodger[2], R. Littleboy[2], I. Delon[1], G. Sagoo[3], S.G. Mehta[1], S-M. Park[1], R. Armstrong[1], G. Woods[1,2], S. Holden[1], M. Kroese[3], R. Sandford[1,2], E. Maher[1,2].* 1) Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom; 2) Dept of Medical Genetics, University of Cambridge, Cambridge, UK; 3) PHG Foundation, Cambridge, UK.

Clinical diagnostic next generation sequencing requires cost effective, efficient and robust methodology. Targeted gene panels are readily available, but significant resources are required to design, optimise and validate each panel, thus restricting the number of panels, number of diseases that can be tested for, and number of patients who can benefit from these tests in an individual laboratory. Whole exome sequencing (WES) overcomes this hurdle, but is considered impractical or too expensive in many health care settings. We have therefore investigated a single test which sequences 4813 clinically relevant genes and offers the best attributes of both panel and WES approaches. We first undertook a technical validation of the Illumina TruSight One sequencing panel by testing 195 samples with mutations in 74 different genes, previously identified by Sanger sequencing. All 188 single nucleotide variants, 20 deletions and 15 insertions were identified correctly. This was followed by sequencing all 4813 genes in 300 patients chronologically referred for diagnostic sequencing of either a large single gene, or a number of genes associated with a genetically heterogeneous condition. Sequencing results were filtered so the analysis focused only on genes determined by Clinical Geneticists to be relevant to the clinical presentation. Among the first 110 newly referred patients, 82 were analyzed for one of 36 different gene panels, ranging from 2-28 genes. 28 patients were analyzed for just single genes. Analysis of results from the remaining 290 samples is ongoing. To date, clear pathogenic mutations were detected in 57% (16/28) of patients analyzed for single genes, and in 28% (23/82) analyzed for gene panels. This allows confirmation of diagnoses for 35% of patients. Mutations were detected in 32 different genes; 8 patients (7.3%) had mutations in 7 genes for which genetic testing is currently not available in UK.

A single sequencing test covering a wider range of genes than currently available by panel or Sanger sequencing has been rapidly validated and introduced into diagnostic service. Sequencing of this clinical exome offers versatility and simplicity, with comparable mutation detection rates to full exome sequencing but at significantly less cost. This test will enable more patients with a wide range of rare genetic diseases to receive rapid molecular confirmation of their diagnosis.

2432S

Implementing an Augmented Clinical Exome and Reference Improvements to Enhance Diagnostic Yield and Discovery. *R. Chen, M. Pratt, D. Church, S. Luo, G. Bartha, J. Harris, S. Garcia, J. Tirsch, M. Clark, C. Haudenschild, N. Leng, J. West.* Personalis, Inc., Menlo Park, Ca.

Clinical exome sequencing is increasingly used for solving diagnostic odyssey cases in children with suspected genetic syndromes and also cancer. The complex process of going from DNA sample to clinical report involves multiple, technologically, scientifically, and medically complex steps; despite early success, significant improvements can be made to increase the overall diagnostic yield of exome sequencing tests. We have developed an augmented exome approach that boosts accuracy and coverage in over 7000 genes that are medically relevant to Mendelian disease, cancer, and pharmacogenomics as referenced by multiple databases and literature sources. The performance of this augmented exome was assessed using three methods: coverage over all exonic bases in our ACE exome assay compared to other standard exomes, comparison against the NIST standard genome, and examination of the structure and coverage of the 7000 medical genes in the latest reference assembly (GRCh38). Of the over 7000 targeted medical genes, our augmented clinical exome completely covers 50% more genes than standard exomes at comparable coverage. Further we demonstrate increased sensitivity and specificity as evaluated against the NIST standard. In evaluating against the reference, over 650 of these medical genes are on 'Fix' patches released as part of incremental reference improvements by the NCBI, suggesting that the structure of these genes were not well described in GRCh37 and need to be addressed as part of clinical informatics pipeline. Over 150 of the patched genes are on existing panels, which may need to be re-assessed in light of new assembly information. An augmented exome sequencing and informatics approach enables clinical grade performance over a large percentage of medically important genes. We report on specific examples of pathogenic variants that would have been missed by standard exomes in clinical cases processed through our testing laboratory.

2433M

Contextualization and recommendation: How doctors and patients discuss whole-genome sequencing results. *J.L. Vassy[1,2], K. Davis[3], J. Oliver Robinson[4], J. Blumenthal-Barby[4], K.D. Christensen[4], R.C. Green[1,5], A.L. McGuire[4], P.A. Ubel[3] for the MedSeq Project.* 1) Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) VA Boston Healthcare System, Boston, MA; 3) Fuqua School of Business and Sanford School of Public Policy, Duke University, Durham, NC; 4) Baylor College of Medicine, Houston, TX; 5) Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA.

Background: There are concerns that non-geneticist physicians are not prepared for the role genomics will play in clinical medicine. It is unknown how they will talk to their patients about whole-genome sequencing (WGS) and its complex results. **Methods**: We have enrolled primary care physicians and cardiologists to participate with their patients in the MedSeq Project: a randomized trial of WGS in clinical care. Physicians undergo brief CME in genomics prior to patient enrollment. Each patient's WGS data are interpreted and reported on a Genome Report, including results for monogenic disease variants, carrier status, complex polygenic traits, and pharmacogenomics. The Genome Report is delivered to the patient's physician, and the two meet for a disclosure visit to discuss its findings. Disclosure discussions are recorded, transcribed, and coded with thematic content analysis to identify emergent themes. **Results**: 16 of 100 planned WGS disclosure visits have occurred to date: 11 in generally healthy primary care patients and 5 in cardiomyopathy patients. We identify 5 major themes. 1) Contextualization: Physicians use additional clinical information, such as age, family history, lack of symptoms, or other test results, to interpret the significance of WGS findings for each patient. 2) Prioritization: Among the many results listed on each patient's Genome Report, physicians tend to emphasize explicitly the 1 or 2 most clinically relevant WGS findings for each patient, identifying "the single biggest thing to come out of this." 3) Limitations of WGS: Some physicians discuss the technological limitations of sequencing and the unknown penetrance of certain variants in cautioning patients against over-interpreting certain WGS results. Some discuss their own inability to interpret certain WGS findings. 4) Misinformation: We have identified a few instances of inaccuracies in physicians' interpretations of WGS results, such as underestimating a patient's risk of having a child with cystic fibrosis. 5) Recommendation: Physicians ultimately synthesize WGS results and all other clinical information to make recommendations for each patient, including genetic testing for family members, additional diagnostic testing, and referrals to subspecialists. **Conclusions:** Physicians and their patients discuss WGS results similarly to other tests in clinical medicine. Concerns about their unpreparedness may be overstated, but efforts to reduce misinformation are warranted.

2434T

Towards highly sensitive diagnostic exome sequencing without the need for confirmations by Sanger sequencing. *K.L.I. van Gassen, M.G. Elferink, P.H.A. van Zon, I.J. Nijman, B. van der Zwaag, J.K. Ploos van Amstel.* Department of Medical Genetics, University Medical Centre Utrecht, Utrecht, Netherlands.

Current best practice genetic diagnostics by massive parallel sequencing includes confirmations of all reported findings by Sanger sequencing. Such confirmations eliminate false positive variant calls and exclude potential sample swaps, but often need the design of new PCR primers. Designing and ordering new PCR primers for confirmations create a laboratory bottleneck, adds substantial capital costs and leads to an increase in test turnaround-time. The proposed workflow utilizes the power of an independent biological replicate without the need for extra target enrichments or sequencing resources and is applicable to most enrichment techniques that support multiplexing. This replicate eliminates the need for Sanger confirmation of most variant calls. Additionally, this workflow should theoretically decrease false positive and false negative variant call rates. Here, we present the details of this workflow. We report on false positive and false negative variant call rates using this approach.

2435S

Biological assays to predict the functional impact of missense mutations: the case of the tumor suppressor gene BRCA1. *G.A. Millot[1,2], P. Thouvenot[1], B. Ben Yamin[1], L. Fourrière[1], C. Houdayer[3,4], D. Stoppa-Lyonnet[3,4], D.E. Goldgar[5], A. Nicolas[1].* 1) UMR3244, institut curie, Paris, Paris, France; 2) Sorbonne Universités, UPMC Univ Paris 06, 4, Paris, France; 3) Service de Génétique Oncologique & INSERM U830, Institut Curie, Paris, France; 4) Université Paris Descartes, Paris, France; 5) Department of Dermatology, University of Utah School of Medicine, Salt Lake City, Utah.

Missense mutations generally come from single nucleotide polymorphisms (SNPs) present in the coding regions of genomes. These mutations trigger the weakest sequence modification in a protein, since a single amino acid is replaced by another one. The resulting effect of this amino acid substitution can range from "no effect" to "complete alteration" of protein function. For this reason, it is difficult to predict the functional consequences of missense mutations. Moreover, the ever-growing amount of exome data will dramatically increase the number of detected missense mutations, and likewise increase the challenge of interpreting their functional impacts. Our work is focused on the tumor suppressor gene BRCA1, whose germline inactivation leads to hereditary breast and ovarian cancers. To date, 668 different germline missense mutations have been identified in BRCA1. However, only a small fraction has been characterized as pathogenic (37) or neutral (103) using classical genetic methods. The 528 additional missense mutations remain unclassified, due to the lack of familial/populational information and the inability to predict their functional impacts. To address this issue, we designed and validated four experimental assays. Using 40 missense mutations, 25 previously classified as pathogenic and 15 previously classified as neutral, by genetic methods, we showed that these assays are efficient in predicting the pathogenicity of BRCA1 missense mutations. We also propose a method to improve the computation of sensitivity and specificity, two parameters critical for functional assays. Altogether, these results contribute to the high-throughput classification of missense mutations, a major challenge for the next decade.

2436M

Complete *APTX* deletion in a patient with ataxia with oculomotor apraxia type 1. *R. van Minkelen[1], M. Guitart[2], C. Escofet[2], G. Yoon[3], P. Elfferich[1], G.M. Bolman[1], R. van der Helm[1], R. van de Graaf[1], A.M.W. van den Ouweland[1].* 1) Department of Clinical Genetics, Erasmus Medical Centre, Rotterdam, Netherlands; 2) Genetic Laboratory, UDIAT-Centre Diagnòstic, Neuropediatrics Unity, Corporació Sanitària Universitària Parc Taulí, Sabadell, Spain; 3) Division of Clinical and Metabolic Genetics, Department of Pediatrics, The Hospital for Sick Children and University of Toronto, Canada.

Background Ataxia with oculomotor apraxia type 1 is an autosomal-recessive neurodegenerative disorder characterized by a childhood onset of slowly progressive cerebellar ataxia, followed by oculomotor apraxia and a severe primary motor peripheral axonal motor neuropathy. Ataxia with oculomotor apraxia type 1 is caused by homozygote or compound heterozygote mutations in *APTX* (chromosome 9p13.3). Case presentation Our patient has a clinical presentation that is typical for ataxia with oculomotor apraxia type 1 with no particularly severe phenotype. Multiplex Ligation-dependent Probe Amplification analysis resulted in the identification of a homozygous deletion of all coding *APTX* exons (3 to 9). SNP array analysis using the Illumina Infinium CytoSNP-850K microarray indicated that the deletion was about 61kb. Based on the SNP array results, the breakpoints were found using direct sequence analysis: c.-5+1225_*44991del67512, p.?. Both parents were heterozygous for the deletion. Homozygous complete *APTX* deletions have been described in literature for two other patients. We obtained a sample from one of these two patients and characterized the deletion (156kb) as c.-23729_*115366del155489, p.?. The more severe phenotype reported for this patient is not observed in our patient. It remains unclear whether the larger size of the deletion (156kb vs 61kb) plans a role in the phenotype; no extra genes are deleted besides the non-coding *APTX* exons 1A and 2. Conclusion Here we described an ataxia with oculomotor apraxia type 1 patient who has a homozygous deletion of the complete coding region of *APTX* instead of homozygosity or compound heterozygosity for *APTX* mutations. We were unable to confirm a more severe phenotype for ataxia with oculomotor apraxia type 1 patients with a complete deletion of *APTX*, however, more research is needed to study the exact breakpoints/sizes of these kind of *APTX* deletions.

2437T

Next generation sequencing of the *HEXA* and *HEXB* genes in African Americans with positive enzymatic carrier screening result for Tay-Sachs disease. *J. Zhang, J. Liao, S. Guha, W. Zheng, L. Edelmann, R. Kornreich, C. Yu.* Mount Sinai Genetic Testing Laboratory, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Tay-Sachs disease (TSD) is an autosomal recessive lysosomal disorder caused by mutations of the *HEXA* gene which results in the deficiency of hexosaminidase A enzyme and devastating neurodegenerative disease. TSD carriers are more prevalent among Ashkenazi Jewish (AJ) individuals with a frequency of ~1:27, and some other genetically isolated populations such as French Canadians in Quebec, Cajuns from Louisiana and the Old Order Amish. The carrier frequency in the general population is ~1:300. Carrier screening for TSD usually includes both an enzymatic assay and a *HEXA* common mutation panel as recommended by American Congress of Obstetricians and Gynecologists (ACOG). It has been noted that the enzyme positive rate in African American (AA) patients is unusually high compared with the carrier frequency in the general population. A retrospective study was performed to analyze the inconclusive rate based on self-reported ethnicity. We found that the inconclusive rate in African Americans (n=173) is 25% compared to 3.6% in AJ (n=1530), 4.5% in Asian (n=313), 4.6% in Caucasian (n=747), and 6.0% in Hispanic (n=341). In addition, 12 AA individuals (1:15) had enzyme results within the carrier range; however, none were positive by targeted mutation analysis. Next-Generation Sequencing of the *HEXA* and *HEXB* genes was performed in nine AA individuals with positive enzyme results either in the inconclusive or carrier range. No pathogenic mutations in *HEXA* or *HEXB* gene were identified; however, two *HEXA* missense SNPs common in AA, c.1195A>G (rs1800430) and c.1306G>A (rs1800431), were present in either the heterozygous or homozygous state in all nine individuals. We hypothesize that these two SNPs are partial pseudodeficiency alleles, which have reduced cleavage activity against the artificial N-acetylgalactosamine substrates used in the in vitro assay, but do not impact the HEXA enzymatic activity in vivo. We are currently testing a larger cohort of AA with normal HEXA activity to demonstrate that these two alleles are associated with the high positive TSD enzyme results in this population.

2438S

Enhanced detection of large indels in diagnostic exome sequencing. *D. Shinde, L. Shahmirzadi, D. El-Khechen, Z. Powis, C. Mroske, D. Salvador, H. Lu, K. Farwell Gonzalez, S. Tang.* Ambry Genetics, Aliso Viejo, CA.

Diagnostic exome sequencing (DES) has been remarkably successful as both a diagnostic and novel gene discovery tool. It represents one of the most comprehensive tests available and requires a robust bioinformatics pipeline for base calling, sequence alignment, variant calling, annotation, filtering, and prioritization. While single nucleotide substitutions are generally efficiently identified in well-covered exonic regions in DES, accurate mapping of indels, especially those larger than 20 nt, is challenging due to complicated gapped alignment and paired-end sequence inference. There are very limited data on the detection limit of indels of different pipelines based on the current exome sequencing platforms. Of 153 positive/likely positive cases in the first 500 unselected DES cases referred to Ambry Genetics, 4 (2.6%) were associated with large indels (> 40 nt) up to 115 nucleotides in size. These large indels were identified by Ambry Variant Analysis (AVA) pipeline and confirmed by Sanger sequencing. In the first family, a maternally inherited heterozygous UBE3A deletion of 89 nt was detected in a pediatric patient with a differential diagnosis of Angelman syndrome (AS). It is presumed that the asymptomatic mother's deletion occurred on her paternally inherited allele, which was inactivated due to imprinting. In a second pediatric patient with suspected arthrogryposis, AVA detected a de novo heterozygous 115 nt deletion in the MYH3 gene. In a pediatric patient with recurrent infections, immunodeficiency and significantly decreased expression of CD127, compound heterozygosity of a splice site mutation and a 58 nt deletion in the RFXANK gene was observed. Lastly, a homozygous 41 nt indel in the PDE6B gene caused by a 47 nt deletion coupled with 6 nt insertion was detected in a 14 year old patient with retinitis pigmentosa consistent with the gene finding. In conclusion, although detection of large indels is inherently difficult for DES, indels larger than 40nt account for 2.6% of the positive cases in our cohort and indels up to 100 nt can be identified by AVA. Our data highlight the importance of an optimized bioinformatics pipeline for the detection of large indels to improve clinical sensitivity and diagnostic yield.

2439M

Efficient diagnostic routing using whole exome sequencing. *M. Weiss, D. van Beek, R. Straver, I. Bakker, H. Meijers-Heijboer, E. Sistermans, Q. Waisfisz.* Clinical Genetics, Genome Diagnostics, VU University Medical Center, Amsterdam, Netherlands.

Whole exome sequencing (WES) can provide a molecular diagnosis in families with an unexplained phenotype. When using this technique in a diagnostic setting it is important to follow a well documented diagnostic routing. Here we present an effective routing and the results of the first 31 families. All candidate families for a diagnostic WES procedure are discussed and selected in a monthly meeting with clinicians, clinical laboratory geneticists and a molecular geneticist. Selection criteria are based on the assumed likelihood of revealing the molecular diagnosis and include clinical phenotype, likely mode(s) of inheritance, available family members, and diagnostic relevance. During this meeting the diagnostic routing is defined which can include: exclusion of other genes in differential diagnosis, exclusion of deletions and duplications (array), revealing the regions of homozygosity (SNP array), and determination of the primary filtering strategy according to the expected inheritance model. Filtering of variants is performed with Cartagenia using four different validated filtering trees: trio analysis de novo, trio analysis recessive/X-linked, multiple affected dominant/mosaic/imprinted, and multiple affected recessive/X-linked. The diagnostic WES procedure was performed for 31 families with different clinical phenotypes. In two families the causative mutation was identified before exome sequencing was initiated. Both cases were from consanguineous families and in each family the candidate gene was identified in a common region of homozygosity. In the remaining 29 families (including 14 trio's, 10 multiple affected, and 5 single cases) WES was performed using different filtering strategies. In 17 families a (probably) pathogenic variant and in 2 families a possible pathogenic variant was detected. Analysis of families with multiple affected family members showed the highest diagnostic yield (73%, 8/11), compared to trio analysis with one patient (43%, 6/14). In two cases the actual mode of inheritance proved to be different from what was originally thought most likely. In 12 families no clinical relevant variant was identified. We have implemented an efficient strategy for using WES in clinical diagnostics. The diagnostic yield is high: 19/31 (61%) families received a probably molecular diagnosis. The diagnostic WES procedure includes a close and structured communication between physicians and the laboratory.

2440T

Performance of in silico sequence conservation tools in predicting the pathogenicity of missense variants in HBOC and Lynch syndrome-associated genes. *I.D. Kerr, K. Moyes, B. Evans, B.C. Burdett, A. van Kan, H. McElroy, P.J. Vail, D.B. Sumampong, N.J. Monteferrante, K.L. Hardman, R.J. Wenstrup, J.M. Eggington.* Myriad Genetic Laboratories, Inc., Salt Lake City, UT.

Non-synonymous missense changes that result in amino acid substitutions in the protein product represent the majority of variants of uncertain clinical significance (VUSs) identified by genetic testing. To analyze whether commonly used in silico tools can accurately characterize the possible disease association of missense mutations, we compared the accuracy of six commonly used algorithms (Align-GVGD, SIFT, PolyPhen-2, MAPP-MMR, SIFT, Grantham Analysis and Condel) using a dataset of 1,118 BRCA1, BRCA2, MLH1, and MSH2 variants previously classified as clinically deleterious or benign by our laboratory's variant classification program. For all algorithms, except Align-GVGD, the false-positive (FP) rate compared to the reference classification was substantially higher than the traditionally accepted threshold for clinical confidence, with a range from 30.6% - 58.5% for BRCA1, 27.1% - 40.1% for BRCA2, 17.9% - 67.9% for MLH1 and 17.1% - 56.1% for MSH2. Although the FP rates using Align-GVGD for all four genes were lower, including values of 2.2% for BRCA1 and 7.9% for BRCA2, the sample size was too small to provide robust analysis due to exclusion of 750 variants that were used to train the algorithm. The high FP rates for Condel, which classifies variants based on a weighted average of scores from five in silico tools, suggests that the use of multiple models is not significantly more accurate than any of the individual models in isolation. The results of this study suggest that none of the commonly used in silico tools achieve the traditionally accepted minimum threshold of specificity for the clinical use of predictive tools.

2441S

Detection of Pathogenic Mutations in Moderate Penetrance Breast Cancer Genes Significantly Increases the Number of Patients Identified as Candidates for Increased Screening. *E. Rosenthal, H. McCoy, K. Moyes, B. Evans, R. Wenstrup.* Myriad Genetic Laboratories, Salt Lake City, UT.

Hypothesis/Purpose: We sought to establish the clinical utility of multigene hereditary cancer panels for the identification of patients who may benefit from interventions to reduce cancer risk, focusing on 3 genes in which pathogenic mutations are estimated to carry a >20% lifetime risk for breast cancer: CHEK2, PALB2 and ATM. This level of risk meets professional society recommendations for initiating breast screening at younger ages and the use of MRI in addition to mammography. We determined the proportion of women identified as candidates for modified screening through genetic testing who would not have been identified with family history. Methods: We used the Claus tables to evaluate the reported family histories of women in whom pathogenic mutations were found in CHEK2, PALB2 and ATM through clinical testing with a 25-gene hereditary cancer panel. We determined the proportion of these women who would have been identified as having a >20% lifetime breast cancer risk based on family history. Results: Among 9201 patients tested between 09/04/2013 and 04/17/2014, 174 female patients were identified with a single mutation in either: CHEK2 (n=71), PALB2 (n=39) and ATM (n=64). Three patients carried mutations in 2 of the genes. Excluding 5 women who also carried a mutation in BRCA1 or BRCA2, only 15 (8.7%) of the remaining 172 women reached the >20% threshold for lifetime breast cancer risk using the Claus tables. By comparison, among this same group of 9201 tested individuals, 275 had a pathogenic mutation in BRCA1 or BRCA2. Therefore, inclusion of the 3 moderate penetrance genes improved the sensitivity of the testing, as defined by the identification of a genetic finding associated with an established breast cancer screening recommendation, by 62.5%. Conclusions: Clinical testing including the moderate penetrance genes CHEK2, PALB2 and ATM significantly increases the likelihood of identifying women who can benefit from modified medical management strategies that would not have been applied based on family history alone.

2442M

Update on evaluation of ACMG Recommended Incidental Findings in Clinical Whole Exome Sequencing. *P. Vitazka, J. Neidrich, J. Scuffins, B. Friedman, J. Tahiliani, H. Hanson Pierce, D. Copenheaver, N. Smaoui, J. Juusola, G. Richard, S. Bale, E. Haverfield.* GeneDx, Gaithersburg, MD.

In March, 2013 The American College of Medical Genetics and Genomics (ACMG) issued recommendations to include reporting of incidental findings (IF) on all individuals who receive clinical whole exome sequencing (WES) or whole genome sequencing (WGS). These recommendations advised actively reviewing 56 genes primarily associated with an increased risk for cancer or cardiac disease for the presence of known pathogenic (KP) or expected pathogenic (EP) mutations, since early intervention is likely to reduce or prevent accrued morbidity or early mortality. The ACMG estimated that ~1% of patients undergoing WES will have reportable incidental findings. At GeneDx, IF analysis has been offered since July 1, 2013. Since then, 2148 unrelated probands underwent WES, 433 (20.2%) of whom opted out of receiving IF. The recommended 56 genes were assessed for KP and EP mutations present at ≤1% frequency in the 1000 Genomes database or our internal GeneDx annotation database, excluding previously identified polymorphisms and large copy number mutations. All variants were evaluated for evidence of pathogenicity in the literature (HGMD database), and consistency with the ACMG reportable mutation spectrum for the gene. Of the 975 completed tests, we have reported 86 IF in 81 (8.3%) cases. Of the 86 pathogenic variants, 61 (70.9%) were in cardiac-related genes, 9 (10.5%) were in cancer-related genes, and 16 (18.6%) mutations were identified in genes associated with either hypercholesterolemia or malignant hyperthermia. IF were most commonly reported in MYBPC3 (19 mutations), SCN5A (10), KCNQ1 (8), and KCNH2 (7). No reportable variant was found in 34 (61%) of the 56 genes recommended by ACMG. Although previous pilot studies indicated that probands undergoing WES may have a high number of reportable ACMG IF (16-24%), these frequencies have not been replicated in our current study. Based on a much larger dataset, we have now found that the frequency of reportable ACMG IF is ~7-9% of cases undergoing WES, which is ~7-9x higher than the initial estimate provided by ACMG. This number is likely more accurate estimate of IF but may increase slightly in the future as more variants are evaluated for pathogenicity and the literature continues to expand.

2443T

Targeted gene panel sequencing using multi-parallel single-plex PCR amplification for the detection of somatic mutations. *K. Yap, K. Kiyotani, K. Tamura, M. Montoya, Y. Nakamura.* Department of Medicine, Section of Hematology/Oncology. The University of Chicago, Chicago, IL, USA.

The use of whole exome sequencing has been instrumental in the discovery of novel variants and somatic mutations that may be involved in the development of human diseases. Generally, non-synonymous variants which give rise to protein coding changes or protein loss are more likely to be the culprit, indicating the positive utility of exome sequencing. However, the broad target regions of whole exome sequencing usually dictates the sacrifice of overall sequencing read coverage, reducing the ability to detect low frequency variants, which is especially relevant for somatic mutation detection in cancer. In this study, we selected 10 bladder cancer samples that were previously whole exome sequenced, for deeper sequencing using a targeted gene panel of 60 cancer genes (1070 amplicons) developed on the Wafergen Smartchip TE system. Using a multi parallel single plex PCR-based approach, the design avoids the undetectable fall-outs and primer inefficiencies that typically occur in a multiplex primer pool PCR reaction. Additionally the chip-based design improves the throughput of the number of multi-parallel PCR reactions that is possible in a plate-based design. We found that all the variants (33 somatic mutations) that were called on exome sequencing can be confirmed by this targeted gene panel. Furthermore, 10 somatic mutations were additionally detected by the targeted gene panel due to the significantly improved overall read coverage. Conversely, there were a number of false positive variants that arose in this targeted gene sequencing approach due to the PCR-based amplification steps. This issue can be overcome by conservative filtering criteria and disregarding variants that were supported by sequencing reads with strong strand biases. Overall, targeted gene panel sequencing will significantly reduce sequencing cost and increase coverage in target gene regions of interest. This will allow for the detection of low frequency somatic mutations in heterogeneous cancer samples, which is instrumental for disease monitoring. However, additional bioinformatics processing will need to be carried out on data sets derived from targeted gene panels to account for the erroneous variants that may arise due to PCR amplifications.

2444S

Clinical and technical evaluation of NGS based testing for hereditary cancer syndromes. *S. Lincoln[1], A. Kurian[2], A. Desmond[3], M. Gabree[3], Y. Kobayashi[1], M. Anderson[1], G. Nilsen[1], S. Yang[1], R. Hart[1], K. Jacobs[1], K. Shannon[3], L. Ellisen[3], J. Ford[2].* 1) InVitae, San Francisco, CA; 2) Stanford University School of Medicine, Stanford, CA; 3) Massachusetts General Hospital, Boston, MA.

Introduction: Next-generation sequencing (NGS) of gene panels and exomes is gaining clinical acceptance, although questions remain about the sensitivity, specificity and clinical implications of these tests. Expanding on our recently published work (Kurian et al, J Clin Oncol, 2014) we considered whether NGS can both replace and supplement traditional *BRCA1/BRCA2* testing in patients indicated for hereditary breast/ovarian cancer testing. **Methods:** Over 900 patients indicated for *BRCA1/BRCA2* testing under clinical management guidelines were recruited and tested with a 220-gene NGS panel. In this analysis we focused on both sequence variants and copy-number variants (CNVs) detected by NGS in 29 moderate to high penetrance cancer risk genes. These patients also had traditional genetic testing results (e.g. from Sanger sequencing and QPCR) available for comparison. **Results:** 261 alterations (196 pathogenic and 65 others) were reported in the traditional data, and all were detected by NGS when the corresponding test was ordered. In this set are 141 insertions, deletions, complex events and CNVs, with particularly technically challenging variants including indels up to 126bp and CNVs as small as one exon. Specificity was also high: all NGS variants for which we sought orthogonal confirmation (n>2000) were confirmed, including 42 pathogenic variants not previously reported. No single calling algorithm achieved this performance but rather a combination did (GATK, Freebayes, PolyMNP, CNVitae and split-read detection). By testing for genes beyond BRCA1/2, diagnostic yield increased over 50%. As expected, clinical actionability of these findings varied, with various changes in care recommended when we were able to recontact patients. Communication of this information proved clinically feasible and was appreciated by the patients. **Conclusions:** NGS can be a viable replacement for traditional genetic testing techniques for hereditary cancers and other syndromes, with the additional benefit of cost-effectively increasing diagnostic yield. Orthogonal confirmation of clinical NGS results remains a strongly recommended practice, although the high concordance of traditional and NGS data suggests that the cost-benefit of confirmation merits careful consideration over time. **Note:** All of the variants in this study and their interpretations will be released to public databases by the time of the meeting.

2445M

Beyond BRCA1 and BRCA2: results from screening 94 genes in 200 patients with familial breast and ovarian cancer using panel sequencing and custom array-CGH. *A. Rump[1], E. Schrock[1], K. Hackmann[1], L. Mackenroth[1], A. Beyer[1], D. Huenich[3], U. Markwardt[3], N. DiDonato[1], B. Klink[1], P. Wimberger[2], K. Kast[2].* 1) 1Institut für Klinische Genetik, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany; 2) 2Abteilung für Gynäkologie und Geburtshilfe, Universitätsklinikum Carl Gustav Carus, Technische Universität Dresden, Germany; 3) 3Zentrum für Informationsdienste und Hochleistungsrechnen (ZIH), Technische Universität Dresden, Germany.

Background Breast and ovarian cancer (BC/OC) predisposition has been associated with a number of high- and low-penetrance susceptibility genes. With the advent of NGS-based panel sequencing, testing all these genes became feasible in addition to the analysis of BRCA1, BRCA2, RAD51C and CHEK2. Here we report on the results of custom array-CGH for deletion/duplication analysis and panel-based screening of 94 genes that have been associated with hereditary cancer predisposition. **Method** Selection criteria for the 200 patients included in this study were defined by the German Consortium for Breast and Ovarian Cancer. High risk patients with previously excluded mutations in BRCA1 and BRCA2 were also included. Target enrichment was performed with the Illumina TruSight cancer panel, which includes 94 genes associated with both common (e.g., breast, colorectal) and rare cancers. All 94 genes were additionally analyzed with a customized, high resolution array CGH. **Results** In 28 % of the patients, BRCA1 and BRCA2 variations have been found. These were either clearly pathogenic protein truncating mutations or rare, unclassified missense variations with high probability of effect. In 30 % of the patients we found nonsense-mutations or unclassified missense variants in low penetrance susceptibility genes, especially NBN, CDH1, ATM and PALB2. TP53 mutations were revealed in 2 % of our cases. Interestingly, one of these TP53 mutations was found as a 10 % mosaic in blood cells from a patient with no familial history of cancer but late onset of both BC and OC. The frequency of the TP53 mutation in these tumors is currently being investigated. In one patient with a familial history of both BC and colon cancer a pathogenic mutation in MLH1 could be identified, along with a frame-shift mutation in BRCA1. Complementary custom array-CGH in all 200 patients identified deletions in ATM, BRCA1, CHEK2 or RAD51C in 3 % of the cohort. Despite comprehensive testing, 37 % of the patients did not reveal any convincing mutation, neither on nucleotide level nor on genomic level. **Conclusion** The extension of mutation screening beyond BRCA1 and BRCA2 reveals disease-causing mutations in high-penetrance genes, like TP53, as well as mutations in low-penetrance susceptibility genes, especially CDH1, ATM and NBN. However, the enormous number of unclassified sequence variants and the detection of mutations in “non-breast-cancer” genes pose a huge challenge for genetic counselling.

2446T

Reporting Candidate Genes: Identifying a Potential Genotype-Phenotype Correlation via Whole Exome Sequencing. *J. Neidich, P. Vitazka, J. Tahiliani, K. Retterer, J. Juusola, S. Suchy, D. McKnight, R. Bai, N. Smaoui, G. Richard, S. Bale, E. Haverfield.* GeneDx, Inc., Gaithersburg, MD.

Whole Exome Sequencing (WES) has become an effective diagnostic tool for the identification of the molecular basis of human genetic disorders. WES allows the investigation of a patient's clinical phenotype through the evaluation of specific variants found by the simultaneous analysis of approximately 20,000 genes in the human genome. The process begins with assessing the over 6000 genes already associated with human diseases. Resources including the Human Gene Mutation Database (HGMD), population frequency databases, gene-specific databases, and in silico pathogenicity prediction tools can be used to evaluate variants. New potential genetic etiologies not yet linked with human genetic diseases (candidate genes) may also be interrogated. The clinical utility and research potential of reporting out variants in candidate genes has yet to be established. During the initial 28 months of our WES program, approximately 7% (157/2242) of the reported variants have been in candidate genes. All reported candidate genes had animal models or pathway, expression, or functional studies tying the phenotype to the gene. Findings reported in candidate genes were retrospectively reviewed to determine whether any new information linking these genes to a genetic disorder became available after the initial report. These candidate genes were crosschecked against HGMD to assess how many of these genes had been newly added to the database since the report date. Of these, 3% (13/431 newly added genes) have since been added to HGMD as genetic causes of human disease. The associated disorders include neuropathies, intellectual disability, kidney, cardiac, and metabolic disorders, and congenital anomalies, including well-described syndromes such as Noonan, Prader-Willi, and Joubert syndromes. This result suggests there is value to reporting findings in candidate genes. Reevaluation of candidate genes on a periodic basis may be helpful in clarifying the cause of a phenotype with unknown genetic etiology. With this additional information, patient counseling should emphasize that genetic information is improving over time as a component of the benefits and limitations of WES. Ongoing review of already reported results is important to maximize the benefits of the existing technology and to ensure appropriate genomic diagnosis is provided.

2447S

Diagnostic sequencing in integrated clinical and research laboratory setting for 100 families at the Dorrance Center for Rare Childhood Disorders. *S. Szelinger[1,2,3], K.M. Ramsey[1,2], A.L. Siniard[1,2], I. Schrauwen[1,4], J.J. Corneveaux[1,2], A.A. Kurdoglu[1,2], J. Krate[1,5], B.E. Hjelm[1,2], I. Malenica[1,2], R. Richholt[1,2], M. de Both[1,2], S. Rangasamy[1,2], M.J. Huentelman[1,2], V. Narayanan[1,2], D.W. Craig[1,2].* 1) Neurogenomics Division, Translational Genomics Research Institute, Phoenix AZ, USA; 2) Dorrance Center for Rare Childhood Disorders, Translational Genomics Research Institute, Phoenix AZ, USA; 3) Molecular and Cellular Biology Interdisciplinary Graduate Program, School of Life Sciences, Arizona State University, Tempe AZ, USA; 4) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; 5) University of Arizona College of Medicine-Phoenix, Phoenix, AZ, USA.

Integration of a clinical diagnostic center and a genomic research lab provides a unique opportunity to improve the life of children with rare, undiagnosed genetic disorders. To date, we enrolled 225 families and sequenced the whole-exome, whole-genome of 150. Study design included sequencing whole-exome of family trios, constant re-analysis of genomic data with updated annotation and prioritization methods, application of mRNAseq from whole blood for challenging cases and for validation of candidate causal variants, and the establishment of patient specific fibroblast cell lines for functional studies. We report a concise summary of the first 100 families enrolled. Enrollment criteria included a previously undiagnosed condition, ambiguous genetic origin, and negative, or inconclusive genetic tests. Most patients exhibited some form of neurological phenotype and were characterized as one of following condition: Neurodegeneration, Epilepsy, Movement disorder, Mitochondrial disorder, Aicardi Syndrome, Aicardi Goutieres Syndrome, Autonomic Dysfunction, Demyelinating disease, Cerebral Palsy, Leukodystrophy, GI dysmotility, Hearing loss, and Neuromuscular dysfunction. We applied Illumina TruSeq Exome Enrichment capture method, HiSeq2000 platform with a 2×100bp paired-end sequencing set up as standard. In summary, we sequenced 60 family trios, 15 parent-proband duo, 9 singletons, and 16 families with multiple affected and/or siblings. In 16 families, whole genome sequence of the proband and/or the parents were also obtained. In 28 families Illumina TruSeq RNA sequencing was additionally performed, and from 11 families, we obtained fibroblast tissue for cell culture. After filtration and prioritization, we categorized variants by their likelihood as: 1. Presumed Causal 2. Likely Causal/Major Contributor, 3. Candidate, 4. Unknown. In approximately 35% of cases a presumed causal or likely causal variant was found, and in about 9% of cases, we identified a candidate variant. Post diagnosis, we helped facilitate enrollment a number of diagnosed families in NIH funded clinical trials, or connected them with experts in the field, or provided alternative treatment options. We developed mouse models, and large-scale drug screens with previously approved FDA drug compounds for selected causal genes. Continuous outreach to the families and close integration of diagnostic and research approach was able to provide conclusion to the diagnostic odyssey for many families.

2448M

An intronic deleted mutation in the COL3A1 gene affecting exon skipping causing vascular Ehlres-Danlos syndrome. *A. Watanabe[1, 2], B.T. Naing[1, 3], K. Akutsu[4], T. Shimada[1, 2], T. Okada[1].* 1) Dept Biochem & Molec Biol, Nippon Med Sch, Tokyo, Japan; 2) Div Clinical Genetics, Nippon Med Sch Hosp, Tokyo, JAPAN; 3) Dept Molec Med & Anatomy, Nippon Med Sch, Tokyo, Japan; 4) Div Intensive and Cardiac Care Unit, Nippon Med Sch Hosp, Tokyo, Japan.

Vascular type of Ehlers-Danlos syndrome (vEDS), also known as EDS type IV (MIM#130050), is a life-threatening autosomal dominant inherited connective tissue disorder mainly caused by mutations in type III collagen, *COL3A1* gene. The main types of mutations in vEDS are a single amino acid substitutions for glycine in the GLY-X-Y repeat of the triple helical region or invariant splice sites. Recently, premature translation termination mutations are reported affecting nonsense-mediated mRNA decay of mutated allele. Here, we described a 26 y/o female with clinical features and family history of vEDS and characterized the functional consequences of an intronic *COL3A1* deleted mutation. The coding regions and flanking intronic bases of the *COL3A1* gene were screened by high resolution melting curve analysis method using genomic DNA. The patient is a heterozygote for a four bp deletion at 5' splice site of intron 8 with in genomic DNA (c.636+5+4del GTAA/IV8) that do not affect the consensus GT or AG splice sites and changed from G to C transversion at Position +5 of the splice donor site in the intron. A minigene splicing assay demonstrated that this deletion is sufficient to cause a skipping of exon 8. G at Position +5 of the splice donor site is supportive of an importance for splicing of *COL3A1*. Intronic mutations may be responsible for vEDS in some families with otherwise negative mutation screening of *COL3A1*.

2449T

The Relationship between Blood Index and Thalassemia Disease. *N. Ghazavi, M. Behnam, F. Hosseini, A. Hejazifar, A.R. Salehi Chaleshtori, M. Salehi, N. Mansouri.* Medical Genetic Center of Genome, Isfahan, Iran.

Background and Aims: Thalassemia is a severe hemoglobin disease, recognizing by CBC and hemoglobin electrophoresis primarily. It is possible that CBC and hemoglobin electrophoresis were performed for the index case who recognized as α-Thalassemia but finally experimental studies clearly demonstrated that it was different type of thalassemia. In the present study, we report a case of β-thalassemia mutations which is called -101 C>T relative to the transcription start site of β-globin gene, and is a β+-Thalassemia case, but it has been observed with different indices. Methods: The patient was admitted based on hematologic indices as α-Thalassemia, and diagnostic tests for α-Thalassemia, including GAP PCR, ARMS-PCR for α-Thalassemia were performed and all examinations were normal. For increased confidence, sequencing for patient's β-globins' gene was performed. We also observe more cases with the same index. Results: Our findings, present the incidence of mentioned mutation in β-globin gene in β+ patient, in contrast to the hematologic indices for β-Thalassemia heterozygote. Moreover, this case has no resemblance to other β+-Thalassemia cases and it can be mistaken as α-Thalassemia. (For example: The - 88 C> A mutation relative to the transcription start site, which several of our indices present beta-Thalassemia). Conclusion: As mentioned above, our outcomes are extremely similar to the results of CBC and hemoglobin electrophoresis of α-Thalassemia, but the patient was β-Thalassemia. Such conditions as cited above are hazardous and end to affected neonates if the other partner is β-Thalassemia patient. In this condition, Sequencing of β-globins for one of the couples could be helpful.

2450S

Investigation of Genomic Deletions and Duplications by Custom MLPA in a Cohort of 338 Patients with Obesity, Developmental Delay, Hyperphagia, and Additional Features. *C.S. D'Angelo[1], C.V. Monica[1], I.E.C. Claudia[1], A.K. Chong[2], R.B. Debora[2], M.L. Charles[3], P.K. Celia[1].* 1) Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of Sao Paulo, Sao Paulo, Brazil; 2) Department of Pediatrics, Children Institute, School of Medicine, University of Sao Paulo; 3) Department of Medical Genetics, School of Medicine, University of Sao Paulo.

Obesity is highly heritable and a significant health problem for people with intellectual disability (ID). The recent implementation of whole-genome chromosomal microarray analysis (CMA) has resulted in the discovery of novel and rare CNVs implicated as risk factors for obesity, often also present with ID. However, the strong phenotypic overlap between syndromic forms of obesity poses challenges to accurate diagnosis, and many different individual cytogenetic and molecular approaches may be required. Multiplex dependent-probe amplification (MLPA) enables the simultaneous analysis of multiple targeted loci in a single test, and serves as an important investigation tool in situations where specific microdeletion and/or microduplication syndromes are suspected. Our aim was to design a synthetic probe set to screen for deletions and duplications at previously described loci associated with obesity in a cohort of 338 patients with syndromic obesity of unknown etiology who have had a normal methylation test for PWS or in whom no 1p36 deletion could be found. Eighteen alterations were detected using the synthetic MLPA probe set. Ten of the detected alterations were delineated by chromosomal microarray, while the remaining alterations were fine-mapped by additional MLPA probes incorporated into commercial kits. We identified nine known CNVs representing clinically well defined microdeletion syndromes with obesity as a clinical feature: in 2q37 (4 cases), 9q34 (1 case) and 17p11.2 (4 cases). We also identified three deletions and a reciprocal duplication in the DiGeorge syndrome locus at 22q11.2. Additionally, we have identified two recurrent small deletions within the 22q11.2 'distal' locus associated with a variable clinical phenotype and obesity in some individuals. Three individuals were found to have a recurrent CNV of one susceptibility loci at 1q21.1 'distal', 16p11.2 'distal', and 16p11.2 'proximal'. The overall detection rate with the synthetic MLPA probe set was about 5.3% (18 out of 338). Our experience leads us to suggest that MLPA could serve as an effective alternative first line screening test to chromosomal microarrays for syndromic obesity, allowing for a number of microdeletion syndromes loci (e.g. 1p36, 2p25, 2q37, 6q16, 9q34, 11p14, 16p11.2, and 17p11.2) known to be clinically relevant for this patient population to be interrogated simultaneously. Financial Support: CEPID-FAPESP, CNPq.

2451M

Prenatal detection of chromosomal aberrations and its reflection at adult age. *B.B. Ganguly[1], S. Mandal[1], N.N. Kadam[2].* 1) MGM Center for Genetic Research & Diagnosis, MGM Institute of Health Sciences, Mumbai, India; 2) Department of Pediatrics, MGM Institute of Health Sciences.

Heritable chromosomal abnormalities have been reported in ~50% spontaneous abortion and ~5-10% of live pregnancies with prevalence of trisomy 21. Intervention of prenatal diagnosis is generally considered in suspected cases with high risk triple screen, significant family history and/or advanced maternal age. In our previous data, Down syndrome was recorded in ~4% cases, while balanced translocation and inversion were present in 5% among 140 pregnancies. The cases with balanced rearrangements might achieve a long life span since such alterations do not contribute to major phenotypic or clinical manifestation. Therefore, chromosome aberrations are expected in ~10% of children including 5% of adults; however, these aberrations can be prevented from transmission through generations with a straightforward approach of prenatal chromosomal analysis by conventional techniques. In the present report, we present data on chromosomal status in 269 fetal samples of live pregnancies. CVS or amniotic fluid was processed for FISH and/or conventional analysis of karyotypic status. Interphase cells were considered for FISH whereas metaphases were processed for GTG-banding and karyotyping. Mitotic metaphases were obtained from long-term culture of the specimens following standard technique. IKAROS and ISIS software (MetaSystems, Germany) was utilized for conventional karyotyping and FISH analysis. The cases were categorized into three age groups of 10 years interval and the abnormalities detected by FISH and G-banding was presented in different age groups. The incidence abnormality was maximum in the advanced age group as expected; however, the youngest age group showed higher number of structural aberrations over group II. Conventional G-banding result appeared with 71.5% structural aberrations, which was not detectable by FISH. The present report highlights the importance of karyotyping over interphase FISH. Also the study recommends the intervention of karyotyping at fetal age with a view to prevention of transmission and reduction of chromosome aberrations among adults which would further reduce the burden of birth defects and/or reproductive failure.

2452T

45,X (30%); 46,X,i(Xq) (60%) mosaicism. Case report. *M. Pérez Sánchez[1, 3], A. Mora Guijosa.[1, 3], M. Lopez Melchor.[1, 3], A.R. González Ramírez.[1, 2].* 1) UGC Análisis Clínicos, Hospital Virgen de las Nieves, Granada, Spain; 2) Fibao. Hospital Clínico San Cecilio, Granada, Spain; 3) Instituto de Investigación Biosanitaria de Granada, Granada, Spain.

The commonest structural rearrament of the X is 46,X,i(Xq). An i(Xq) chromosome theoretically consists of wo copies of the entire long arm and no short arm. The gross banding patterns of most isocromosomes fits with this interpretation but in detail the situation is more complicated. Some isocromosomes have one centromere and are perfectly symmetrical, while others are symmetrical but have two centromeres close to each other, and still others have a clearly asymmetrical central portion, raising the posibility of a more complex abnormality. On the other hand, Turner mosaicism are present in an important group of Turner syndrome with a wide clinical features ranging from full-blown Turner syndrome to normal phenotype. The karyotype is the elecction technique for these cases, but for a finest detection of the rearrament point, the microarrays CGH will be of elecction. Here we present a case of a 19 years old woman that was referred for genetics studies. The clinical findings were amenorrhea with treatment response, normal secundary sexual caracteristics, hormonal status at normal levels, the right ovary was not detected and the left one was normal when magnetic resonance studies were done. No appreciable mental retardation were detected. Chromosome culture and karyotyping were realized by standard thecnics and showed karyotype of 45,X (30%); 46,X,i(Xq) (60%) mosaicism. To determine the Xp region deleted and the Xq region duplicated, GH-array 180 K (PerkinElmer Platform) was performed, with a resul of total loss of 55,17 Mb in the Xp22.33-p11.21 region (genomic coordinates: chrX296520-55466477), a partial duplication of 2,77 Mb in the Xp11.21-p11.1 (genomic coordinates: 55556815-58324786) and partial duplication of the entire Xp arm. The results obtained with CGH-Aray are in concordance with the karyotype results, but with the possibility of detec the exact rearragement point. The presence of the Xp región near to the centromere can explain the mild Turner clinical features in this patient. As a conclusión, when a rearrangement is detected by karyotype, the CGH-Array can be indicated to detect the exact rearrangement point.

2453S

Comparison of whole genome amplification (WGA) methods for detection of genomic aberrations using low coverage massive parallel sequencing. *D. Deforce[1], L. Deleye[1], D. De Coninck[1], C. Christodoulou[3], A. Dheedene[2], T. Sante[2], B. Heindryckx[3], P. De Sutter[3], B. Menten[2], F. Van Nieuwerburgh[1].* 1) Lab of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium; 2) Center for Medical Genetics, Ghent University, Ghent, Belgium; 3) Department of Reproductive Medicine, Ghent University, Ghent, Belgium.

Whole genome amplification (WGA) is currently performed on biopsied blastocyst trophectoderm cells to subsequently perform aCGH to screen for large chromosomal aberrations in a Pre-implantation Genetic Diagnosis (PGD) setting. The different currently available WGA methods lead to amplification bias resulting in over- and under-represented regions in the genome. Current WGA methods, such as PicoPLEX (Rubicon Genomics) and subsequent aCGH analysis, make it possible to detect duplications and deletions at a 10 Mb resolution.

An alternative to aCGH is low coverage Massive Parallel Sequencing (MPS). Genomic regions for which a significantly higher or lower number of sequences are generated can be called as duplications and deletions. Unlike with aCGH, there is no reference DNA used within the MPS assay that can be used for a within-assay normalization of the representation bias. WGA representation bias might thus be more critical when using MPS.

Recently, a new Multiple Annealing and Looping Based Amplification Cycles (MALBAC) WGA method has been published, claiming unparalleled performance. This method has not yet been studied in a PGD setting. The goal of this study was to compare the well established PicoPLEX and MALBAC WGA for their ability to produce optimal PGD by MPS results from a limited number of cells.

Six repeats of samples consisting of 1 cell, 3 cells and 5 cells were collected from the Loucy lymphoblastoid cell line using micromanipulation techniques. Three repeats were amplified using PicoPLEX (Rubicon Genomics), and three using MALBAC. The amplification products were used to create Illumina sequencing libraries and were sequenced on an Illumina HiSeq1500. Results show a comparable performance between the PicoPLEX and the MALBAC WGA. By subsampeling the generated sequences, we also were able to determineminimal coverage to achieve a 10 MB resolution.

2454M

Supporting the free exchange of clinical laboratory variant data through VariantWire. *S. Baxter[1,2], S. Aronson[2], L. Edelmann[3], M. Lebo[1,4], J. Lerner-Ellis[5,6,7], E. Lyon[8,9], M. Oates[2], H. Rehm[1,4,10], VariantWire Consortium.* 1) Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, MA; 2) Partners HealthCare Personalized Medicine, Cambridge, MA; 3) Mount Sinai Genetic Testing Laboratory, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, NY; 4) Department of Pathology, Brigham & Women's Hospital, Boston, MA; 5) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; 6) Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, University of Toronto, Toronto, ON, Canada; 7) Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 8) Department of Pathology, University of Utah School of Medicine, UT; 9) ARUP Laboratories, Salt Lake City, UT; 10) Pathology, Harvard Medical School, Boston, MA.

With the emergence of next generation sequencing technology, the quantity of sequencing data being produced by genetics laboratories is exponentially increasing. In the last decade, not only has the volume of clinical genetic tests been on the rise, but the average number of genes covered by those tests has also increased dramatically. The expanding scope of genetic tests is leading to more data being generated and an increased rate of variants of unknown significance being returned to clinicians. Siloed data sets restrict our understanding of the spectrum of variation for a particular disease and result in higher rates of variants of uncertain significance. Data sharing is proving to be an essential element for overcoming these challenges of the genomic era. VariantWire, a consortium supported by GeneInsight, provides a data-sharing network that allows clinical genetic testing labs to share variant and gene interpretations, and the evidence behind those interpretations, in real time. Data sharing on VariantWire occurs in a secure environment and all case information in the network is de-identified. VariantWire is governed by the VariantWire Committee, which is comprised of one member from each participating entity (listed in author affiliations). The VariantWire committee guides the content of the data sharing policy and reviews all applications from laboratories seeking to join the network. As of May 21st 2014, VariantWire is sharing 23,318 variants across 336 genes. Of the shared variants 2,973 are classified as clinically significant (pathogenic or likely pathogenic) for over 100 diseases. Another 3,907 variants are classified as unknown significance and 16,438 were determined to be either likely benign or benign. Finally, 24 somatic variants in the EGFR gene are involved with drug response. Two hundred and seventeen variants have been identified in more than one lab, with 61 of those variants being classified as either pathogenic or uncertain significance. Of those 61 clinically significant variants seen more than once, only two variants had inconsistent classifications across labs. By enabling real-time data-sharing, labs using VariantWire are assured that they have the most up-to-date knowledge from other participating labs. Finally, the GeneInsight team aids in helping labs submit their variant data to ClinVar, a freely available database, ensuring the broadest sharing of data to further our clinical understanding of the human genome.

2455T

Developing Exclusion Datasets for Genome Filtering in the MedSeq Project. *K. Machini[1,2], R. Shakhbatyan[1], H. McLaughlin[1,2], O. Ceyhan Birsoy[1,2], D. Metterville[1], M. Lebo[1,2], R.C. Green[1,2,3], H.L. Rehm[1,2].* 1) Partners Healthcare Personalized Medicine, Cambridge, MA; 2) Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 3) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

A major obstacle in the clinical application of whole genome sequencing (WGS) is the difficulty in systematically interpreting the enormous amount of data generated. This includes analyzing a genome for the return of incidental findings based upon the interpretation of previously published variants as well as the interpretation of novel loss of function variants in disease-associated genes. Within the framework of the MedSeq Project, 100 participants, half healthy and half with cardiomyopathy, are undergoing WGS. Sequencing is performed at Illumina's CLIA-certified laboratory and alignment, variant calling, annotation and variant review are performed at the Partners' Laboratory for Molecular Medicine. First, annotated variants with a minor allele frequency (MAF) <5% are filtered to identify those classified as DM or DM? in the Human Gene Mutation Database (HGMD). A separate filter identifies nonsense, frameshift, and canonical splice-site (+/-1,2) variants with a MAF <1% from a list of disease-associated genes. Although this is a common strategy employed by laboratories performing incidental findings analysis, it is hindered by the plethora of variants with false claims of pathogenicity or in genes with insufficient evidence for disease causality. After full interpretation of the first 27 genomes, involving review of 1022 rare variants with claims of pathogenicity, only 74 variants (average of 2.7 per genome) reached the evidence level to be clinically reported. In the course of excluding the majority of variants, we have developed a series of filters. First, 19% of variants can now be excluded using a dataset of common variants encountered in >10% of individuals in our cohort. Second, 14% can be excluded using quality metrics derived from analyzing Sanger confirmation data to improve quality thresholds. Third, 27% of variants can now be excluded using a dataset of genes with claimed medical associations that have insufficient evidence. Finally, 15% of reported pathogenic variants can now be excluded because of evidence for a benign interpretation. This has allowed a dramatic reduction in genome review time which will continue to be reduced as further genomes are interpreted and these exclusion datasets are continually populated and queried during our review process. We propose that laboratories share their exclusion datasets and approaches to enable a community resource for collectively improving the process and efficiency of genome interpretation.

2456S

Diagnostic Exome Sequencing provides diagnoses among patients with abnormal brain MRI findings. *K.D. Farwell Gonzalez[1], L. Shahmirzadi[1], D. El-Khechen[1], Z. Powis[1], K. Radke[1], C. Mroske[1], D. Shinde[1], S.K. Gandomi[1], C. Alamillo[1], E.C. Chao[1,2], R.M. Baxter[1], B. Tippin Davis[1], S. Tang[1].* 1) Ambry Genetics, Aliso Viejo, CA., USA; 2) Department of Pediatrics, University of California, Irvine, Irvine, CA 92697.

Diagnostic exome sequencing (DES) is successful in solving the diagnostic odyssey for 30-40% of undiagnosed patients with a broad range of underlying Mendelian disorders. Among the first 500 reported DES families, 168 (34%) had a previous abnormal brain MRI. Positive findings were uncovered in 66 of these families (39%). Positive and likely positive alterations within clinically characterized genes were identified in 55 of the 168 families, for a molecular diagnostic rate of 33% among characterized genes. A novel gene finding was uncovered among 11 families (7%). The diagnostic rates among patients in this cohort are similar to the overall detection rates among all 500 referred patients (39% overall positive rate and 30% among characterized genes). Among the 66 positive findings, 43 (65%) were autosomal dominant, 12 (18%) were autosomal recessive, and 11 (17%) were X-linked molecular defects. The 66 positive patients with abnormal MRI findings were significantly less likely to have an autosomal recessive finding (18%) as compared to the entire positive cohort (51/163; 31%) (p=0.05). The most commonly observed MRI findings among the patients with positive findings were cerebellar hypoplasia, delayed myelination, cerebral atrophy, and agenesis of the corpus callosum. Several well-known molecular diagnoses were provided and/or confirmed including Leigh Syndrome, Genetic Prion Disease, and spinocerebellar ataxia. These data highlight the utility of DES in providing the most comprehensive molecular diagnosis given the diversity of genetic findings including 7% within novel genes.

2457M

Identification of a Novel De Novo Mutation Associated with PRKAG2 Cardiac Syndrome and Early Onset of Heart Failure. *M. Jurkowska[1], R. Szymanczak[1], A. Wasowska[1,2], A. Mlodzinska[1,2], M. Kania[1], M. Krawczyk[1], K. Spodar[1], A. Boguszewska-Chachulska[1].* 1) Genomed Health Care Center, Genomed SA, Warsaw, Poland; 2) Polish-Japanese Institute of Information Technology, Warsaw, Poland.

Cardiomyopathies are a group of severe cardiac diseases with a strong genetic background. More than 100 genes have been associated with them. Mutations in some of the genes can manifest as various forms of cardiomyopathy, and a significant intra- and inter-family variability exists due to penetrance and genetic modifiers. Therefore, effective genetic testing is moving from single gene and subgroup testing to large gene panels covering most of the genes causing cardiomyopathies. Purpose: A family was referred to our center whose <1 year-old boy has died in consequence of hypertrophic cardiomyopathy. The family requested testing to establish the genetic background of their offspring's clinical symptoms and to be provided with genetic counselling concerning their other child's (16 years old now) risk as well as with a risk assessment for future pregnancies. Methods: Frozen blood sample was used as a source of DNA for testing. Two NGS-based approaches were applied: 1.) HaloPlex CardioMyopathy Research Panel (ILM) that screens 34 genes involved in various cardiomyopathies 2.) Whole exome sequencing (WES) with HaloPlex Exome Target Enrichment System KIT (ILM) with subsequent analysis of 106 laboratory-selected genes reported to be involved in syndromic and non-syndromic cardiomyopathy. Results: For the cardiomyopathy gene-panel, 99.35%; of all the bases had a coverage of ≥50 reads per nucleotide and 99.48%; had ≥10. The screening results were inconclusive - no pathogenic variant was found. For 106 selected genes, the WES mean coverage was 27.5 reads/bp, 56.86%; of the bases were covered >20x and 69.63%; of the bases >10x. Despite a much less efficient coverage, a heterozygous *de novo* variant in the *PRKAG2* gene, not reported in HGMD, was identified and confirmed by Sanger sequencing. The p.Phe293Val (c.877T>G) variant was found uniquely in the index case and was classified as 'likely pathogenic' in accordance with current recommendations. *PRKAG2* pathogenic variants influence cardiac metabolism but pathological glycogen storage in heart due to mutated PRKAG2 may be potentially attenuated or significantly reversed pharmacologically. Conclusion: Despite being superior to WES in coverage efficacy and facility of result interpretation, particular NGS panels risk to be too narrow for providing with proper genetic service in the genome era, when effectiveness of testing influences not only genetic counselling, but also therapeutic options.

2458T

Targeted massively parallel sequencing in molecular diagnosis: a Brazilian report. *M. Lazar, K.M. Rocha, G.L. Yamamoto, M. Aguena, V. Takahashi, N. Lourenço, M. Varela, S. Ezquina, D.R. Bertola, R. Pavanello, M. Vainzof, M. Zatz, M.R. Passos-Bueno.* University of São Paulo, São Paulo, Brazil.

Massively parallel sequencing (MPS) has emerged as a powerful tool for investigating nucleic acids and has been applied not only for research purposes, but also for diagnosis of human disorders. The Human Genome and Stem Cell Research Center (HUG-CELL, University of São Paulo, Brazil) has implemented this methodology through a customized panel of genes associated with skeletal dysplasias, neuromuscular diseases and neurodevelopmental disorders (Nextera target enrichment for MiSeq platform - Illumina). In order to evaluate the efficiency of targeted MPS for concluding molecular diagnosis, we analyzed 92 patients presenting phenotypes compatible with neuromuscular disorders (n=50) or skeletal dysplasias (n=42). Using an average coverage of 100x to 200x, it was possible to identify pathogenic variations in 62% of individuals with skeletal dysplasias and 65% of patients diagnosed with a form of neuromuscular disorder. Particularly, in a subgroup of 14 patients clinically diagnosed with dystrophinopathy and previous normal MLPA results, the mutation detection rate was around 85%. In contrast, we obtained low detection rate (33%) for a subgroup of 9 patients affected by non-specific forms of skeletal dysplasias, indicating that high mutation detection rate is obtained if a detailed clinical evaluation and restricted diagnosis are taken into consideration. Application of this customized panel in our testing routine has provided remarkable advantages for patients affected by genetic heterogeneous disorders, where accurate categorization by clinical evaluation only is challenging. Overall, besides the high sensitivity, our experience has shown that targeted MPS is time- and cost-effective, which are mandatory criterions for molecular diagnosis. For this reason, in cases with well-defined diagnostic hypothesis, targeted MPS should be considered the first-stage in molecular investigation instead of exome sequencing. Not only has this approach higher costs, but undirected search for deleterious mutations increases frequency of incidental findings, which implies ethical dilemmas that have to be exhaustively discussed among scientific community members before defining a consensus. Financial support: Fapesp/CEPID and CNPq/INCT.

2459S

Prevalence of medically actionable findings: a summary of clinical Whole Exome Sequencing cases. *Z. Niu[1], J. Beuten[1], M. Leduc[1], W. He[1], J. Zhang[1], P. Ward[1], A. Braxton[1], T. Vaughn[1], D. Muzny[3], F. Xia[1], R. Person[1], S. Plon[1,2], J. Lupski[1], R. Gibbs[1,3], A. Beaudet[1], Y. Yang[1], C. Eng[1,2].* 1) Molecular Human Genetics, Baylor College of Medicine, Houston, TX; 2) Texas Children's Hospital, Houston TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

Clinical Whole Exome Sequencing test is a powerful and effective tool to identify the genetic etiology in patients with a variety of clinical presentations. This genome wide analysis has also drawn great attention to the returning of incidental findings (we prefer the term medically actionable findings), which may greatly impact the care of the patient (and potentially family members), if the findings can lead to prevention, early diagnosis, and medical interventions. The American College of Medical Genetics and Genomics (ACMG) issued guidelines for the return of incidental findings on a core group of 56 actionable genes from clinical whole exome or genome sequence to patients. From the 2000 sequential clinical Whole Exome Sequencing cases completed at the Medical Genetics Laboratory and Whole Genome Laboratory of Baylor College of Medicine, 95 medically actionable findings were reported for an overall rate of ~4.8%. Of these, 60 findings were reported in genes on the ACMG recommended list, and 35 findings were reported in genes which met the criteria for medically actionable by expert opinion from the clinical and diagnostic team. When the consent form was updated with an opt-out from actionable findings in non-ACMG recommended gene, 2 patients out of 190 opted out of additional reporting. Majority of families requested to receive all aspects of actionable findings. 24 findings were reported in hereditary cancer predisposition genes in patients without cancer as an indication: five *BRCA1*, eight *BRCA2*, three *PMS2*, two *VHL* and single cases of *CDH1, MSH6, RAD51D, MUTYH, RET and PALB2.* 46 of the 95 mutations occurred in genes for which mutations can cause susceptibility to disorders of the cardiovascular system. Our protocol returns medically actionable findings for the proband first. Followup testing of the parents for the medically actionable findings can be ordered after disclosure of the proband's result at no cost. 33 requests from 19 families of the 95 patients with incidental findings have been received to date. In multiple cases, parental testing was requested for one parent only, typically when phenotype related to the actionable finding was noted by review of family history. Our early experience exemplified the extent of medically actionable findings uncovered from clinical Whole Exome Sequencing. Studies are needed to further evaluate the long term impact on patient care and at-risk family members when a medically actionable finding is reported.

2460M

Identification of Complex Rearrangements of the MECP2 Gene Requires a Combination of Molecular Diagnostic Techniques. *S. Ordorica, D. Lahey, O. Jarinova*, N. Carson*.* Molecular Genetics, Regional Genetics Program, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada.

The methyl-CpG-binding protein 2 gene (MECP2; MIM#300005) is a chromosome binding protein that acts as a transcriptional repressor, located on the X chromosome. Mutations in this gene can cause mental retardation, Rett Syndrome (RTT; MIM#312750) and have also been associated with Autism Spectrum Disorders. While missense and truncating mutations are most common, complex rearrangements also occur. Here we describe three cases with complex rearrangements involving the MECP2 gene that were identified using MLPA analysis and required further molecular investigations. Case 1: MECP2 MLPA analysis in a female patient showed a reduced signal (~75% signal) for all MECP2 probes and a heterozygous deletion of CDKL5 (MIM 300203) and ARX (MIM 300382) that reside on the p-arm of the X chromosome and are implicated in Rett Sydrome and Mental Retardation. Consistent with these findings, microarray analysis showed a loss of a portion of the p-arm of the X, and a mosaic loss of the remaining X chromosome. Case 2: MECP2 MLPA analysis in a male patient showed a triplication of exons 1 and 2 and a duplication of exon 3 and 4 of the MECP2 gene. Consistent with these findings, microarray analysis showed a triplication and a duplication of the corresponding regions on Xp28. Since the patient's phenotype was indicative of the presence of two copies of MECP2 (ie: MECP2 duplication syndrome, MIM 300260), a third partial copy is likely to be non-functional. Case 3: MECP2 MLPA analysis in a female patient showed a partial deletion of exon 4. Consistent with the MLPA findings, targeted sequencing of the MECP2 gene revealed a complex rearrangement involving exon 4 of the MECP2 gene: c.[523_1075delinsCG;1161_1188del], p.[Lys175_Ala358delinsArgfs; Pro388_Ser396del]. This mutation event is consistent with the presence of deletion hot spots in this gene and is predicted to be deleterious to MECP2 gene function. The three cases presented here illustrate that while MECP2 analysis is commonly conducted using Sanger sequencing and MLPA, a combination of molecular techniques is sometimes required to confirm the presence of complex rearrangements involving the MECP2 gene and provide a clear diagnostic result.

2461T

Pathogenic mutations in genes responsible for Maple Syrup Urine Disease type 1A *(BCKDHA)*, type 1B *(BCKDHB)*, and type 3 (*DLD*) determined in a large pan-ethnic cohort. *C. Perreault-Micale, A. Frieden, S. Hallam, V. Greger.* Good Start Genetics, Cambridge, MA. 02139.

Maple Syrup Urine Disease (MSUD) is an autosomal recessive metabolic system disorder that is due to dysfunction of the branched-chain alpha-keto acid dehydrogenase (BCKD) enzyme complex. This complex is composed of subunits encoded by 4 different genes. Founder mutations in 3 of these genes, *BCKDHB* and *DLD* in the Ashkenazi Jewish (AJ) and *BCKDHA* in the Mennonites, lead to an increased prevalence of MSUD in these populations. We sequenced the coding region and intron-exon borders of these three genes in a pan-ethic population of nearly 23,000 individuals referred to us for carrier screening using next generation DNA sequencing (NGS) to identify unique variants (number of different variants detected among the carriers), known variants (those previously cited in a publication or public database), and novel variants predicted to be pathogenic (nonsense mutations, conserved splice site mutations, or indels not divisible by 3). We found 16 carriers of 12 unique variants (7 known and 5 novel) in *BCKDHA*, 40 carriers of 15 unique variants (7 known and 8 novel) in *BCKDHB* and 22 carriers of 6 unique variants (2 known and 4 novel) in *DLD*.

Overall, 78 carriers of a pathogenic MSUD mutation were identified. The AJ founder mutations, p.Arg183Pro and p.Gly278Ser in *BCKDHB*, and p.Gly229Cys in *DLD*, accounted for the majority of carriers in each subtype, even though only 6 carriers reported AJ descent. In contrast, only three carriers of the *BCKDHA* founder mutation, p.Tyr438Asn, were identified; most carriers for this MSUD subtype had other mutations. Interestingly, we identified about twice as many carriers of pathogenic *BCKDHB* variants compared to the other two genes. Four out of 6 unique *DLD* variants were novel variants that we predict to be pathogenic on the basis of their truncating nature. In contrast, the groups of *BCKDHA* and *BCKDHB* unique variants consist of about half known and half novel truncating variants. In summary, we have determined the mutation prevalence and spectrum of MSUD types 1A, 1B and 3 in a large pan-ethnic population and show that a large number of novel pathogenic variants would have gone undetected using an alternative technology that is not as comprehensive as NGS.

2462S

Molecular Genetic Determination of Maple syrup urine disease cases from southwest Iran. *A. Sedaghat[1], G. Shariati[2,3], A. Saberi[2,3], M. Hamid[2,4], M. Mohebbi[5], H. Galehdari[6].* 1) Endocrinology Dept., Jundishapour University of Medical Science, Ahvaz, Iran; 2) Narges Genetic Diagnostic Lab, Ahvaz, Iran; 3) Genetic Dept., Jundishapour University of Medical Science, Ahvaz, Iran; 4) Pasteur Institute, Tehran, Iran; 5) Internal Medicin Dept., Jundishapour University of Medical Science, Ahvaz, Iran; 6) Genetic Dept., Shahid Chamran University, Ahvaz, Iran.

Maple syrup urine disease (MSUD) affects an estimated 1 in 185,000 infants worldwide, and is most frequent metabolic disorder in southwest Iran that ranges from very mild to severe form. The frequent occurrence of MSUD (like other disorders with autosomal recessive pattern) is probably due frequent consanguineous marriage. In last twelve month we analyzed 20 families with MSUD affected children. Most our cases have been preliminarily diagnosed by tandem mass as helpful tool for this propose. Mutations in four genes BCKDHA, BCKDHB, DBT, and DLD can cause MSUD. But, we find recurrently changes in the two genes BCKDHA and BCKDHB responsible for outcome of the disease. In summary, we found in 8 cases mutations in the BCKDHA gene and in other 9 cases nucleotide changes in the BCKDHB gene. Three cases were negative, despite screening of all 4 responsible genes as have been mentioned. The negative results can be explained by two reasons: wrong diagnosis or existence of the 5th undetected gene in the pathogenesis of the MSUD. However, some detected mutations were novel in analyzed cases. In cases of novel changes in one gene, we additionally screened the other three genes to assert our results. Further, we used some predicting program such as "Predict SNP", "I-Taisser", and "Mutation Tester" to underline the pathogenicity of determined changes.

2463M

A homozygous frameshift insertion in the MRPS34 gene identified in a family with two affected boys Suffering from progressive retinal dystrophy. *G. Shariati[1,2], A. Saberi[1,2], M. Hamid[1,3], M. Sedaqat[2,4], M. Mohebbi[1,5], H. Galehdari[6].* 1) Narges Medical Genetics Lab, Ahvaz, Iran, PhD, MD; 2) Genetic Dept., Jundishapour Medical science University, Ahvaz, Iran, PhD; 3) Pasteur Institute, Tehran, Iran, PhD; 4) Endocrinology Dept, Jundishapour Medical science University, Ahvaz, Iran, MD; 5) Internal Medicine Dept, Jundishapour Medical science University, Ahvaz, Iran, MD; 6) Genetic Dept, Shahid Chamran University, Ahvaz, Iran, PhD.

Progressive retinal dystrophy covers a wide range of eye disorders including syndromic and non syndromic retinitis Pigmentosa. To date, more than 100 disease causing genes have been identified in different populations. Recently, a couple (first cousin) from southwest Iran referred us with two affected boys (8 & 6 years old) suffering from progressive retinal dystrophy as preliminary diagnosis. Genomic DNA of affected individuals was subjected for next generation sequencing (NGS) leading to the detection of numerous candidate genes including a homozygous insertion (c.320_321insAGGT) causing frameshift at codon 107 (p.Q107fs). Targeted sequencing of the MRPS34 gene exon 1 showed that both parents are heterozygous for this mutation. The MRPS34 gene codes for the mitochondrial ribosomal protein 34 that is highly expressed in eye. No clinical assertion has been found for the mentioned mutation, but we consider it as a strong candidate causative for progressive blindness in the family.

2464T

Beyond Bias: Broadening the phenotype of genetic disorders using very large gene panels. *C. Stanley, J. Eggington, P. Carmody, H. Hornung, A. Zare, S. Wong, J. Dickerson, E. Kazarinoff, K. Wright, E. Nelson, D. Neilsen, T. Foss, J. Warner, K. McKernan, R. Boles.* Clinical Genomics, Courtagen Life Sciences Inc, Woburn, MA.

Ascertainment bias is a sampling bias in genetics, where a non-random sample is collected in such a way that some members of the intended population are less likely to be included than others. In genetics, the consequence is that the most severely affected individuals (e.g. those who present in the emergency rooms) are the first to be described in the literature for most genetic disorders. This bias impacts the development of diagnostic criteria and effectively removes patients with milder or atypical presentations from diagnosis and restricts the phenotypic spectrum of disease, literally marginalizing many individuals with the same disorder and consequently leaving them without a diagnosis. This approach was dictated in part by the cost associated with Sanger sequencing, where clinical diagnostic justification was required to order a gene test. Recently, massive parallel sequencing brought both a radical drop in cost-per-gene and the advent of large comprehensive gene panels, allowing individuals who do not classically fit diagnostic criteria to be evaluated for genes with or without high clinical utility. In patients without the classic clinical criteria, or those with a milder phenotype, genetic testing may provide the definitive diagnosis. Large gene panels also allow primary care physicians to order genetic tests on patients who only meet a subset of diagnostic criteria, moving genetics to the beginning of the testing process and effectively turning the traditional paradigm on its head. For the patient, this leads to a shortcut to what was otherwise a long maze of referrals and circumvents the need for a clinical diagnosis prior to testing in many cases. For the geneticist, this provides much more information regarding the clinical spectrum of genetic disease. We report the results from over twenty cases of diagnoses confirmed by genetic panels (most de novo) identified by very large gene panels targeting patients with epilepsy, developmental delay, and mitochondrial disorders and the implications of identifying the genetic cause earlier than would have been achievable if clinical diagnostic criteria were used alone. We will also highlight how early identification has important clinical management implications for patient care.

2465S

Development of a Hereditary Spastic Paraplegia gene panel using exome sequencing. *D.J. Stavropoulos[1], C.R. Marshall[1], L. Lau[1], M. Eliou[1], J. Orr[1], P.J. Sabatini[1], A. Shlien[1,2], G. Yoon[3], P.N. Ray[1,2].* 1) Molecular Genetics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Canada; 2) Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Canada; 3) Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada.

Recent breakthroughs in next generation sequencing technologies have generated a major transformation in service delivery of genetic medicine. The ability to generate high quality sequence from thousands of genes in one experiment enables the development of comprehensive gene panel tests for genetically heterogeneous disorders, which are unattainable by Sanger sequencing. We have developed and clinically validated an in silico gene panel of 48 genes for hereditary spastic paraplegia (HSP) using exome sequencing. HSP represents a heterogeneous group of neurodegenerative disorders characterized by stiffness (spasticity) and progressive weakness (paraplegia) of the lower limbs, and may be associated with ataxia, intellectual disability, seizures, peripheral neuropathy, and visual defects. An accurate genetic diagnosis will reduce the number of patients who undergo serial investigations (such as multiple brain MRIs) or invasive investigations (such as muscle biopsy) in an attempt to establish a diagnosis. Confirmation of a genetic diagnosis is essential for comprehensive testing and genetic counseling for at-risk family members, and will make possible the availability of prenatal testing for at-risk couples. The previous strategy in the Molecular Diagnostic Lab has been to perform Sanger sequencing for nine genes with a diagnostic yield of ~ 44% (Yoon et al. 2013 Neurogenetics. 14:181-8). With implementation of exome sequencing we are able to widen our scope to all 48 genes known to be associated with HSP, and can readily expand the diagnostic panel as new HSP-related genes are discovered. This strategy also facilitates the development of additional disorder-specific gene panels with the same standardized workflow. We evaluated the performance characteristics of our exome sequencing workflow, and found high reproducibility and accuracy, with 100% detection of HSP genetic variants in our validation samples. We also explore the advantages and technical challenges of this strategy to further develop clinically relevant gene panels for genetically heterogeneous disorders.

2466M

Exome-based Neurological Region of Interest Assay for Identifying Pathogenic Sequence Variants. *J. Thompson[1], W. Gillett[1], R. Lopez[1], J. Luchetti[1], P. Milos[1], P. Park[2], Y. Shen[1], N. Vena[1], T. Yu[1].* 1) Claritas Genomics, Cambridge, MA; 2) Harvard Medical School, Boston, MA.

Identification of the genetic basis for neurological disorders is made difficult by the large number of genes that can lead to such phenotypes. Further exacerbating the situation is the highly variable presentation different mutations in the same genes can show. These issues make it challenging to decide which gene(s) to analyze in detail for proper diagnosis. Sequencing individual candidate genes frequently results in a frustrating cycle of negative results that is both time consuming and expensive. Gene panels currently available typically look at dozens of genes at a time but these are difficult to maintain given the complex phenotypes observed in neurological disorders and rapid pace of new gene discovery. To circumvent these issues, Claritas developed an approach utilizing complementary sequencing technologies to analyze the exons and adjacent intronic sequence from an exome-based region of interest. This ROI includes hundreds of genes previously demonstrated as causal in a variety of neurological disorders. The phenotype-driven set of genes makes analysis faster and more economical than analysis of a whole exome and also minimizes the challenges of presenting incidental findings. By taking advantage of the power of combining the knowldedge of clinical symptoms with the scale of massively parallel sequencing technology, patients are able to benefit with more directed and rapid results. DNA variants from the two platforms are compared to confirm novel pathogenic variants. In cases where coverage is missing on one of the platforms, possibly pathogenic variants are confirmed using Sanger sequencing. Confirmed variants are then analyzed by a team of medical directors and genetic counselors to determine clinical relevance with results reported back to physicians and patients.

2467T

RYR1 [Arg2454His] gene mutation identified in a family associated with malignant hyperthermia - an undergraduate research project. *J. Li, D. Caporale.* Biology, University of Wisconsin-Stevens Point, Stevens Point, WI.

Malignant Hyperthermia (MH) is a rare life-threatening dominant disorder that causes extreme fever, muscle rigidity, acidosis, and rapid heart rate, when exposed to general anesthesia. Researchers have previously identified four amino acid replacement mutations on the RYR1 gene in MH patients: [Arg44Cys], [Arg533Cys], [Val2117Leu] and [Arg2454His]. The RYR1 gene, located on chromosome 19, codes for membrane proteins on the surfaces of muscle cells that regulate calcium flow, which is essential for muscle contraction and relaxation. These mutations can cause symptoms of MH because they cause significant changes in the structure of the calcium channel proteins, rendering them nonfunctional. The purpose of my undergraduate research study was to screen for the aforementioned four point mutations of the RYR1 gene that may be the cause of MH in a German family that has an individual with this disorder and to identify family members that carry the mutation. With IRB approval, 15 participants filled out a survey describing whether or not they have experienced any signs of MH. DNA was extracted and purified from 10 hair follicles each. The four segments of the RYR1 gene were amplified in separate reactions by the polymerase chain reaction (PCR), cycle-sequenced using dye-labeled dideoxyterminators, and then separated by capillary electrophoresis. The DNA sequences of the family members, represented as electropherograms, were compared to a wild type DNA sequence provided by NCBI GenBank in order to detect the presence of any of the four point mutations mentioned above. Sequence comparisons revealed that four family members tested positive for the Arg2454His mutation. These findings correlated with all four experiencing symptoms of MH. A recommendation was made to family members to be screened for the RYR1 [Arg2454His] mutation, to verify these results. Currently, there are only two genotyping companies in the U.S. that provide genetic testing of the RYR1 gene, which is very costly. However, since the specific mutation has been identified, family members would only need to have their DNA screened for this mutation, which would help lower costs considerably.

2468S

An exome sequencing strategy to diagnose lethal autosomal recessive disorders. *S. Ellard[1,2], E. Kivuva[2], P.D. Turnpenny[2], M. Parker[3], A.-M. Bussell[2], K. Stals[2], R. Caswell[1], H. Lango Allen[1].* 1) University of Exeter Medical School, Exeter, United Kingdom; 2) Royal Devon & Exeter NHS Foundation Trust, Exeter, United Kingdom; 3) Sheffield Children's Hospital, Sheffield, United Kingdom.

Rare disorders resulting in prenatal or neonatal death are genetically heterogeneous. Some affected fetuses can be diagnosed by ultrasound scan but often not until mid-gestation and limited fetal DNA may be available. We used a strategy for diagnosing autosomal recessive lethal disorders in non-consanguineous pedigrees with multiple affected fetuses. Exome sequencing identified genes with rare heterozygous non-synonymous coding or splicing variants in both parents and putative mutations were tested for co-segregation. In 8 couples of European ancestry we found on average only one gene where both parents were heterozygous for different rare potentially deleterious variants. A proof of principle study detected heterozygous DYNC2H1 mutations in a couple whose 5 fetuses had short rib polydactyly. Prospective analysis of three couples with multiple pregnancy terminations for presumed autosomal recessive fetal malformations identified RYR1 mutations c.14130-2A>G and p.Ser3074Phe in one couple, GLE1 mutations p.Arg569His and p.Val617Met in the second couple and MRSP22 mutations p.Arg170His and c.878+1G>T in the third. All six affected fetuses were compound heterozygous for the respective mutations, providing molecular genetic diagnoses consistent with the fetal phenotypes. This highlights the difficulty of diagnosis as neither RYR1, GLE1 nor MRSP22 had been considered as candidate genes. Exome sequencing of parental samples can be an effective tool for the diagnosis of lethal recessive disorders in outbred couples enabling management for future pregnancies.

2469M

Molecular diagnosis of congenital hyperinsulinism improves medical management and long-term outcome. *G. Alkorta-Aranburu, A. Knight Johnson, L. Ma, V. Nelakuditi, S. Das, D. del Gaudio.* Human Genetics, University of Chicago, Chicago, IL., USA.

Congenital hyperinsulinism (CHI) is the most common cause of persistent hypoglycemia in children, which if unrecognized, may lead to severe brain damage. A prompt diagnosis and initiation of appropriate treatment is essential to prevent long-term neurological complications. Over the past 20 years there has been remarkable progress in understanding the molecular basis of hyperinsulinism (HI) and mutations in ten genes are currently known to cause HI in ~50% of the cases suggesting that continuing research will identify new genes. Inactivating mutations of beta cell ATP sensitive K+ channel genes *ABCC8* (SUR1) and *KCNJ11* (Kir6.2) cause the most common and severe type of HI, which is unresponsive to diazoxide, a channel agonist. Focal HI, also diazoxide-unresponsive, is due to the combination of a paternally-inherited *ABCC8* or *KCNJ11* mutation and a paternal isodisomy of the 11p15 region, which is specific to the islets cells within the focal lesion. It is important to differentiate these two types as surgery can cure focal but not diffuse hyperinsulinism, which frequently requires pancreatectomy. Molecular genetic testing and imaging with 18F-DOPA-PET/CT can help to diagnose diffuse or focal forms of HI.

We developed a comprehensive 2-tiered testing algorithm for the molecular diagnosis of CHI to direct the clinicians along a process of exclusion, so that rarer forms of HI would only be considered once commoner and diazoxide-unresponsive forms are ruled out. Tier-1 includes mutation analysis of genes associated with diazoxide-unresponsive HI, *ABCC8*, *KCNJ11* and *GCK* for which results are obtained within a week, allowing for rapid surgical intervention if required. If a mutation is not identified in tier-1 or the patient is known to be responsive to diazoxide, a tier-2 approach that includes targeted next-generation sequencing and array-CGH deletion/duplication analysis of all ten known genes causing HI can be performed. To date, we have identified the genetic etiology in 3/19 (16%) patients studied and likely causative mutations were identified in *ABCC8*, *GLUD1* and *GCK*. Variants of uncertain clinical significance (VOUS) were also identified in 3 patients. Parental analysis will be performed to clarify mode of inheritance of the mutations and provide insights on the nature of the VOUS. Studies are also ongoing to assess the impact of genetic testing results on patients' medical management and clinical course.

2470T

Carrier Screening in the Sephardic/Mizrachi Jewish Population for Genetic Disorders with Known Founder Mutations. *X. Cai, G. Akler, L. Shi, J. Zhang, G. Diaz, L. Edelmann, R. Kornreich.* Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

The Sephardic/Mizrachi Jewish (SMJ) population includes descendants of Jews from Spain, Southern Europe, North Africa and the Middle East. They are a heterogeneous group defined by a combination of geography, languages and religion. Similar to the Ashkenazi Jewish (AJ) population, recurrent founder mutations associated with autosomal recessive and X-linked genetic diseases have arisen in the SMJ population due to geographic and cultural isolation. Current genetic SMJ screening panels are limited mostly due to heterogeneity. The purpose of this study was to select diseases/mutations that may be appropriate for an SMJ carrier screening panel, and to establish accurate carrier frequencies among distinct SMJ subgroups in the New York metropolitan area. A comprehensive panel of mutations was selected based on literature review regarding diseases reported among SMJ individuals. Saliva samples were collected with informed consent and screening was performed by Sequenom MassARRAY iPLEX platform on over 500 individuals including the Iraqi, Iranian, Syrian, Moroccan, Yemenite, Tunisian, Bukharian, and Egyptian communities. To explore the differences between AJ and SMJ populations, approximately 1500 AJ anonymized DNAs were also screened using the same panel, and the carrier frequencies were compared between these two groups. Forty nine autosomal recessive and two X-linked diseases were selected for screening, including 130 mutations from 51 genes. Four diseases were determined to have a high carrier frequency (>1%), including hereditary inclusion body myopathy 2 (~1 in 10), phenylketonuria (~1 in 20), familial mediterranean fever (~1 in 25) and Wolman disease (~1 in 30). Six diseases had a low carrier frequency (<0.5%), including retinitis pigmentosa 28, Costeff disease optic atrophy, and glycogen storage disease type II with carrier frequencies of about 1 in 250. Acute infantile liver failure, retinitis pigmentosa 25 and glycogen storage disease type III all had lower frequencies of approximately 1 in 500. Several of these mutations are also recurrent in the AJ population. These findings suggest a set of candidate disorders with carrier frequencies high enough to be considered for inclusion in a prenatal SMJ carrier panel.

2471S

Clinical whole exome sequencing in a group of pediatric heterogeneous disorders: Yield and new gene discoveries. *J. Gauthier[1], I. Thiffault[1,2], JF. Soucy[1,2], FH. Hamdan[3], P. Campeau[2,3], M.E. Samuels[3, 5], E. Lemyre[2], A.-M. Laberge[2], C. Brunel-Guitton[2], S. Nizard[2], G.A. Mitchell[1, 2,3], G.A. Rouleau[1,4], JL. Michaud[1,2,3].* 1) Medical Biological Unit, Molecular Diagnostic Laboratory, Sainte-Justine University Hospital Center, Montreal, QC, Canada; 2) Division of Medical Genetics, Department of Pediatrics, Sainte-Justine University Hospital Center, Montreal, QC, Canada; 3) CHU Sainte-Justine Research Center, Montreal, Quebec, Canada; 4) Montreal Neurological Institute and Hospital, Montréal (Québec), Canada; 5) Department of Medicine, Université de Montréal, Montreal (Quebec), Canada.

It is now increasingly recognised that whole-exome sequencing (WES) is a useful diagnostic approach for the identification of disease mutations responsible for various genetic disorders but the yield of WES and the optimal methodology for its use in the clinic are not well defined. In the process of implementing WES in a clinical setting in a universal health care system, CHU Sainte-Justine, a Mother and Child University Hospital Center (Montreal, Quebec, Canada), recently launched a WES pilot project for individuals with rare conditions, unexplained despite extensive investigations. A total of 96 patients, mainly children, were recruited from genetics clinics for whom WES was performed and analysed. For patient recruitment, priority was given to clinical presentations known to have great genetic heterogeneity. We focused on rare variants known/strongly predicted to alter gene function in genes previously associated with monogenic disorders. To date, the overall diagnostic yield in 70 patients with completed analysis is 36% (25/70). It is 37% (18/49) for patients with global developmental delay / intellectual disability, 33% for other patients representing a heterogeneous group of diseases (skeletal dysplasia, retinopathy, deafness, neurodegenerative diseases, etc.). We did not identify any incidental findings but we fortuitously discovered strong candidate genes for several conditions, including Megacystis-microcolon intestinal hypoperistalsis syndrome (MYH11), congenital myopathy (PAX7) and mitochondrial diseases (TRNT1). Therefore, we conclude that our clinical WES approach is associated with a high diagnostic yield.

2472M

Towards the development of a standardized analytical pipeline for clinical whole genome analysis for rare disease diagnosis. *R.K. Iyer[1], A. Khromykh[1], D. Richards[2], G. Eley[1], S. Scott[2], B.D. Solomon[1], J.G. Vockley[1], J.N. Niederhuber[1].* 1) Inova Translational Medicine Institute, Inova Fairfax Hospital, Falls Church, VA; 2) QIAGEN Silicon Valley, Ingenuity Systems Bioinformatics, Redwood City, CA.

For many children with rare, potentially genetic syndromic disease, a clear diagnosis remains elusive, and the diagnostic odyssey continues. The successful application of whole exome or whole genome sequencing (WES/WGS) in causative gene discovery for rare Mendelian disease has the potential to vastly reduce or eliminate such situations. A very significant barrier to the routine clinical use of WES/WGS is the lack of standardized, validated and repeatable and scalable bioinformatics solutions that can facilitate the rapid annotation, filtering and classification of the large number of potentially biologically relevant variants detected by these technologies. We are currently evaluating QIAGEN's new evidence-based clinical decision support software, Ingenuity Clinical, as a potential solution for clinical interpretation and reporting by analyzing a set of whole genome data generated on 30 family trios centered on probands, afflicted with severe congenital malformation(s) that are not consistent with a known clinical diagnosis. There is no clear family history of disease, and standard genetic testing was negative. The Ingenuity Clinical software platform classifies variants using draft ACMG 2.0 assessment guidelines based on a pre-curated knowledge base of biomedical literature and clinical evidence that allows standardized and scaled annotation, filtering, classification and reporting of observed clinically relevant variants. For cases in which a genetic cause for disease is suspected based on family history, but no causal variant has been previously identified, we use QIAGEN's Ingenuity Variant Analysis to identify gene variants that may perturb biological pathways/networks upstream of known causal genes in described cases with overlapping phenotypes. Initial analysis, biological interpretation and clinical assessment of variants through these Ingenuity platforms has unequivocally identified a mutation in one proband that is clearly consistent with the clinical presentation, and facilitates a diagnosis in this patient. Initial analysis of several other cases indicates good candidates in all of them, but further assessment is required to determine their validity. The analyses on all 30 probands continue and findings will be presented.

2473T

Population diversity and the genetics of hypertrophic cardiomyopathy. *A.K. Manrai[1, 2], B.H. Funke[3, 4], H.L. Rehm[3, 5], M.S. Olesen[6], B.A. Maron[7], P. Szolovits[8], D.M. Margulies[2], J. Loscalzo[7], I.S. Kohane[1, 2].* 1) Health Sciences and Technology, Harvard-MIT, Cambridge, MA; 2) Center for Biomedical Informatics, Harvard Medical School, Boston, MA; 3) Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, MA; 4) Department of Pathology, Harvard Medical School and Massachusetts General Hospital, Boston, MA; 5) Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 6) Laboratory of Molecular Cardiology, Department of Cardiology, the Heart Centre, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; 7) Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 8) Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

BACKGROUND—Risk stratification for hypertrophic cardiomyopathy (HCM) is an exemplar of the clinical gains attainable by targeted genetic testing. After decades of research, sequencing laboratories now identify a causal mutation in over one third of referred patients. Using these findings, clinicians routinely assess risk for the patient's relatives and may even tailor therapy for rare patients. However, the clairvoyance offered by genetic testing comes with a cost—when variants are misclassified, there is potential for harm. METHODS—Using publicly accessible large-scale exome data, we identified variants previously considered causal of HCM that were overrepresented in the general population. We studied these variants in diverse populations using complementary sequencing data, and reevaluated their initial ascertainments in the medical literature. We reviewed patient records at a leading genetic testing laboratory for all occurrences of these variants during the near decade-long history of the laboratory. RESULTS—Multiple patients, notably all of African or unspecified ancestry, received incorrect positive reports of variants classified as "pathogenic" or "presumed pathogenic" and later revised to "benign." All identified variants were significantly more common in African Americans than European Americans ($P < 0.001$). We show that if moderately diverse control sequencing data had been available to test previous pathogenicity assertions, these variants would likely have been classified sooner as benign, possibly avoiding multiple misclassifications in African-ancestry individuals and their families. We identify issues in the ascertainment process that may have led to these errors in the medical literature. CONCLUSIONS—Diverse population data can be used to identify ancestry informative yet clinically uninformative markers. These findings expand upon current guidelines, which recommend using ethnically matched controls to interpret variants. As diverse sequencing data becomes more widely available, we expect variant reclassifications to increase, particularly for groups that have historically been less well studied.

2474S

Prevalence of variants of unknown significance in a next-generation sequencing panel: an experience with autism spectrum, intellectual disability, and dysmorphic features disorders. *M. Nelson, J.E. Williamson, J.J. Evans, J.A. Carstens, E.E. Kaspar, E.J. Rief, J.L. Berry, A.C. Carter, D.L. Bishay, R.L. Barbar, J.N. Sanmann, B.J. Dave, R.T. Hagelstrom, W.G. Sanger.* Human Genetics Laboratory, Munroe Meyer Institute, University of Nebraska Medical Center, Omaha, NE.

Accurate interpretation of genetic test results to assess the phenotypic effects of genomic variants is necessary for effective patient management. Frameshift, nonsense or canonical splice site changes that yield a truncated protein product are more likely to represent deleterious variants with stronger association to disease phenotype; whereas, missense and intronic changes more often lack clinical and functional data for classification as either pathogenic or benign. Genetic test results identifying a pathogenic or likely pathogenic variant are characterized as 'positive'; those identifying variants of unknown clinical significance (VUSs) are 'inconclusive'; and those with only benign or likely benign (LB) variants are generally considered 'negative'. We analyzed the frequency of VUSs and LB changes and the specific genes and regions involved in more than 150 patients referred for a 49 gene next-generation sequencing autism/intellectual disability/multiple anomalies (A/ID/MA) panel. Approximately 1/3 of the patients had at least one VUS and more than 1/2 had at least one LB variant. While the majority of VUS and LB variants were non-recurrent and/or missense changes, greater than 30 % were represented by only 4-6% of the genes included in the panel (e.g. *KMT2D* [MIM 602113], *SHANK2* [MIM 603290], *ERCC6* [MIM 609413], *ASPM* [MIM 608716]). Within these genes, clusters of VUSs and LB variants, some within protein domain structures, were present. Fewer than 7% of patients with VUSs or LB variants had identifiable pathogenic or likely pathogenic changes. Expanding our understanding of VUSs and LB variants in A/ID/MA disorders will facilitate their reclassification. Consequently, this will increase specificity of the tests and reduce the burden resulting from information of unknown significance on the physicians, genetic counselors and patients.

2475M

Comprehensive Analyses of Causative Variants in Hereditary Nephrotic syndrome (NS) in Children: Frequency, Clinical Utility and Phenotype- Genotype Correlation. *F.E. Orkunoglu-Suer, M. Meservey, M. Jaremko, M. McCarthy, V. Storozuk, C. Hoffman, K. Liaquat, J. Lapierre, W. Zhenyuan, S. Batish, J.J Higgins.* Quest Diagnostics, Athena Diagnostics Inc, Worcester, MA., US.

Background: Hereditary nephrotic syndrome (HNS) is a genetically heterogeneous disorder representing a spectrum of inherited renal diseases that cause podocyte anomalies and resultant proteinuria. Mutations in the *NPHS1, NPHS2, WT1, PLCE1, and LAMB2* genes cause the onset of HNS in neonates, infants, and children, with progression to end-stage renal disease in later life. The relative frequencies of causative gene variants in children with HNS are unknown. **Methods:** We analyzed the frequencies of pathogenic variants (PVs) and variants of unknown significance (VUS) in a targeted panel of five genes in 273 patients referred to a clinical laboratory for HNS testing. Mutation analysis was carried out by direct sequencing of the *WT1, LAMB2, NPHS1, NPHS2, and PLCE1* genes in all cases. The study included 10 prenatal samples, samples from various childhood age groups (82 neonatal, 146 infant, 16 child), and 15 syndromic cases that were consecutively submitted. **Results:** Sequence variants (total 100 PV, 188 VUS) were detected in 159 patients. Variants were most frequent in the *NPHS1* gene (50 PV, 82 VUS), followed by the *PLCE1* gene (14 PV, 12 VUS). Variants were less frequent in the WT1 (15 PV, 14 VUS), NPHS2 (17 PV, 11 VUS), and *LAMB2* (4 PV, 42 VUS) genes. VUS were found in multiple genes in 26 cases. **Conclusions:** An analysis of 273 samples submitted for HNS testing shows that over half have variants of interest. This study confirms the clinical value of genetic testing for HNS using this panel.

2476T

Targeted Gene Panel for Diagnosis of Nephrotic Syndrome in Pediatric Patients. *J. Rousseau, L. Ambrogio, A. DiLucia, S. Hintzen, M. Kennedy, E. Mauceli, P. Milos, S. Reilly, T. Ross.* Claritas Genomics, Cambridge, MA.

The prevalence of idiopathic nephrotic syndrome (NS) is approximately 16 cases per 100,000 individuals. Recent studies show that much of the observed disease incidence can be accounted for by mutations in just 29 genes, most of which are expressed in the kidney and which play a variety of biological roles. Until recently, children with NS, steroid-resistant NS (SRNS), proteinuria, and focal segmental glomerulosclerosis (FSGS) underwent sequential, gene-by-gene testing because there were no clinical tests that simultaneously assessed mutation load across the multiple genes involved in these closely-related disorders. To allow for more rapid and cost-effective diagnosis of these patients, we validated an assay that allows detection of mutations in all exons and adjacent intronic sequences of the 29 genes most commonly found mutated in these disorders. This panel is expected to identify disease-causing mutations in 15% of cases that have a general diagnosis of chronic kidney disease and will offer a significantly higher detection rate for children with proteinuria specifically. Ordering guidance for the assay is thus based on a relatively straightforward phenotypic presentation of high protein levels in the urine, and given the panel's speed, clinical providers can now bring treatment options to their patients more quickly than with previous testing options.

2477S

A novel *EDA* splice site mutation cause hypohidrotic ectodermal dysplasia in a heterozygous female with severe phenotype. *C. Weng[1], T.Y Wei[1], F. Jin[4], P. Yu[1], M. Qi[1,2,3].* 1) School of Medicine, Zhejiang university, Hangzhou, China; 2) James D Watson Institute of genome Sciences, Hangzhou, China; 3) Department of pathology and laboratory Medicine, University of Rochester, NY, USA; 4) Department of Reproductive Endocrinology, Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, China.

Hypohidrotic Ectodermal dysplasia (XHED[MIM 305100]) is a type of X-linked genodermatosis characterized by the abnormal development of sweat glands, teeth and hair. Our study aims to characterize the pathogenesis of a heterozygous female diagnosed with typical HED at molecular level. The proband underwent complete physical examination. Mutation screening was performed by Sanger sequencing on associated genes. In silico analysis was used to predict putative effect of the identified mutation, and in vitro exon trapping assay to analyze mutant transcript processing. PCR-based human androgen receptor gene assay (HUMARA) was applied to study the skewed X chromosome inactivation pattern. A novel mutation c.925-2A>G at the end of intron 7 of *EDA* gene(MIM 300451) was identified. This mutation is expected to influence the function of that splicing acceptor. The exon trapping assay indicated the first 32 bp sequence of exon 8 had been skipped abnormally in the splicing process. The resulted premature stop codon truncated 71 amino acid including the tumor necrosis factors (TNF) domain at C-terminal-end. This may lead to the deformation of the inner anti-parallel β-strand, weakening the hydrophobic interfaces and losing a number of hydrogen bonds. X chromosome HUMARA assay for the affected proband and her mother of proband, who is also a mutation carrier but with normal phenotype, showed an obvious skewed X chromosome inactivation pattern. Conclusion: A splice site mutation resulting in partial exon skipping was found in a typical hypohidrotic Ectodermal dysplasia patient. The mutation probably leads to a truncated protein with part of C-terminal skipping out, affecting the interaction properties of the *EDA* protein. The skewed X chromosome inactivation pattern is associated with the severity of the phenotype.

2478M

Follow-up of the first 250 clinical WES cases: periodic re-analyses revealed additional molecular diagnoses. *Y. Yang[1], F. Xia[1], R. Person[1], N. Niu[1], M. Leduc[1], J. Beuten[1], D.M. Muzny[2], S.E. Plon[1], J.R. Lupski[1], A.L. Beaudet[1], R.A. Gibbs[1,2], C.M. Eng[1].* 1) Dept Molec & Human Gen, Baylor College of Medicine Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine Houston, TX.

We recently reported clinical whole exome sequencing (WES) results for the first 250 unrelated patients received between October 2011 and June 2012 (Yang, et al NEJM 2013;369:1502). The patients mostly had neurological manifestations suggesting potential genetic etiology but no definitive diagnosis could be rendered despite extensive workups. Molecular diagnoses by WES were made in 62/250 patients (25%), including 4 patients with 2 molecular diagnoses resulting in blended phenotypes. The initial 25% diagnostic rate was higher than several other comparable physician ordered genetic tests including chromosomal microarray analysis and most targeted gene tests. Further, more diagnoses from the existing WES dataset are expected as new disease genes are published. We have been following up the 250 cases since the completion of the clinical reports in 2012 by performing re-analyses of the original WES variants every six months using a list of newly published disease genes. This effort, which was free of charge, resulted in additional molecular diagnoses in 4 AD disorders (genes: *AHDC1*, *ASXL3*, *KCNT1*, *KMT2A*), 2 AR disorders (genes: *SERAC1*, *SFXN4*) and 2 X-linked (XL) disorders (gene: *WDR45*, seen in 2 unrelated patients). In addition to findings in new disease genes, molecular diagnoses were also confirmed in 2 more cases by add-on parental studies and phenotype correlations. The follow-ups improved the WES diagnostic rate from 25% to 28% by adding 10 more diagnoses. Interestingly a case previously reported to carry a *de novo* truncating mutation in *ANKRD11* was now found to carry an additional *de novo* truncating change in a new disease gene *AHDC1*, making this case the fifth case with two diagnoses out of the 250 samples. It should be noted that the current process, which focused on truncating mutations and only considered missense changes with compelling evidence, was not a full-scale re-analysis. Nevertheless, every new diagnosis required ~ 2 hours on activities such as literature review, inheritance pattern and disease mechanism verification, phenotype and data evaluation, report update and communication with referrals. There is also significant time spent on evaluating and later ruling out other candidate cases. For a full-scale re-analysis, more automation is needed and the test may be offered for a fee. In our laboratory, WES patients without a clinical diagnosis were also offered the option of entering research studies for novel disease gene discoveries.

2479T

Identification of 290-bp deletion as a first report on the beta- globin gene in South of Iran. *M. Hamid[1], GH. Shariati[2], L. Dawoody Nejad[1], H. Galehdari[2], AH. Saberi[2], B. Kaikhaei[3], M. Mohammadi -Anaei[2].* 1) Department of Molecular Medicine, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran; 2) Narges Medical Genetics & PND Laboratory, No. 18, East Mihan Ave, Kianpars, Ahvaz ,Iran; 3) Research Centre of Thalassemia and Hemoglobinopathies, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

In this study we describe the first report of 290-bp deletion (c.-176_92+25del) in four individuals from three unrealated families in Khuzestan province with Arabic ethnic background. They were referred for carrier detection as part of a national program for the prevention of thalassemia with anysocytosis, microcytosis and hypochromia. All the Individuals had elevated Hb A2 and normal HbF levels. One of the individual, Offspring of K.B, was 5 years old girl inherited both defects from her parents. Physical examination of this patient showed paleness, slightly hepato- and splenomegaly and she was not transfusion dependent. The 290 bp deletion characterized by DNA sequencing and MLPA test. The deletion removes the promoter region, the entire exon 1 and 5 end of IVS-1. The MLPA results confirmed the deletion by probs ranging from prob 21(Promotor) to prob 1(HBB intron 1). Diagnose of this β-thalassemia deletion can be extremely useful for prenatal diagnosis in the couple especially in the consanguineous and ethnic marriage.

2480S

PCR-based method for complete HLA gene sequencing and capture-based method for entire HLA region sequencing. *K. Hosomichi, I. Inoue.* Division of Human Genetics, National Institute of Genetics, Mishima, Shizuoka, Japan.

Human leukocyte antigen (HLA) is a group of genes that are extremely polymorphic among individuals and populations and has been associated with more than 100 different diseases and adverse drug effects. HLA typing is accordingly an important tool in clinical application, medical research, and population genetics. To date, several high-throughput HLA typing methods using next-generation sequencing (NGS) have been developed. We have previously developed a phase-defined HLA gene sequencing method using the MiSeq sequencer. Our sequencing protocol and pipeline provided essentially complete phase-defined HLA gene sequences without referring IMGT/HLA database; however, it required complicated and labor-intensive workflows especially in the library preparation step. Most importantly, the method was not well adapted for processing multiple samples. Here, we developed a Bead-based Normalization for Uniform Sequencing (BeNUS) procedure using three steps of bead purification. BeNUS can easily and precisely normalize the molar concentrations of up to 96 samples. We applied long-range PCR to amplify *HLA-B* for 96 samples followed by transposase-based library construction and 300 bp paired-end sequence reads with the MiSeq. After sequencing, we observed low variation in read percentages (0.2% to 1.55%) among the 96 demultiplexed samples. On this basis, all the samples were amenable to haplotype phasing using our phase-defined sequencing method. Our HLA sequencing method optimized for 96 multiplexing samples is highly time and cost effective and is especially suitable for automated multi-sample library preparation and sequencing. In addition, we established a sequence capture method for the 3.8 Mb entire HLA region for sequencing and targeted-bisulfite sequencing to detect genetic (SNVs, small and large indels) and epigenetic (methylation variable positions) information together with the HLA haplotype map information. Only one capture step was enough to prepare the two different types of libraries, targeted sequencing and targeted-bisulfite sequencing, for 12 samples as a pooled sample. After one MiSeq run of 350 bp + 250 bp paired-end reads, more than 97% of the target region was covered by at least 20-fold in all samples, although the variation in read percentages among 12 samples was observed as 4.8% to 11.5%. Our future plan is to set these methodologies for practical application of medical research and clinical application.

2481M

Identifying disease causing variants ranging from SNVs, small InDels, single exon to whole gene deletions in the RB1 gene through a single NGS-based test. *D.L. Abramovitz[1], S. Agarwal[2], A.U. Mannan[2], A.J. Pandian[3], R. Gadkari[2], R. Ramalingam[2], S. Sankaran[2], V. Veeramachaneni[2], P. Ramamoorthy[2], R. Hariharan[2].* 1) Strand Genomics, San Francisco, CA; 2) Strand Centers for Genomics & Personalized Medicine, Bangalore, Karnataka India; 3) Vision Research Foundation, Sankara Nethralaya, Chennai, Tamil Nadu, India.

Introduction - NGS (next-generation sequencing) based tests provide clinical diagnostic labs with an attractive alternative to the traditionally employed sequential testing strategies. With the right combination of test design and analysis tools a single NGS test can identify genetic abnormalities arising from single nucleotide variations (SNVs) and structural variations (SVs) including small and large indels and copy number variations (CNVs). At Strand Centers for Genomics and Personalized Medicine, we have run hundreds of diagnostic tests ranging from whole exome to disease specific panel based NGS tests. We illustrate the clinical utility of combining efficient panel based tests with the right analysis tools in providing quick and economic diagnosis in a representative set of retinoblastoma cases. **Materials and Methods** - The samples were tested using either Illumina's TruSight Cancer panel (consisting 94 genes) or our custom designed Eye Disorders panel. In both panels, the RB1 gene is included. Paired end sequencing was done to an average coverage of >100X, the analyses were done using AVADIS® NGS software and StrandOmics™ was used for the interpretation and reporting of the variants. **Results** - We were able to provide a positive diagnosis in all retinoblastoma cases identifying the causative variants ranging from SNVs to small deletion to whole gene deletion. Performing CNV analysis using normalized read depths and z-scores from all samples previously run on the panel, we are able to pick up statistically significant changes causing heterozygous deletion of the whole RB1 gene as well as the deletion of a single exon of the gene. This NGS based method was more sensitive than FISH (fluorescence in situ hybridization) in identifying deletions. As in one case, previous analysis by FISH had identified a 13q14 deletion, however the NGS test revealed a heterozygous splice site variation as the causative variant. **Conclusions** - With the combination of panel based sequencing, AVADIS® NGS and StrandOmics™, we illustrate the ability to quickly identify disease causing variants which range from SNVs to whole gene deletions. These methods can be extended to other multi-gene diseases where sequential testing could cause significant delays in diagnosis.

2482T

Diagnoses by Clinical Exome Testing Suggest Wider than Expected Phenotypic Spectra on New Disease Genes: Implications for Choosing Testing Strategy and Interpretation of Results. *F. Xia[1], E.P. Simpson[2], C.W. Brown[1], M.F. Wangler[1], J. Zhang[1], M.S. Leduc[1], Z. Niu[1], R. Person[1], D.M. Muzny[1], A.L. Beaudet[1], R.A. Gibbs[1], C.M. Eng[1], Y. Yang[1].* 1) Baylor College of Medicine, Houston, TX; 2) Houston Methodist Hospital, Houston, TX.

Recently discovered disease genes account for a significant portion (~25-30%) of diagnoses made by clinical WES. However, some of the new genetic disorders do not have characteristic symptoms and the phenotype spectra are yet to be expanded with additional studies. The lack of knowledge and recognition in disease phenotypes can pose a challenge in making clinical diagnoses and choosing specific genes for testing, which is illustrated here with two examples. 1) Within two years, a previously asymptomatic 40 year-old female progressed to dysphagia, loss of speech, spastic quadriparesis and became non-ambulatory. WES detected a heterozygous p.A781V variant in *CSF1R*, which was recently associated with hereditary diffuse leukoencephalopathy with spheroids (HDLS). The same variant had been reported to be segregating with disease in multiple affected members of a different affected family. The age of onset and clinical presentation of HDLS is highly variable and its associated phenotypes overlap with at least 17 other genetic disorders. The deceased father and two paternal cousins of our proband were clinically diagnosed with Alzheimer, cerebral atrophy and Parkinson diseases respectively. In light of the molecular diagnosis in the proband, it is possible that those affected relatives had the same disorder. 2) A 6 year-old male had a normal growth history followed by sudden deceleration. He has additional clinical features including autism, intellectual disability, hydrocephalus, macrocephaly with frontal bossing, and a small cyst in the parieto-occipital region by MRI. By WES, a *de novo* p.Q882X pathogenic variant in a new disease gene *AHDC1* was detected. Four cases of *AHDC1*-related disease had been previously reported and each shared features of hypotonia, intellectual disability, expressive language disorder and global developmental delay. This proband showed a different regression pattern on growth, relatively normal muscle tone, and was cognitively more challenged than previous cases. Our experience indicates WES is particularly effective in making rapid diagnoses soon after disease genes are discovered. Further delineation of disease phenotypes and potential expansion of phenotypic spectra will require careful examinations of aggregated WES data and close collaborations between the diagnostic labs and referring physicians in order to distinguish causal from chance correlations between disease gene and phenotype.

2483S

Comprehensive Evaluation of Congenital Immunodeficiency by Next Generation Sequencing. *E.B. Gorman[1], H. Yu[1], S. Chen[1], A. Mangubat[1], K. Bagley[1], J. Wang[1], K.E. Berge[2], H. Sorte[2], J.R. Lupski[1], J.S. Orange[3], L.F. Forbes[3], I.C. Hanson[3], A. Stray-Pedersen[3], V.W. Zhang[1], L.J. Wong[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, United States; 2) Department of Medical Genetics, Oslo University Hospital, Norway; 3) Center for Human Immunobiology, Immunology Allergy Rheumatology, Texas Children's Hospital, Feigin Center, 1102 Bates, Suite 303, Houston, TX, United States.

Background: Congenital immunodeficiency is an inherited disorder of the immune system. Severe combined immunodeficiency (SCID), characterized by profound deficiencies of T cells and/or B cells at the time of birth, is the most severe form of congenital immunodeficiency. If not treated promptly, affected patients usually do not live beyond infancy due to severe, recurrent infections. Genetic heterogeneity of SCID frequently delays the diagnosis, which is crucial for life-saving treatment and optimal management. Methods: All targeted exons and at least 20 bp of flanking genomic sequences of 46 genes involved in SCID are enriched using SeqCap EZ solution-based capture and are subjected to Massively Parallel Sequencing (MPS) on Illumina HiSeq2000. Results: The target gene capture/MPS provides an average coverage of 1000X. Exons with insufficient coverage (<20X) or high sequence homology (pseudogenes) are complemented by PCR/Sanger sequencing using gene-specific primers to ensure the 100% coverage of all targeted regions. In a pilot study, deleterious mutations were detected in fourteen out of seventeen patients analyzed (82%). Identified mutations include compound frameshift mutations in CORO1A, a novel indel mutation in JAK3, a novel 70-bp duplication in a RNA gene, RMRP, novel missense mutations in TBX1, ZAP70, and X-linked FOXP3, as well as reported mutations in DCLRE1C and RAG1. Deep coverage of next generation sequencing revealed a c.361dupA frameshift mutation of IL7R at 82% of heterozygosity in the blood sample of a female patient. Further analysis of a buccal sample from the same patient after hematopoietic stem cell transplantation confirmed the presence of the wild-type allele at a low level (~20%), indicating possible somatic mosaicism for that patient (both in blood and in other tissue). Conclusion: High throughput deep sequencing analysis greatly increases the diagnostic yield of congenital immunodeficiency. Establishing a molecular diagnosis of immunodeficiency enables early immune reconstitution through proper treatment and guides a better management for improved long-term quality of life.

2484M

Utility of Expanded Gene Panels in Clinical Diagnostics: A Tale of a Laboratory's Experience with Two Autism Testing Panels. *R.T. Hagelstrom, M. Nelson, E.J. Rief, J.J. Evans, J.M. Carstens, J.L. Berry, J.E. Williamson, E.E. Kaspar, A.C. Carter, D.L. Bishay, R.L. Barbar, B.J. Dave, J.N. Sanmann, W.G. Sanger.* Human Genetics Laboratory, Munroe-Meyer Institute for Genetics and Rehabilitation, University of Nebraska Medical Center, Omaha, NE.

The advent of next generation sequencing (NGS) panels has become common place in clinical genetic testing strategies. Clinicians have the ability to order a single test that will interrogate tens to hundreds of genes in a time period, and at a cost, that was typical for single gene testing only a few years ago. Like all diagnostic tests, however, NGS panels should be evaluated for their clinical utility, a measurement that weighs all benefits and drawbacks of a particular test. Some general aspects of clinical utility include: analytic validity (how well the test can predict the presence or absence of a variant), clinical validity (how well the variant being analyzed is related to the presence of a specific disease), clinical utility (whether the test can provide information about diagnosis, treatment, management, or prevention of the disease), as well as the ethical, legal, and social implications (ELSI) of the test. Our laboratory has performed NGS testing for patients with autism, intellectual disability, and/or multiple congenital anomalies for several years. During this time, we have utilized two separate panels, a 49-gene panel, and then later an 86-gene panel. The larger panel increased our pathogenic variant rate by approximately 4-fold while only increasing our variants of uncertain significance (UCS) rate by approximately 1.5-fold. These results suggest that the clinical utility of the larger panel is increased over that of the smaller panel.

2485T

Important Factors in Designing Accurate and Reliable Next-generation Sequencing (NGS) Assays. *P. Mueller.* Dept NCEH/DLS/NSMBB, CDC, Atlanta, GA.

Problem: NGS applications are increasing in newborn screening (NBS), including reducing false-positives, confirming positive screens, identifying carriers, and providing timely information for appropriate initiation of treatment. NBS laboratories, commercial providers, and researchers are developing NGS assays for genes, gene panels, and the whole human exome and genome, however, there are challenges in designing accurate and reliable assays. Design and methodological challenges: New data from the GENCODE Project, an international consortium, is providing information on a large number of pseudogenes in the human genome. These pseudogenes are distributed throughout the genome and most have high homology to their parent functional genes. These data indicate that approximately 17% of the roughly 20,000 human genes have pseudogenes. The very mutations that are sought in sequencing the parent genes can be present in the normal sequence of the pseudogenes. Therefore, it is likely that false reports of mutations in the parent genes will result from these homologous regions unless appropriate methods are used. This has implications for all standard approaches to targeted, whole exome, and whole genome NGS. Additionally, special care is needed in designing 1) primers and hybrid capture arrays, 2) depth of coverage, 3) and data analysis to maximize the accuracy and reliability of NGS. Current approaches and tools: Available approaches and tools include specific long-range amplicons, software for determining primer and oligonucleotide specificity, designing appropriate coverage for the quality score used, and appropriate data trimming. Implications: New approaches are needed to improve the accuracy and reliability of NGS especially when screening newborns, children and adults, because the use of this technology in a non-symptomatic population increases the risk of a false result compared to sequencing a clinical population symptomatic or diagnosed with genetic disease.

2486S

Proband whole-exome sequencing as a cost-effective diagnostic strategy for suspected Mendelian disorders. *J. Thevenon[1,2,3], J. St-Onge[2,3,4], Y. Duffourd[2,3], A. Masurel-Paulet[1,2], S. El Chehadeh-Djebbar[1,2], C. Juge[4], P. Callier[5], M. Lefebvre[1,2], M. Chouchane[6], V. Darmency-Stamboul[6], M. Milh[7,8], F. Feillet[9], C. Thauvin-Robinet[1,2,3], L. Faivre[1,2,3], J-B. Rivière[2,3,4].* 1) Centre de Génétique, CHU de Dijon, Dijon, France; 2) 1.Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement (TRANSLAD), Centre Hospitalier Universitaire Dijon, F-21079 Dijon, France; 3) Equipe d'Accueil 4271, Génétique des Anomalies du Développement, Université de Bourgogne, F-21079 Dijon, France; 4) Laboratoire de Génétique Moléculaire, Plateau Technique de Biologie, Centre Hospitalier Universitaire Dijon, F-21079 Dijon, France; 5) Laboratoire de Cytogénétique, Plateau Technique de Biologie, Centre Hospitalier Universitaire Dijon, F-21079 Dijon, France; 6) Service de Pédiatrie 1, Centre Hospitalier Universitaire Dijon, F-21079 Dijon, France; 7) Service de Neurologie Pédiatrique, Hôpital de la Timone, Assistance Publique des Hopitaux de Marseille, F-13005 Marseille, France; 8) Unité Mixte de Recherche 910, Institut National de la Santé et de la Recherche Médicale, Aix-Marseille Université, F-13005 Marseille, France; 9) Service de Médecine Infantile 1, Centre de Référence des Maladies Héréditaires du Métabolisme, Centre Hospitalier Universitaire Brabois-Enfants, F-54511 Vandœuvre-lès-Nancy, France.

Background: Whole-exome sequencing (WES) is a powerful tool to identify the molecular basis of clinically and genetically heterogeneous suspected Mendelian disorders. When trio-based WES strategy was preferred for gene identification, we considered proband WES as a cost-effective diagnostic strategy. **Methods:** We present a sample of 41 subjects with epileptic encephalopathy and/or severe intellectual disability (18 females and 23 males), with no diagnosis after the realization of all the conventional indicated investigations. Subjects and relatives were informed and consented for exome testing. WES was performed on index cases. We developed a standardized bioinformatic, interpretive, and validation pipelines for diagnostic WES. Variants were considered when affecting a gene referenced in OMIM as related to the subject disorder. Prioritizations of the variants lead to control familial segregation of pathogenic or probably pathogenic variants, with a maximum of 5 per subject. If no plausible variant was evidenced and if an inheritance mode could be inferred from the familial history, the whole data were considered for interpretation. **Results:** Overall, median coverage was of 86 fold and 93% of the RefSeq exons were covered at least by 10 reads. This strategy allowed the identification of 4 de novo variants (namely affecting *SHANK3*, *ARID1B*, *DYRK1A*, and *TBR1*), 2 X-linked maternally inherited variants (affecting *CUL4B* and *SLC16A2*), 2 autosomal recessive variants with consistent familial segregation (*SCN10A* and *GFER*). The diagnostic yield of this strategy was of 19% (8/41). Interpretation of the whole data in the remaining subjects lead to the identification of variants recurrently affecting *SLC13A5* gene in 2 multiplex families. Besides, candidate genes with consistent familial segregation were identified but not considered as diagnostic results. At last, anonymized pathogenic and candidate variants were uploaded on dedicated repositories (namely ClinVar, www.ncbi.nlm.nih.gov/clinvar/ and Phenome Central, https://phenomecentral.org/). **Conclusion:** In this sample, proband WES was performed as a last resort for 41 subjects. Overall diagnostic yield of this study was of 24% (10/41) and will increase with the scientific community using international repositories. Finally, a medico-economic study is being performed on this sample to evaluate this strategy's cost-effectiveness.

2487M

Whole exome sequencing in patients with intellectual disability. *M. Mila[1,2], MI. Alvarez-Mora[2,1], U. Liljedahl[3], O. Karlberg[3], L. Rodriguez-Revenga[1,2], A. Mur[4], J. Rosell[5,2], E. Guillen[6,2], M. Brannvall[3], AC. Syvänen[3], I. Madrigal[1,2].* 1) Biochemistry and Molecular Genetics Department, Hospital Clínic and IDIBAPS, Barcelona, Spain; 2) for Biomedical Research on Rare Diseases (CIBERER), ISCIII,Madrid,Spain; 3) Department of Medical Sciences, Molecular Medicine, Uppsala University, Uppsala, Sweden; 4) Servicio de Pediatría, Hospital del Mar, Departamento de Pediatría y Obstectricia de la UAB Barcelona, Spain; 5) Universitario Son Espases, Palma de Mallorca, Spain; 6) Genetic Unit, Hospital Clínico Universitario Virgen de la Arrixaca, Murcia, Spain.

Intellectual disability is characterized by substantial limitations in intellectual, functioning and adaptive behavior, diagnosed before 18 years of age. It affects 1-3 per cent of the general population with 50-60 per cent of patients remaining undiagnosed due to the great complexity and the high heterogeneity of the genetic basis. The aims of this study were: to identify mutations in known intellectual disability genes, to identify new genes responsible for intellectual disability and to provide genetic counseling to families. We sequenced 32 individuals of self-reported European ancestry from eight families presenting with moderate to severe intellectual disability using the Illumina HiSeq 2000 Sequencing System. We found four families with novel indel mutations in genes associated with intellectual disability: a compound heterozygous mutation (c.5998delCT; c.10475delAA) in the *VPS13B* gene; a deletion in the *UBE3A* gene (c.2013delG) resulting in an imprinting defect; one autosomal dominant mutation in the *DYNC1H1* gene (c.4461dupC) and one splicing mutation affecting intron 11 of the *IQSEC2* gene (ChrX:53267490_53267491delGT). We also identified two missense mutations in the X chromosome: a novel mutation in the *SMC1A* gene (c.1405C>T) and one previously described in the *OCRL* gene (c.1567G>A). In the remaining 2 families, several candidate variants were identified and require additional studies. In conclusion, these results show the high heterogeneity and difficulty in the clinical diagnosis of intellectual disability and demonstrate the high performance, reliability and cost-effectiveness of this methodology (75 per cent diagnostic success) which is now currently applied in clinical diagnosis. Acknowledements: FP7/2007-2013, grant agreement n°262055 (ESGI project) Premio discapacidad 2012 Fundació Agrupació Mutua.

2488T

Does increase in genomic microarray resolution result in increased diagnostic yield? *S. Costa, A. Krepischi, C. Rosenberg.* Genetics and Evolutionary Biology Dpt., University of Sao Paulo, Sao Paulo, Brazil.

We present data from 1,098 children with intellectual disability, obtained using different microarray platforms with increasing resolution (Table). Genomic imbalances were classified as causative (>1Mb or fulfilling criteria from Genetics in Medicine (2011) 13, 680-685) or Vous (Variants of Uncertain Clinical Significance).*

Array resolution	Patients (n)	Frequency of normal array	Total frequency of alterations	>1Mb	<1Mb causative	<1Mb Vous
44K	57	52.6%	47.4%	35.1%	10.5%	1.8%
60K	780	69.6%	30.3%	21.9%	5.1%	3.3%
180K	191	66%	34%	18.8%	5.2%	10%
850K	70	54.3%	45.7%	21.4%	4.3%	20%
Total	1,098	67.1%	32.9%	22%	5.4%	5.5%

Except for 44K platform (more stringent patient selection) we detected higher frequency of alterations with increased resolution. The difference was due to higher detection of Vous, rather than of causative variants. These results question the benefits obtained by higher resolution arrays in the diagnosis of these patients. Grants: FAPESP/CAPES.

2489S

Investigation by array comparative genomic hybridization (aCGH) in patients with syndromic retinal dystrophies: Preliminary results. *L.R.J. da Silva[1,2], F. Blanco-Kelly[1,3], N. Reyes-Rodriguez[1], C. Vélez-Monsalve[1], M.J. Trujillo-Tiebas[1], I. Sánchez-Navarro[1], A. Avila-Fernandez[1,3], R. Riveiro-Alvarez[1,3], M. Cortón[1,3], C. Ayuso[1,3].* 1) Department of Genetics, Fundación Jiménez Díaz University Hospital (IIS-FJD, UAM), Madrid, Spain; 2) Universidade de Mogi das Cruzes, São Paulo, Brazil; 3) Centre for Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, Madrid, Spain.

Syndromic retinal dystrophies (SRD) are a group of inherited rare diseases with clinical and genetic heterogeneity and high disability, and great social and welfare impact. The current molecular study is complex and lengthy, with approximately 100 genes involved which make finding therapies and clinical management patients difficult. The aCGH is a technology whose implementation is allowing a change in the genetics diagnosis of various diseases. Furthermore, the investigation of amplifications, deletions and chromosomal rearrangements can detect causative factors in the intellectual disability and/or development delay associated with congenital abnormalities. The purpose of this study is to establish the molecular basis of SRD that may be caused by genomic rearrangements using aCGH technology to complete the characterization of patients. Among our cohort of SRD Spanish families, 14 unrelated cases were selected for this study. All of them had undergone previous mutation screening for genes involved in SRD and displayed intellectual disability, suggesting genomic rearrangement. Agilent SurePrint G3 human CGH+SNP microarray 2X400k was performed for all patients, followed by a custom Agilent CGH +SNP 8X60k and other techniques to confirm the results positives. Our study allowed us to preliminary characterize 4 of 14 (29%) with a suspected duplication, mechanism previously unreported for SDR. Microduplications in 1q21.1, 15q13.3, 14q24 and 11q13.2 involving genes associated with intellectual disability, neuropsychiatric disorders and retinal dystrophy were detected. No genomic rearrangement was observed in 8 (57%) patients. Furthermore large loss of heterozygosity (LOH) regions were observed in 2 additional patients (14%) on different chromosomes and different sizes (one patient showed a LOH of 17Mb and the other patient showed 8 different LOH) pointing to possible involved new genes to be further screened. Not only is this study important to help the search for new genes and pathogenic mechanisms associated with SRD, but also to provide a better prognosis and diagnosis, contributing to a appropriate genetic and reproductive counseling. Keywords: aCGH, syndromic retinal dystrophies, intellectual disability.

2490M

Retrospective comparison of SNP microarray and cytogenetics analyses of products of conception. *J.B. Schleede[1], R.D. Burnside[1], J.H. Tepperberg[1], V. Jaswaney[1], I.K. Gadi[1], H. Risheg[2], K.K. Phillips[1], A.L. Penton[1], E. Keitges[1], H. Taylor[1].* 1) LabCorp, Durham, NC., Select a Country; 2) Dynacare/LabCorp, Seattle, Washington.

Spontaneous miscarriage is a common reproductive problem that occurs in approximately 15% of all clinically recognized pregnancies. Since chromosomal abnormalities account for half of the observed first trimester reproductive failures, accurate diagnosis of chromosomal imbalance is paramount to comprehensive patient care. In this report we compare the efficacy of SNP microarray analysis to conventional cytogenetics for the identification of chromosome abnormalities associated with pregnancy loss. Conventional cytogenetic analysis of products of conception (POCs) is dependent upon tissue growth which is subject to contamination, growth failure and overgrowth of maternal tissue. We found that using microarray technology decreased failure rates from 28.7% to 1.4% and reduced the number of artifactual normal female karyotypes (71.1% female vs. 28.9% male) resulting in nearly complete balanced gender ratios. Microarray analysis also markedly improved the detection of chromosomal abnormalities (38% vs. 20% of all samples received). Another advantage over conventional cytogenetics is that SNP microarray genotyping enables the identification of normal karyotype complete moles and differentiation of digynic vs. diandric XXY triploid pregnancies. Accurate diagnosis of diandric partial moles and complete moles identifies individuals at risk of developing gestational trophoblastic neoplasia. The genotyping array also identifies the presence of maternal cell contamination, twin admixture and reveals the presence of UPD and mosaic trisomies associated with incomplete correction.

2491T

Limb-girdle muscular dystrophy 1G: how frequent are mutations in this gene? *V. Nigro[1,2], M.S. Naslavsky[3], N.M. Vieira[3,4,5,6], F. Kok[3,7], M. Vainzof[3], R.C. Pavanello[3], L.M. Kunkel[4,5,6], R. Tupler[8,9], M. Zatz[3].* 1) TIGEM & Patologia Generale, Seconda Univ di Napoli, Napoli, Italy; 2) Lab of Medical Genetics - Seconda Università di Napoli, Napoli, Italy; 3) Human Genome and Stem Cell Center, Biosciences Institute, University of São Paulo, Brazil; 4) Program in Genomics, Department of Pediatrics, Children's Hospital Boston, Boston, USA; 5) Department of Genetics, Harvard Medical School, Boston, MA, USA; 6) The Manton Center for Orphan Disease Research, Children's Hospital Boston, MA, USA; 7) Dept. of Neurology, University of São Paulo Medical School, São Paulo, Brazil; 8) Department of Science of Life, University of Modena and Reggio Emilia, Modena, Italy; 9) Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, USA.

Limb-girdle muscular dystrophies (LGMD) are a heterogeneous group of genetically determined muscle disorders with a primary or predominant involvement of the pelvic or shoulder girdle musculature. More than 30 genes with autosomal recessive (LGMD2A to LGMD2U) and autosomal dominant inheritance (LGMD1A to LGMDH) have been mapped/identified to date. We have recently identified the gene responsible for LGMD1G which had been previously mapped at 4q21 in a Caucasian-Brazilian family. Subsequently, we mapped a Uruguayan family with patients displaying a similar LGMD1G phenotype at the same locus. The inheritance is autosomal dominant, with no evidence of clinical anticipation but asymptomatic carriers were identified in both families. The initial symptoms were proximal lower limbs involvement and weakness followed by upper limbs proximal weakness. Limitation of finger and toe flexion was also a characteristic of this condition. The age of onset varied from 15 to 53 years old (36.6 ± 13.7). Interestingly, 6 out of 11 clinically affected individuals from the Uruguayan family and 2 from the Brazilian family presented cataracts with onset before age 50. None of the patients had myotonic phenomena. Muscle histological features revealed variable findings apparently related to the progression of the disease, with a mixed myopathic/neurogenic pattern. Whole genome sequencing identified, in both families, mutations in the HNRPDL gene. HNRPDL is a heterogeneous ribonucleoprotein (hnRNP) family member, which participates in mRNA biogenesis and metabolism. The coding mutations were found in the same aminoacid residue p.D378. These findings prompted us to search for mutations in a cohort of 400 Italian adult patients with a nonspecific limb-girdle presentation or facioscapulohumeral muscular dystrophy (FSHD), for which a conclusive diagnosis was not yet reached. We developed a targeted NGS analysis of the coding regions of 89 disease genes, among which the HNRPDL. We discovered two adult patients carrying new missense variants not shared by control samples or WES data. Interestingly, the variants replace Leu38Phe and Leu42Phe, but further testing is necessary to understand whether these variant may have functional impact. These data suggest that LGMD1G should have a frequency among LGMD patients <1%, in the Italian population. We are currently screening patients with LGMD or atypical FSHD to estimate the frequency of HNRPDL mutations in the Brazilian population.

2492S

An alternative method for the analysis of deletions/duplications with MLPA©using the QIAxcel© Advanced System. *S.B. Fischer[1], S. Herms[1,3], M. Attenhofer[1], K. Heinimann[1,2], I. Spier[3], S. Aretz[3], S. Cichon[1,2], P. Hoffmann[1,3].* 1) Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland; 2) Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel, Switzerland; 3) Institute of Human Genetics, University of Bonn, Bonn, Germany.

Multiplex Ligation-dependent Probe Amplification (MLPA) from MRC-Holland© is the gold standard for the detection of exonic deletions/duplications in genetic diagnostic testing. MLPA is based on a multiplex PCR, but different from a classical multiplex PCR it is a two-step process involving a hybridisation/ligation reaction for generating the amplification template as the first step. Every exon of a specific target gene is covered by one or more probe-pairs which, when both probes bind to the exon, are ligated. In the second step, the ligated probe-pairs will then be amplified and subsequently analysed using a capillary electrophoresis system. The resulting electropherogram shows a specific peak pattern for each sample. The ratio of the peak height between target and reference probes or even reference samples is used for the detection of exonic deletions/duplications. The fragment analysis is usually performed using capillary (Sanger) sequencing, which is a costly and time consuming method. In this pilot-study we present a faster and more cost effective alternative approach for the MLPA analysis by using the QIAxcel© Advanced System from Qiagen©. The QIAxcel© Advanced System is an automated DNA and RNA analyser, replacing the traditional gel analysis of DNA and RNA. For our pilot study we used two different clinical samples: a) 12 FAP (familial adenomatous polyposis coli) patients with known mutation status as well as 5 healthy controls; b) 5 HNPCC (hereditary non-polyposis colorectal cancer, Lynch syndrome) patients with known mutation status as well as 3 healthy controls. For the MLPA analysis we used the SALSA MLPA P043 APC Probemix© for the FAP samples and the SALSA MLPA P003 MLH1/MSH2 Probemix© for the HNPCC samples. The visualisation of each MLPA-generated amplification product was performed using a Sanger sequencer as well as the QIAxcel capillary electrophoresis system. Utilizing a newly developed protocol we were able to detect exonic deletion/deletions with the same resolution and accuracy as with Sanger sequencing. Based on our preliminary results using the QIAxcel Advanced system allows us to have reliable, fast and cost effective MLPA analysis.

2493M

The SickKids Genome Clinic: a model for paediatric diagnostic and predictive genomic medicine. *M.S. Meyn[1,2,3,4,5], S. Bowdin[2,4,5], N. Monfared[2], D. Merico[1,6], D.J. Stavropoulos[7,9], M. Girdea[1,8], R.K. Jobling[4], R. Hayeems[2,17], M. Szego[10,12], G.D. Bader[3,13], R.D. Cohn[1,2,3,4,5], S.W. Scherer[1,2,3,6,11], J.A. Anderson[14], I. Cohn[15, 16], S. Ito[15], R. Zlotnik Shaul[5,10,14], M. Brudno[1,8], C. Shuman[1,2,3,4], C.R. Marshall[2,6,7], P.N. Ray[1,2,3,6,7].* 1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 4) Division of Clinical and Metabolic Genetics, Dept of Paediatrics, The Hospital for Sick Children, Toronto, Ontario, Canada; 5) Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada; 6) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 8) Department of Computer Science, University of Toronto, Toronto, Ontario, Canada; 9) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, Ontario, Canada; 10) Joint Centre for Bioethics, University of Toronto, Toronto, Ontario, Canada; 11) McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 12) Centre for Clinical Ethics; Providence Healthcare, St. Joseph's Health Centre and St. Michael's Hospital; Toronto, Ontario, Canada; 13) The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada; 14) Department of Bioethics, The Hospital for Sick Children, Toronto, Ontario, Canada; 15) Division of Clinical Pharmacology and Toxicology, Department of Paediatrics, The Hospital for Sick Children, Toronto, Ontario, Canada; 16) Program in Physiology and Experimental Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 17) Program in Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada.

The transformative potential of whole genome sequencing (WGS) lies in its use in individualized genomic medicine, where knowledge of an individual's genomic variants guides health care decisions throughout life, in order to anticipate, diagnose and manage disease. To pilot the implementation of genomic medicine in paediatrics we developed the SickKids Genome Clinic, a multidisciplinary test bed that supports a wide range of research into the clinical uses of WGS, from development of new bioinformatics pipelines and counselling models to health economics and bioethics studies. Our project treats an individual's genome as a source of genetic information to be repeatedly queried over time in order to manage overall health, rather than a one time test performed to diagnose a pre-existing disorder. We are recruiting 150+ children/year who are undergoing diagnostic molecular testing for suspected genetic disorders. DNA and RNA are collected from the child and both parents. Participating children undergo WGS after their parents are counselled about WGS and have declared their preferences for learning their child's pharmacogenetic variants, adult-onset secondary (incidental) medically-actionable variants and carrier status variants.

We enrolled 174 families in our first year of operation. Sequencing now has been completed for 50 children and variant analysis is underway using separate bioinformatics pipelines for primary disorder variants, structural variants, secondary variants and pharmacogenomics variants. Five of the twelve genomes analyzed to date have yielded variants that explain the presenting phenotype, four of which were not detected by the child's clinical genetic testing.

Secondary/incidental medically actionable variants are a major focus of study. ~60% of couples choose to learn their child's adult-onset and carrier variants. For those children we screen variants in >2800 disease genes listed in the NIH Clinical Genomic Database. Using stringent criteria we currently are identifying an average of 7 carrier variants and 1 medically actionable secondary variant/child for Sanger validation and subsequent return.

Our initial results support our hypothesis that WGS is superior to conventional multi-gene panels and clinical microarrays for identifying pathologic variants and can yield reportable carrier and medically actionable variants on every child.

2494T

Progress of Noninvasive Prenatal Testing (NIPT) of Mendelian disorders. *F. Chen[1], Y. Gao[1,2], H. Jiang[1], W. Xie[1], HJ. Jiang[1], J. Lu[1], XC. Li[1], W. Wang[1,2].* 1) BGI-Shenzhen, Shenzhen, Gangdong, China; 2) BGI Health, Shenzhen, Gangdong, China.

Objectives Noninvasive prenatal testing (NIPT) of Mendelian disorders has been reported as case study this year. However, most previous publications focused on sex-linked disorders and autosomal dominant monogenic diseases by assessing the fetal gender and detecting paternal inherited mutation in the maternal plasma. Only few studies showed the feasibility to detect maternal inherited mutations and practical application for autosomal recessive monogenic diseases. Our study aim to report the efficiency of a robust method of NIPT to detect sex-linked, autosomal dominant and recessive Mendelian disorders. Methods In 2012-2013 we recruited eleven families who have a proband child and genetic counseling was provided to each couple. Informed written consent was obtained from each couple. Participants as follows: 1) two families are carriers of hearing loss mutation; 2) seven families have boys with DMD; 3) one family has a 2-year old girl with maple syrup urine disease (MSUD); 4) one family with congenital adrenal hyperplasia (CAH); 5) one family with ichthyosis. Blood samples were obtained at 10-16th gestational weeks. NIPT of Mendelian disorders was performed with a haplotype-assisted strategy previously described, to determine whether the fetus is affected with the same mutation as the proband child or not. Amniocentesis was performed to get the fetal tissues and Sanger sequencing was used to confirm the results. Results Totally in five families, the fetuses were identified to inherit the same alleles as the proband. In three families, the fetal escaped from inheriting any mutations from the parents. In the rest three families, the fetal was identified to inherit the mutations either from the mother or the father. NIPT was finished in 10 days and the results were accordant with Sanger sequencing. Conclusions Noninvasive prenatal testing by maternal plasma DNA sequencing has great potential to supplement the current workflow for prenatal diagnosis of Mendelian disorders, not just for sex-linked, autosomal dominant, but also autosomal recessive. NIPT prior to invasive procedures in the second trimester will offer earlier and trustful information for the pregnant women and help to ease maternal anxiety at very early time.

2495S

Increased yield and detection of confined placental mosaicism by chromosome microarray testing of chorionic villus samples. *J. Reiner, H. Mei, N. Cohen, M. Babcock, A. Babu, S.A. Scott, L. Edelmann.* Mount Sinai Genetic Testing Laboratory, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029.

Early detection of chromosomal abnormalities and genomic imbalances remains a high priority in prenatal care. As such, first trimester chorionic villus sampling (CVS) has preferentially been performed for pregnancies with increased risk for genetic abnormalities despite the possibility of discrepancies arising from confined placental mosaicism (CPM). Although karyotype analysis historically has been the gold-standard for prenatal cytogenetic testing, chromosomal microarray analysis (CMA) increasingly is being used in conjunction with karyotyping and recently was recommended as a first-tier test for cases with abnormal ultrasound findings by the American Congress of Obstetricians and Gynecologists. To determine if the increased resolution offered by CMA influences the identification of CPM, we investigated the relative proportion of cases with confirmed CPM as well as the types of placental aberrations detected by conventional cytogenetics and CMA. In total, 4,632 chorionic villus samples were referred to the Mount Sinai Genetic Testing Laboratory from January 2010 to May 2014, and karyotype analysis, aneuploidy FISH (13, 18, 21, X and Y), and/or microdeletion FISH was performed on all cases. CMA was performed in addition to conventional cytogenetic testing on 13.8% (n=639) of these cases. The total diagnostic yield in this CVS cohort was 6.6% (n=304), which includes 21 suspected pathogenic copy number variants (CNVs) or variants of uncertain clinical significance that were detected exclusively by CMA. Of the 304 total cases with abnormal results, 45 patients (14.8%) elected to have a follow-up amniocentesis for confirmatory testing. Presumed CPM was concluded in 28 cases (62.2%), which represent 0.60% of the cohort. Importantly, eight cases of CPM harbored mosaic structural abnormalities, two of which involved confirmed interstitial CNVs detected only by CMA. Taken together, these data indicate that in addition to being able to detect submicroscopic CNVs, the increased resolution of CMA improves the detection of CPM during prenatal CVS testing. However, these data also suggest that the number of cases that require follow-up by amniocentesis may increase due to previously unidentified mosaic structural abnormalities.

2496M

Sample Identification SNP Panel for Exome, Transcriptome and Whole Genome Sequencing. *K. Duvefelt[1], A. Fungmark[1], T. Svensson[1,3], P. Magnusson[2], V. Wirta[2,3], D. Ekman[3], R. Andeer[3], J. Gruselius[3], M. Käller[3], J. Kere[1,2].* 1) Karolinska University hospital, Stockholm, Sweden; 2) Karolinska Institutet, Stockholm, Sweden; 3) Science for Life Laboratory, Stockholm, Sweden.

A panel consisting of 54 single nucleotide polymorphisms (SNPs) has been developed to enable unique identification of human samples. It will serve as a quality control step ensuring that sequence data generated in high throughput sequencing laboratories corresponds to the originally submitted sample. The panel is designed to analyze exome, transcriptome, and whole genome sequenced samples. The markers were selected based upon the following criteria: -Bi- allelic variations, excluding A-T and G-C transversions -Synonymous variants -Variants informative in the Swedish population -Analyzed successfully in HapMap CEU population -Minor allele frequency >0,3, CEU population -At least 140 bp from exon boundaries -Located towards the 3' end in coding part of transcript -Avoiding regions associated with copy number variations -Not in repetitive sequences -Not in MHC region -In regions with high sequencing coverage -Within transcripts expressed in many tissues -Eight markers for gender identification (ZFXY gene)

128 SNPs was selected using these criteria, successful design were obtained for 112 variations resulting in four pools of SNPs for the Sequenom Iplex Gold multiplex chemistry. Validation was performed on 66 samples, 44 belonged to 14 HapMap trios. Mendelian inheritance correctness, concordance with HapMap genotypes, success rate, background, HWE and robustness of genotyping scoring were evaluated. Based on this data 98 variations in four pools were selected to analyze 90 samples already high-throughput exome sequenced (Agilent Sure Select v5, Illumina HiSeq 2500 rapid run mode) as well as 64 twin samples, formerly scored for zygosity and gender, from the Swedish Twin registry. Analysis has been performed for 16 of the sequenced samples using data from 95 markers that fulfilled the quality criterias. In these variations the concordance between genotyping and sequencing was 100%. All gender markers gave concordant gender identification. Based on percent shared alleles all zygosity pairs scored correctly in the zygosity analysis as well as giving consistent gender identification. Two of the pools described which encompass 54 SNPs including three gender markers will be further used for sample identification.

2497T

Detecting copy number variations on the X chromosome in Chinese children with intellectual disability. *H. Xie[1], X.Y. Li[1,2], J. Guo[1], L.W. Wang[3], J. Wang[3], E.Z. Li[3], J.M. Zhong[2], T. Zhang[1], Y.P. Shen[4], B.L. Wu[4], X.L. Chen[1].* 1) Capital Institute of Pediatrics ,Beijing Municipal Key Laboratory of Child Development and Nutriomic, BeiJing, China; 2) Department of Neurology, JiangXi Children's Hospital, NanChang, China; 3) Department of Neurology, Affiliated Children's Hospital of Capital Institute of Pediatrics, Beijing, China; 4) Department of Laboratory Medicine, Children's Hospital Boston, Boston, MA, 02115, USA.

X-linked intellectual disability (XLID) is a genetically heterogeneous disease characterized by cognitive impairment and reduced adaptive skills. XLID is caused by copy number variants (CNVs) and single nucleotide variants (SNVs) on the X chromosome. To date, >100 XLID genes have been identified, while many more still remain unknown. In this study, seven CNVs on the X chromosome were detected from 212 Chinese children with intellectual disability using array comparative genomic hybridization (aCGH), presenting the yield of 3.3% in Chinese patient cohort. Among them, five CNVs were intellectual disability-related CNVs involving known XLID genes; three CNVs were maternally inherited and two were de novo after analyzing five available parental samples. Four known intellectual disability-related CNVs were observed segregation in the family members, including: recurrent MECP2 duplication, recurrent Xp22.31 duplication harbouring STS, Xp22.13 duplication harbouring NHS, and Xq25-26.1 deletion harbouring OCRL. A novel Xp11.4-11.23 deletion containing CASK, MAOA and KRBOX4 was detected do novo in a girl patient. The MECP2 duplication was identified in a four-generation family, in which five male ID patients and four normal females carry MECP2 duplication. The customer designed X chromosome chip and real-time quantitative PCR confirmed all nine family members carry MECP2 duplication, not MECP2 triplication. Our study promoted to identify the novel XLID genes and mechanisms involved, and showed the phenotypic and genetic heterogeneity of XLID in Chinese patients.

2498S

Miami Otogenetic Program: Implementing genomic medicine in care of patients with impaired hearing. *XZ. Liu[1,2], D. Yan[1], D. Tekin[1], M. Grati[1], S. Blanton[2], M. Tekin[2].* 1) Departments of Otolaryngology-Head and Neck Surgery, University of Miami Miller School of Medicine, Miami, Florida 33136, USA; 2) Dr. John T. Macdonald, Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL 33136, USA.

The Miami Otogenetic Program has provided a unique platform to carry out translational research on delivering genetic services to deafness patient care. Using target-enrichment/next generation sequencing (NGS), we are determining 1) the overall frequencies of different forms of genetic deafness, 2) identifying new genes for autosomal recessive non syndromic hearing loss (ARNSHL) and autosomal dominant (NSHL), and 3) creating important Genomic Deafness Database (GDD) and Personalized Sequence Profile (PSP) for the clinical care of deaf patients where data is ranked based on its clinical validity and utility. We have collected a unique cohort of multiplex families derived from three unique sources from USA, China, and Turkey, suitable for determination of molecular epidemiology of hereditary deafness and for new gene identification using "target-enrichment" methods and NGS. Our interdisciplinary and collaborative team will conduct outcome evaluation of genetic service on deafness patient care in our diversity populations. We have established the Miami Otogenetic Program including the research and the clinical components. The infrastructure of our multidisciplinary otogenetics team has being presented along with our utilization of testing algorithms when evaluating patients with sensorineural hearing loss (SNHL). We have collected DNA samples from over 800 probands from multiplex families with no mutations in the common deafness genes from the three unique cohorts. A total of 60% of small multiplex families are identified to have mutations in the known deafness genes in a pilot study and the remaining 40% have mutations in other yet-unidentified deafness-causing genes. Creation of the Genomic Deafness Database (GDD) and Personalized Sequence Profile (PSP) is in process. Using these high-throughput technologies, we have identified several new genes for non-syndromic deafness. Hearing rehabilitation and counseling of patients with genetic causes of hearing loss are provided. The combined target-enrichment/NGS and whole exome sequencing (WES) is a powerful tool in the identification of new deafness genes in small multiplex families and large multi-generational families. The multidisciplinary team approach is an effective way to bring the sequencing data to clinical practice for the clinical diagnosis and management of deaf and hard-of-hearing families.

2499M

Whole-Exome Sequencing to decipher the genetic heterogeneity of hearing loss in a Chinese family with deaf by deaf mating. *J. Qing[1,3], D. Yan[1], Y. Zhou[3], Q. Liu[1], WJ. Wu[3], ZA. Xiao[3], YY. Liu[3], J. Liu[3], LL. Du[3], DH. Xie[3], XZ. Liu[1,2,3].* 1) Departments of Otolaryngology-Head and Neck Surgery, University of Miami Miller School of Medicine, Miami, Florida 33136, USA; 2) Dr. John T. Macdonald, Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL 33136, USA; 3) Department of Otolaryngology- Head and Neck surgery, Institute of Otology, the Second Xiangya Hospital, Central South University, Changsha, Hunan, 410011, China.

Hereditary deafness is one of Mendelian diseases with a highly genetically heterogeneous trait. For many decades, linkage analysis combined with candidate genes approach have been the main tool to elucidate the genetics of hearing loss. However, this study design is costly, time-consuming and unsuitable for small families, mainly due to the inadequate numbers of available affected individuals, locus heterogeneity, as well as assortative mating. Whole exome sequencing (WES) has now become technically feasible and cost-effective method for detection of disease variants underlying Mendelian disorders due to the recent advances in next-generation sequencing (NGS) technologies. In the present study, we have combined both the Deafness Gene Mutation Detection Array and WES to identify deafness causative variants in a large Chinese composite family with deaf by deaf mating. The simultaneous screening of the 9 deafness common mutations (*GJB2*: c.35delG, c.176 del16, c.235delC, c.299-300delAT; *GJB3*: c.538C>T (p.R180X); *SLC26A4*: c.IVS7-2A>G, c.2168A>G (p.H723R); *mtDNA 12S rRNA*: c.1555A>G, c.1494C>T) in Chinese populations, using the allele-specific PCR based universal array, resulted in the identification of the c.1555A>G in the mitochondrial DNA 12S rRNA in affected individuals in one branch of the family. We then subjected the mutation-negative cases to WES and identified novel causative variants in the *MYH14* and *WFS1* genes. This report confirms the effective use of a NGS technique to detect pathogenic mutations in affected individuals who were not candidates for classical genetic studies.

2500T

Comprehensive Characterization of AML Clinical Samples Using MyAML™: A Novel AML Targeted Sequencing Strategy. *A.R. Carson[1,2], B.A. Patay[2], S.M. Graham[2], A.R. Cubbon[3,4], J.E. Miller[1,2,3,4].* 1) Invivoscribe, Inc., San Diego, CA; 2) Genection, Inc., San Diego, CA; 3) LabPMM, LLC, San Diego, CA; 4) LabPMM, GmbH, Martinsried, Germany.

While next generation sequencing (NGS) is a valuable tool in cancer genomics, most cancer panels target variant hotspots in multiple cancer types. This often compromises the sensitivity and specificity required to accurately identify low-frequency mutations in specific cancers within their clonal contexts. To mitigate this limitation, we designed MyAML™: a comprehensive and specific NGS sequencing strategy used to identify mutations in, and stratify patients with, acute myeloid leukemia (AML). MyAML utilizes oligonucleotide baits that target coding exons of 193 genes and potential genomic breakpoint hotspots within 36 somatic gene fusion partners known or predicted to be involved in AML pathogenesis. We sequenced anonymized AML patient samples from Invivoscribe's biobanks using the MyAML panel. Following DNA hybridization, target loci were sequenced on an Illumina MiSeq utilizing v3 chemistry with the 600-cycle kit to an average depth >500x. Using customized bioinformatics, we performed thorough mutation detection analyses to precisely identify and characterize single nucleotide variants (SNVs), indels, and structural variant breakpoints. We also calculated variant allelic frequencies to investigate potential aneuploidy, loss of heterozygosity and clonality. Within the AML patients, we detected known pathogenic mutations, including known and novel missense SNVs, internal tandem duplications (ITD) in *FLT3* over 300bp in size, and insertions in *NPM1*. Variants found in *FLT3*, *NPM1*, *CEBPA*, *DNMT3A*, *IDH1*, *IDH2*, and *KIT* were validated using either capillary electrophoresis or targeted amplicon NGS. Overall, more than 95% of the mutations detected by MyAML at allelic frequencies as low 5% were validated. In addition to SNVs and indels, we also identified potentially pathogenic gene fusions in patient samples, which were confirmed by FISH or cytogenetic microarrays. In order to advance personalized cancer care, diagnostic assays must be designed to maximize sensitivity and specificity such that they accurately identify all driver mutations in their clonal context. This is critical for predicting response and recurrence with various therapies. MyAML is the only comprehensive panel that optimizes variant detection by specifically targeting the genes and gene fusions known and predicted to be involved in AML pathogenesis. We demonstrate the utility of MyAML as a highly sensitive and accurate sequencing strategy for the comprehensive analysis of AML patients.

2501S

Mutations in the *PNPLA8* gene encoding the mitochondrial calcium-independent phospholipase A2 in a patient with lactic acidosis, spasticity, abnormal gait, dystonia and complex partial seizures. *I. Thiffault[1], C.J. Saunders[1, 4], S.H. Moon[2], X. Liu[2], K. Coffman[3], J.B. LePichon[3], E. Taboada[4], L.D. Smith[1], E.G. Farrow[1], N. Miller[1], S.F. Kingsmore[1, 4], R.W. Gross[2].* 1) Center for Pediatric Genomic Medicine, Children's Mercy Hospitals, Kansas City, MO, USA; 2) Division of Bioorganic Chemistry and Molecular Pharmacology, Department of Medicine, Washington University School of Medicine, St-Louis, MO, USA; 3) Division of Neurology, Dept Pediatrics Children's Mercy Hospitals, Kansas City, MO, USA; 4) Division of Pathology, Dept Pediatrics Children's Mercy Hospitals, Kansas City, MO, USA.

Mitochondrial disorders are a group of clinically heterogeneous diseases, commonly defined by a lack of cellular energy due to OXPHOS defects. The majority of proteins involved in mitochondrial metabolism and maintenance are encoded by nuclear genes, with many yet to be associated with human disease. We performed whole exome trio sequencing on a young girl with mitochondrial myopathy manifested by progressive muscle weakness, hypotonia, spasticity, seizures, poor weight gain, and lactic acidosis. She was found to be compound heterozygous for two frameshift mutations, p.Asn112HisfsX29 and p.Leu659AlafsX4 in the *PNPLA8* gene, which encodes mitochondrial calcium independent phospholipase A2γ (iPLA2γ; MIM#612123). iPLA2s functions in a variety of pathways, including cellular growth, lipid homeostasis, and second messenger generation, exerting their function through the catalysis of the cleavage of acyl groups on glycerophospholipids. The clinical presentation of our patient is reminiscent of the phenotype reported in the mice null for the Pnpla8 gene. The Pnpla8 -/- mouse has cognitive dysfunction, decreased exercise endurance, mitochondrial dysfunction, enhanced insulin sensitivity, growth retardation, cold intolerance, and mild heart failure. As observed in the Pnpla8 -/- mouse model, there are unique morphologic changes observed in mitochondria of this patient's muscle biopsy, which showed an abnormal concentric disarray of internal cristae and globular dense osmiophilic inclusions. In addition, there was a significant increase in secondary lysosomes, many of which contained degenerating mitochondria. Variable fiber size was found with isolated small atrophic fibers of all subtypes; no ragged red fibers were identified. Electron transport chain function was not affected. Oxidative enzyme reactions (SDH and COX) showed no alterations in activity. These findings are consistent with a mitochondrial myopathy. Western blots showed dramatic decreases in multiple bands known to be absent in the Pnpla8 -/- mouse in muscle biopsy tissue from the patient, as would be expected from a null genotype/frameshifts. Although other iPLA2 -related diseases have been identified, such as infantile neuroaxonal dystrophy (MIM#256600) and neutral lipid storage disease with myopathy (MIM#610717), this is the first report of *PNPLA8*-related disease in a human. We suggest *PNPLA8* join the increasing list of genes involved in lipid metabolism associated with neuromuscular diseases.

2502M

Development of a Next Generation Sequencing Panel for Clinical Diagnosis and Prognostication in Hematologic Neoplasm Patients. *M. Middha, L.A. Frederick, Z. Tu, S.A. Henke, B.A. Dukek, X. Wu, L.M. Peterson, E. Klee, R. He, D.S. Viswanatha.* Mayo Clinic, Rochester, MN.

Next generation sequencing (NGS) allows investigation of multiple genes in a single test at a reduced cost compared to traditional sequencing methodologies. We design a targeted OncoHeme NGS Panel of 37 genes known to have mutations in AML, MDS, MPN, CLL, and DLBCL diseases. The genes were selected based on current literature and potential for clinically actionable findings. Here we evaluate the analytical components of the test life cycle process for developing this panel. Using a cross-platform sequencing strategy we sequenced 31 samples on Illumina MiSeq and Ion Torrent PGM instruments. To assess the effect of targeted enrichment, 20 and 11 samples were sequenced using Ampliseq PCR product library with TruSeq Nano adapters and Agilent Sure Select capture methods respectively. Ampliseq PCR products with Ion Torrent adapters were also sequenced on the PGM for all 31 samples. CLCbio Server software was used to analyze all the sequencing data. PGM data was also analyzed using TorrentSuite software to compare performance with CLCbio. We obtained about 5 million sequenced MiSeq reads per sample (average coverage ~4000x) yielding ~55 variants including Single Nucleotide Variants (SNV) and Insertions/Deletions (INDEL). The variants were detected at a minimum coverage of 100x and a minimum variant coverage of 10x. An average of ~1.7 million sequenced reads per sample from Ion Torrent (average coverage ~2000x) yielded ~38 variants using TorrentServer and ~300 variants using CLCbio. In-frame Internal Tandem Duplication (ITD) events ranging from 3 bp to more than 400 bp have been reported in the literature in 10-20% of AML patients, most ITDs are no more than 100bp. However, medium sized INDEL (20-70bp) are challenging to detect in NGS data. CLCbio identified these medium sized INDEL (up-to 42bp) for Ion Torrent sequenced samples but not for the MiSeq sequenced samples. We therefore also analyzed MiSeq data using PINDEL tool that successfully reported these medium sized INDEL. We also evaluated the effect of PCR duplication bias on variant calling using the functionality of 'mark duplicates' and 'remove duplicates' in Samtools, Picard and CLCbio plugin. To allow for testing multiple mutations with high accuracy and sensitivity we here establish an evaluated, well-defined approach for analysis and interpretation of an NGS clinical panel.

2503T

Detection of copy number variations in breast cancer samples using single-nucleotide polymorphism-targeted massively multiplexed PCR. *R.K. Swenerton, B. Hoang, J.E. Babiarz, B.G. Zimmermann, T. Constantin, E. Kirkizlar, N. Wayham, M. Rabinowitz, M. Hill.* Research and Development, Natera, Inc., San Carlos, CA.

Evaluation of copy number variation (CNV) in tumor tissues typically involves SNP microarray or aCGH. These methods have high whole-genome resolution, but require large amounts of input material, have high fixed costs, and have sub-optimal performance on formalin-fixed paraffin-embedded (FFPE) samples. Here, we employ a 28,000-plex SNP-targeted PCR methodology using next-generation sequencing (NGS), targeting chromosomes 1, 2, 13, 18, 21 and X, and regions 4p16, 5p15, 7q11, 15q, 17p, 22q11, and 22q13, to detect CNVs from fresh and FFPE samples. The assay was validated using genomic DNA from 96 human samples with known karyotype, including 71 with deletion or duplication of a region, using a minimum of 3 samples per region; 100% accuracy for copy number was observed. Single-molecule sensitivity was established by analyzing single cells. We determined the utility of this method to detect CNVs in breast cancer using both breast cancer cell lines and fresh frozen (FF) or FFPE breast tumor samples. Analysis of 5 breast tumor cell lines and matched non-tumor cell lines revealed multiple amplifications and deletions (median: 13) in all 5 tumor samples, whereas CNVs were absent in the matched non-tumor control samples. Of 31 FF and 2 FFPE breast tumor samples, 90.3% (28/31) of FF and 100% (2/2) of FFPE samples showed full or partial CNVs in at least 1 and up to 15 regions; of the 30 samples with detected CNVs, 93.3% had a CNV of either 1q or 17p, the two most common breast cancer CNVs represented on this panel. Significantly, CNVs were detected in a large proportion of samples across all evaluated tumor stages: 100% (7/7) of Stage I, 77.8% (7/9) of Stage IIA, 90.0% (9/10) of Stage IIB, and 100% (7/7) of Stage III tumors. Evidence of tumor heterogeneity was also observed. Although not investigated here, the very low DNA sample input requirements will allow this assay to determine tumor heterogeneity at the single-cell level. An assay, using our methodology, analyzing all chromosome arms would require a sequencing cost of approximately $5 per sample. Thus, this method offers a powerful, efficient, and scalable approach for investigating large numbers of FF or FFPE samples.

2504S

Automated miRNA expression profiling in FFPE tissue using nuclease protection coupled with next generation sequencing. *D. Thompson, I. Botros, M. Rounseville, H. Harrison, P. Roche.* HTG Molecular Diagnostics, Tucson, AZ.

Diagnostic pathology primarily uses formalin-fixed paraffin embedded (FFPE) tissues. Large FFPE tissue collections - both current and archival - are available, often with links to patient outcomes. A limitation to the utilization of this resource for molecular pathology is the difficulty of nucleic acid extraction from FFPE samples. Crosslinking-induced fragmentation of nucleic acids in FFPE tissue makes extraction and downstream analysis time-consuming and difficult, especially when scaling up into routine screening.

We have developed EdgeSeq, a coupling of our RNA extraction-free nuclease protection assay (qNPA) with next generation sequencing (NGS) - mediated quantification. Library preparation occurs in two simple steps: automated nuclease protection, performed on the HTG Edge Processor, followed by limited PCR cycles to prepare libraries for NGS. No RNA extraction or enzymatic processing of the sample is necessary. Sequencing of the resulting libraries allows counting for quantification of RNA in the sample. The advantages of NGS for detection include a large dynamic range with high sensitivity.

We initially used the EdgeSeq technique to develop a miRNA whole-transcriptome assay, based on Version 20 of miRBase. Application of the assay shows highly reproducible miRNA profiles are obtained from a variety of sample types including plasma, FFPE, frozen tissues, and cell lysates, with CVs of less than 10%.

EdgeSeq shows excellent correlation (R>0.97) when measuring matched frozen and FFPE cancer tissues. Importantly, poorly-expressed miRNAs are not “lost” when moving into FFPE. The high correlation demonstrates that FFPE tissue can be used in the EdgeSeq assay with full confidence that the results mirror expression in the original (generally unavailable) tissue sample. Further, studies with various FFPE samples demonstrate that tissue-specific miRNA markers are appropriately detected. The EdgeSeq miRNA assay is an excellent tool for NGS-based profiling of miRNA expression, especially for higher-throughput screening of FFPE tissues.

2505M

Non-invasive cell-free tumor DNA-based detection of breast cancer-related copy number variations. *B.G. Zimmermann[1], E. Kirkizlar[1], M. Hill[1], T. Constantin[1], S. Sigurjonsson[2], B. Hoang[1], N. Chopra[1], M. Rabinowitz[1].* 1) Research and Development, Natera, Inc., San Carlos, CA, USA; 2) Statistics, Natera, Inc., San Carlos, CA, USA.

Breast cancer screening involves mammography, which has high false positive rates and misses some cancers. Analysis of tumor-derived circulating cell-free DNA (ctDNA) for cancer-associated copy number variations (CNVs) may allow for earlier, safer, and more accurate screening. Here, we employed a single-nucleotide polymorphism (SNP)-based massively multiplex PCR (mmPCR) approach to screen for CNVs in ctDNA isolated from the plasma of breast cancer patients. The mmPCR assay targeted 3,168 SNPs on chromosomes 1, 2, and 22q, which often have CNVs in cancer (e.g., 49% of breast cancer samples have a 22q deletion). Here, we present case studies of six breast cancer patients: one stage IIa, four stage IIb, and one stage IIIb, with ctDNA fractions of 0.58-4.33%. Each tumor sample had CNVs on one or more of the targeted chromosomes. CNVs were identified in the corresponding plasma samples, including in one stage IIb sample with a ctDNA fraction of 0.58%; detection only required 86 heterozygous SNPs. This demonstrates that focal or whole chromosome arm CNVs, both common in cancer, can be readily detected. To further evaluate sensitivity, we generated 22 artificial mixtures containing a 3Mb 22q CNV, with ctDNA fractions of 0.85-14.7%. The method correctly detected CNVs in 100% of these samples with under $100 sequencing cost per sample. In numerous cancers, such as breast and ovarian, a significant proportion of tumors may only be detectable by CNVs instead of point mutations. Together, this supports that this SNP-based mmPCR approach may offer a cost-effective, non-invasive method for detecting these cancers.

2506T

High sensitivity coupled with low false positives through use of a position and sequence specific error-model in in NGS based somatic DNA sequence variant calling in a targeted cancer panel. *D. Brinza, M. Andersen, B. Johnson, N. Naturaj, K. Norman, K. Rhodes, M. Shahbazian, S. Sadis, P. Williams, P. Wyngaard, Y. Zhan, A. Lee, J. Veitch.* ThermoFisher Scientific, South San Francisco, CA.

High sensitivity combined with very low false positive variant calls are critical requirements in an NGS cancer panel. For the Oncomine® Cancer Research Panel targeting about 3800 known cancer specific variants covering 143 genes and using the Ion PGM™ Sequencer we develop an error model that achieves an average of 0.1 false positives per sample rate in the hotspots and a total of about 5 false positive variants per sample in 250 KB of non-hotspot positions when assessed with 1000 Genome samples. A DNA sample mixture was made by combining cell lines and synthetic linear DNA containing known cancer variants so that 334 hotspots were present in a single sample at 10%; allele frequency. This sample was used to create an Ion AmpliSeq™ library and was sequenced on a Ion 318™ Chip using the Ion PGM™ Sequencer. Using our new error model, a sensitivity of >97%; for SNPs and 75% for indels was observed.

2507S

Detection, estimation and correction of technical effects in copy number estimation using NGS in a targeted cancer panel. *Y. Zhan[1], B. Johnson[1], Z. Zhang[1], C. Van Loy[1], P. Williams[1], P. Wyngaard[1], A. Cani[2], D. Hovelson[2], A. McDaniel[2], C. Liu[2], S. Tomlins[2], S. Sadis[1], M. Andersen[1], J. Veitch[1].* 1) ThermoFisher Scientific, South San Francisco, CA; 2) University of Michigan, Molecular and Cellular Pathology.

We demonstrate a metric “Median Absolute Pairwise Difference” applied to assessment of the per-sample reliability of copy number estimation using the Oncomine® Cancer Research Panel that includes amplicons targeting 49 genes implicated in cancer copy number gains and 26 tumor suppressor genes with cancer related copy number losses using the Ion PGM™ Sequencer. We demonstrate the need for correction of per-sample technical effects including those based on per-amplicon GC content and length. Using mixtures of ATCC cell lines with known copy number changes we observe that detection of genes with copy number >=8 and tumor cellularity >= 30% can be identified with a sensitivity greater than 98% when the correction methods are applied.

2508M

MLPAseq: Assaying genomic copy number variation using multiplex ligation-dependent probe amplification paired with high-throughput sequencing. *D.J. Kvitek, S.A. McCalmon, S. Lin, B. Hsieh, N. McDonald, K.B. Jacobs, J.S. Paul, E.C. Olivares.* Invitae, San Francisco, CA.

As large-scale gene panels have emerged in the clinical genetic diagnostics space, it has become essential to simultaneously query the mutation state of large numbers of genes quickly, with high precision and low cost. Multiplex ligation-dependent probe amplification (MLPA) is the gold standard technique for detecting genomic copy number variation (CNV) in the clinic, but traditional MLPA is limited to detection of 40-50 targets per reaction because individual target peaks must be resolvable from one another on a capillary electropherogram. Here we report a novel extension of the original MLPA technology that enables detection of CNV events using high-throughput sequencing instrumentation, which we term MLPAseq. We demonstrate the simultaneous interrogation of ~1000 targets across 48 samples in a single Illumina Miseq run, and detect duplication and deletion events found in common reference samples and in clinical samples from our diagnostic laboratory. MLPAseq displays high sensitivity and specificity for single- and multi-exon events across >40 clinically relevant genes, including *ATM*, *BRCA1*, *BRCA2*, *CFTR*, *DMD*, *FANCA*, *MEN1*, *MLH1*, *MSH2*, *NPHP1*, *PKHD1*, *PLP1*, *PMP22* and *SMN1*. Specifically, we have used MLPAseq to identify CNVs in *BRCA1* and *BRCA2* with >99% sensitivity and >99.5% specificity, as verified by array comparative genomic hybridization (aCGH) and hybridization capture methods. MLPAseq is low cost, and both laboratory and data analysis workflows are easily automated, making it ideal for primary CNV detection or secondary CNV call confirmation for exome or Mendeliome-sized gene panels in a clinical genetic diagnostics environment.

2509T

Identification of five G6PD common deficiency variants using a novel SNaPshot method in patients of the province of Chiriqui, Panama. *O.I. Batista[1, 2], R.W. Allen[3].* 1) Ctr Gendiagnostik, David, Chiriquí, Panama; 2) CEGEN, Universidad Autónoma de Chiriquí, Panamá; 3) Department of Forensic Sciences, Oklahoma State University, USA.

The clinical, biochemical and genetic consequences of the mutations of the glucose-6-phosphate dehydrogenase (*G6PD*) gene can be important physiologically. G6PD deficiency is an enzymopathy affecting about 400 million people worldwide and biochemical testing of neonates has shown that G6PD deficiency is the metabolic disorder with the highest frequency in the province of Chiriquí, Panama. Affected individuals are usually asymptomatic, and go through life without being aware of their deficiency. They are, however, at risk of having acute hemolytic crises in response to infection, eating fava beans, and to drugs having a high oxidation potential. A SNP typing strategy was designed to simultaneously distinguish six different mutations that defined five G6PD deficiency variants. The variant with its respective mutations are the following: 1) G6PD A (+) (c.376 A>G, p.Asp126Asn); 2) G6PD A (-) (c.376 A>G/c.202 G>A, p.Asp126Asn/p.Val68Met and c.968T>C, p.Tyr222His; 3) Mediterranean (c.563 C>T, p.Ser188Phe); Canton (c. 1376G>T, p.Arg459Leu) and 5) Kaiping (c.1388G>A, p.Arg463His). This approach is based on the single-base extension of an unlabeled minisequencing primer that anneals one base upstream of the relevant SNP. In the presence of DNA polymerase, a fluorochrome-labeled dideoxynucleotide (ddNTP) gets incorporated into the SNP site. A total of 24 individual with low concentration of G6PD enzyme were screened for the six mutations mentioned before. Genomic DNA was extracted from buccal swabs of the deficient subjects using the NucleoSpin® Tissue Kit (Macherery-Nagel). PCR reactions were performed to amplify the exons containing the SNPs from each DNA sample. A six-plex of minisequencing reactions with the SNaPshot multiplex kit (Applied Biosystems) was carried out and analysis of the extension products was performed using capillary electrophoresis. The results were confirmed by restriction fragment length polymorphism (RFLP) or sequencing. The results shown that 84% of the screened patients presented the selected mutations indicating they are highly informative. The G6PD A- variant (72%) is the most frequent, followed by the Mediterranean (8%) and the Canton (4.0%) variants. 16% of individuals did not present any of the studied mutations and further sequence is recommended for them. The approach used has a higher multiplexing capacity, robustness, and extreme sensitivity.

2510S

Application of Serum miRNA Signature for Minimization of Immunosuppression and Diagnosis of Rejection Following Liver Transplantation. *B.J. Keating[1, 2], B. Chang[2, 1], T. Guettouche[2], S. Asare[3], D. Phippard[3], M. DesMarais[4], A. Shaked[1].* 1) Division of Transplantation, University of Pennsylvania, Philadelphia, PA; 2) Children's Hospital of Philadelphia, Philadelphia, PA; 3) Immune Tolerance Network, San Francisco, CA; 4) UCSF, San Francisco, CA.

Over 7,000 liver transplants are performed in the US each year with similar numbers performed worldwide. While advances in immunosuppression therapy (IST) have occurred over the last two decades, rates of acute cellular rejection (ACR) still remain significant, and co-morbidities from ISTs such as nephrotoxicity and liver damage are still major clinical issues. The ability to subclinically identify liver graft recipients who are on ACR trajectories, and those that can successfully withdraw from IST using minimally invasive, reproducible biomarkers with high specificity and sensitivity would represent a major advancement in personalized patient care. We performed miRNA profiling of 752 transcripts on 318 serum samples from 90 liver recipients transplanted from the NIH (ITN)-030 and CTOT03 studies. 48 recipients were randomized to supervised minimization and IST withdrawal, prior to a clinically indicated biopsy event, and were analyzed for prediction of rejection and for identification of those who may tolerate low, or no, IST. Serum miRNA profiles at time of biopsy from 104 samples with and without biopsy proven ACR were compared in a two-stage study. 15 miRNAs were observed to be significantly associated with ACR diagnosis after multiple testing corrections (FDR-adjusted p-value < 0.05). A logistic regression model consisting of a 3 miRNA panel to differentiate ACR from non-ACR with an AUC of 0.90 (95%CI=0.84-0.95), 92.6% sensitivity & 84.2% specificity [p=0.0001]). This 3 miRNA ACR signature was tested in an independent validation set (sera samples from 19 ACR and 16 non-ACR patients), confirming the performance of the model to differentiate ACR from non-ACR (AUC of 0.89 [95%CI: 0.83 - 0.94], 84% sensitivity and 75% specificity, p = 0.01). Statistically significant alterations in this three miRNA ACR panel preceded the rejection event by up to 40 days. The composite score of another distinct 3 miRNA panel early after initiation of IST minimization (at 75% of pre-withdrawal dose) identified recipients able to withstand significant IST reduction (<25% of pre-withdrawal dose, AUC=0.88 [95%CI: 0.80 - 0.95], sensitivity=0.82, specificity=0.90, p = 0.02). The findings in these prospective clinical trials demonstrate that 2 distinct multi-marker miRNAs signatures from sera can be used to: diagnose ACR up to 40 days before manifestation of clinical symptoms; predict rejection trajectories; & guide personalized minimization of IST.

2511M

Next generation sequencing panel for study of genes associated with Epilepsy using a novel single step enrichment and library preparation technology. *Y. Shevchenko[1], C. Naouar[1], K. Harrif[1], C. Chinault[1], L. Chan[2], M. Sanchez[2], S. Silveria[2], J. Dunne[2].* 1) GeneDx, Gaithersburg, MD, USA; 2) WaferGen BioSystems, Fremont, CA.

Epilepsy is a central nervous system disorder (neurological disorder) in which the nerve cell activity in the brain is disturbed, causing a seizure resulting in abnormal behavior, symptoms and sensations, including loss of consciousness. It is a common disorder that affects at least 0.8% of the population and can be caused by genetic disorders, metabolic diseases, trauma, infection, and structural brain abnormalities. Genetic testing for epilepsies is often complex but can be useful in certain clinical scenarios when it can help in clarifying the prognosis, assist in treatment and management of the patient, and predict the risk of a disease in family members. We set out to create a genetic test that will look at the most common causes of infantile seizures and identify mutations in genes that have treatment modifications issues. We target the test towards infants with sudden onset of seizures (possibly intractable) who will often be in the neonatal intensive care (NICU). The results of the test should be delivered within two weeks, the length of a typical NICU stay. To this end we assembled a panel of 18 genes linked with Mendelian forms of epilepsy. A genetic test based on this panel would be conducted in a CLIA lab, upon the evaluation of assay sensitivity, cost, throughput and ease of workflow. Next generation sequencing (NGS) can reduce cost and overhead in genetic sequencing but requires a similar fast, scalable and inexpensive targeted enrichment technology with sensitivity and specificity appropriate for transfer to a CLIA lab. Technologies exist which enrich targeted regions but many suffer from either protracted, expensive workflows, or uneven coverage making them poor choices to proceed to clinical testing. We tested a novel technology, Seq-Ready™ TE, which utilizes a large set of parallel singleplex PCR reactions to perform simultaneous sample enrichment and library preparation. This combined workflow can reduce total preparation time for 8 Epilepsy samples to < 4 hours with minimal hands on time. Amplicons (median size of 235bp) were designed for the 77kb region of interest (ROI) and control samples were enriched in sets of 8 per chip, extracted and purified prior to paired end 2 x 250 bp sequencing run on a MiSeq (Illumina). We will present the results of validation experiments for the Epilepsy panel using Seq-Ready TE technology.

2512T

PACTS: Development, validation and implementation of a low cost, high-throughput combination carrier screen/PGx using targeted nanofluidic PCR and massively parallel sequencing. *J. Buis, S. Birkeland, H. Wang, T. Reeder, B. Tarrier, T. Mann, J. Stoerker.* Research & Development, Progenity, Ann Arbor, MI.

The rapid decrease in the cost of massively parallel sequencing poses both opportunities and challenges for clinical molecular diagnostics. Advances in the clinical interpretation of sequence data has not been able to maintain pace with the ability to develop accurate and expansive targeted gene sequencing tests, leaving an open question as to what are the best practices for mutation screening and clinical reporting of sequence data. Here, we present data on the development and validation of a low cost, high throughput test using the Fluidigm Access Array™ and Illumina HiSeq™ 2500 Rapid Run technology to create a multi-gene sequencing based carrier screen/pharmacogenetic laboratory developed test for cystic fibrosis, spinal muscular atrophy, 18 common Ashkenazi Jewish Disorders, four cytochrome p450 genes important in common drug metabolism pathways and Factor II and Factor V Leiden. Deep multiplexing (up to 768 samples per run) and use of the rapid run technology has made high throughput sequencing on the Illumina HiSeq2500™ cost effective and fast. By combining multiple tests into one and blinding results of unordered tests in the bioinformatics pipeline, we can achieve additional operational flexibility and efficiencies of scale in a production genomic setting. Further, we present a comparative analysis of three different reporting algorithms for cystic fibrosis carrier screening. We provide data on the use of a limited, phenotypically validated set of 154 variants, an expanded clinically disease associated list of variants, and a general variant calling protocol. The data from the first 10,000 de-identified samples is evaluated to give guidance as to the benefits and drawbacks of the three different reporting paradigms.

2513S

Next-Generation Sequencing of the *"Clinome"* in a hypotensive newborn identifies novel mutations in *ACE* of the renin-angiotensinogen system. *H. Daoud[1], C. Beaulieu[2], O. Jarinova[1], N. Carson[1], B. Potter[3], C. for R. Consortium[1,4], J. Majewski[4], D. Bulman[2], P. Chakraborty[5], T. Lacazze[5], K.M. Boycott[1,2], D.A. Dyment[1,2], J. Richer[1].* 1) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; 3) Department of Epidemiology, University of Ottawa, Ottawa, Ontario, Canada; 4) McGill University and Genome Quebec Innovation Centre, Montréal, Québec, Canada; 5) Department of Pediatrics, University of Ottawa, Ontario, Canada.

Next Generation Sequencing (NGS) has significantly improved our ability to make molecular diagnoses for rare disease and there is significant opportunity for its application in the neonatal intensive care setting (NICU). Here we present the case of a premature newborn with anuria and refractory hypotension. After no response to catecholamines, the neonate was able to maintain blood pressure and kidney function on vasopressin. Next generation sequencing of a comprehensive sequencing panel, targeting >4800 clinically-relevant disease genes (termed the "clinome"), identified compound heterozygous pathogenic mutations in the Angiotensin I Converting Enzyme gene (*ACE*, NM_000789.3); c.820_821delAG [p.Arg274Glyfs*117] and c.3521delG [p.Gly1174Alafs*12]. Mutations in this gene are known to cause Renal Tubular Dysgenesis (RTD; OMIM 267430), that results in failure of lung development and fetal, or neonatal, demise. The few long-term survivors with RTD have experienced chronic or end-stage renal disease. Our patient's prematurity (27 weeks) and favorable response to vasopressin potentially has spared the child significant renal pathology. At 7 months chronological age, the patient's renal ultrasound was normal. One month later, his potassium was normal on the kayexalate previously prescribed and he appeared to have nicely recovered from his neonatal acute kidney disease secondary to hypovolemia. The clinome approach provided an answer in 5 days and its focus on disease-related genes permitted pathogenicity-status of the variants to be readily evaluated, without further research study, which has obvious benefits for patients in the NICU setting.

2514M

MASTR Technology based targeted gene panels for comprehensive diagnostic MPS based analysis of inherited and acquired mutations. *J. Del Favero, S. Berwouts, K. Van der Ven, L. Heyrman, D. Goossens, A. Rotthier.* Multiplicom N.V., Niel, Antwerp, Belgium.

Multiplicom's MASTR (Multiplex Amplification of Specific Targets for Resequencing) assays enable multiplex PCR amplification of all required gene regions in a limited number of PCR reactions, resulting in the generation of highly efficient, low cost assays to establish a wide range of clinical and diagnostic applications. MASTR assays employ a simple two-step PCR protocol enabling specific amplification of the regions of interest followed by the incorporation of molecular barcodes in each amplified product to unambiguously link each read to the sample it originated from. This simple and straightforward two-step protocol enables direct compatibility with all commercially available, bench top massively parallel sequencing (MPS) platforms for cost-effective sequencing. Multiplicom currently offers sixteen targeted gene panels for diagnostic predisposition testing of cancer (e.g. breast and colon cancer) and inherited diseases (CFTR, DMD, Marfan,..). Also, MASTR based cancer panels for detection of somatic mutations in tumor tissue that target important cancer therapy related genes linked to many cancers, including melanoma, lung, colorectal, gastrointestinal, prostate, breast and ovarian cancer. Most panels are developed to amplify the complete coding region of cancer genes: the TP53 gene panel; panel containing BRAF, NRAS, KRAS and a panel containing EGFR, ERBB2, PTEN, PIK3CA and PIK3R1. The ability of these assays to use minimal amounts of precious FFPE material allows unlocking a wealth of genomic information for many tumor types using a reproducible and cost effective method. Furthermore, we developed and employed large targeted gene panels, comprising more than 500 exons, for mutation analyses in conditions such as sudden death. Pre-MPS mutation analysis of these conditions was virtually impossible due to large workload, cost and turn around time. Our data show that these large panels allow any diagnostic lab to perform these tests. Also, we showed that a substantial number of patients contain more than one disease causing mutations, not discovered before as a result of cascade based analysis, enabling better genotype/phenotype correlations. For several of these MASTR panels large validation studies have been performed or are ongoing. Results from these studies will be presented and show that multiplex PCR based gene panels enable comprehensive testing of diagnostic relevant genes which can be readily applied to large cohorts of patients.

2515T

Next-generation sequencing as a genetic diagnostic tool to improve the detection of tuberous sclerosis complex (TSC) causative variants. *H. Li[1], T. Wu[2], P. Chen[1,3,4].* 1) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 2) Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 4) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan.

Tuberous sclerosis complex (TSC) is an autosomal dominant disorder characterized with benign tumor growth in multiple organs including skin, brain, heart, lung, kidney, and other organs causing clinical symptoms such as skin abnormalities, developmental and behavioral problems, and lung and kidney diseases. Identification of pathogenetic mutation in ether *TSC1* gene at chromosome 9q34 or *TSC2* gene at chromosome 16p13 has been determined to cause TSC and is sufficient to make a definite diagnosis of TSC. The molecular genetic testing of *TSC1* and *TSC2* has been incorporated into the diagnostic criteria of TSC. However, the fact that *TSC1* gene consists of 23 exons and *TSC2* gene consists of 41 exons spanning in 53 and 45 kb of genomic DNA regions leads the genetic testing of TSC by traditional Sanger sequencing a time and labor-consuming work. In this study, we tested the use of next generation sequencing as the TSC genetic testing tool by taking the advantage of its high throughput to raise the sensitivity of mosaicisms. To enhance the detection of TSC causative variants including large deletion, insertion and inversion, we performed next generation sequencing with target enrichment strategy. The whole genomic regions containing all the exons, introns and 10 Kb of both 5' untranslated region (UTR) and 3'UTR of the *TSC1* and *TSC2* were captured by costumer designed Roche NimbleGen SeqCap EZ Choice Library. The enriched libraries were paired-end sequenced (2*300cycles) by Illumina Miseq system. Bioinformatics tools including BWA, SAMtools, Picard, GATK and ANNOVAR were applied for variants analysis. SIFT and PolyPhen2 were used to predict the biological significance of the genetic variants. Sequence data were further analyzed by Pindel and Breakdancer software to reveal the precise chromosomal breakpoints of the structural variations such as large insertion/deletion (indel) and inversion. All results were visualized by integrative genomics viewer (IGV). Among the 94 TSC families we examined, causative variants in *TSC1* or *TSC2* genes including 51 single nucleotide variants (SNVs), 20 small indels and 8 large indels were identified in total 79 families. Moreover, 2 lineal relatives from 2 TSC families turned out to be mosaic carriers of *TSC2* mutations were also detected in this study. These data suggest that next generation sequencing serving as a genetic diagnostic tool is economical and sufficient to detect the tuberous sclerosis complex causative variants.

2516S

Comparative study for the evaluation of a new technology for cystic fibrosis screening. *M. Majolini[1], M. Rongioletti[1], F. Papa[1], C. Vaccarella[1], A. Luciano[1], C. Centrone[2], B. Minuti[2], V. Mazzucchi[1], M. Belli[1], F. Torricelli[2], G. Liumbruno[1].* 1) Clinical Pathology Department, San Giovanni Calibita Hospital, Rome, Italy; 2) SOD Diagnostica Genetica, AOU Careggi, Florence, Italy.

Introduction Screening for Cystic Fibrosis Transmembrane conductance Regulator (CFTR) gene mutations, is strongly recommended in infertile couples planning a pregnancy by assisted reproductive technology (ART). This study evaluated the performance of the new Nanochip CF70 kit (Savyon Diagnostic, Israel), a microarray assay, and compared it with the Innolipa kits (Innogenetics, Belgium) Methods We analyzed 392 blood samples with Innolipa and Nanochip technologies that identify respectively 70 and 56 CFTR mutations. Discordant results were analyzed with the Devyser CFTR Core Kit (Devyser, AB, Sweden) based on PCR allele specific technology, MLPA (MRC Holland), Direct Sequencing (DS) on the 3730 DNA Analyzer (AppliedBiosystems), and Sequenom's MassArray system (Diatechpharmacogenetics, Italy) Results Innolipa and NanoChip were concordant for 371/392 samples. 21/392 (0.5%) discordant results were tested with the aforementioned technologies: DS confirmed Innolipa results in 18/21 samples and Nanochip results in 1/21, while Devyser and Sequenom did not recognize some mutations not included in their panels. DS was essential for the identification of two different homozygous deletions; although they were not present in Innolipa panels, in 2/21 samples Innolipa indicated a mutation with the warning no interpretation possible Conclusions In this study the Innolipa assay confirmed its reliability and Nanochip showed that it could become competitive with slight changes to the software.

2517M

Clinical application of an anion exchange HPLC column that distinguishes DNA methylation status. *K. Miyake[1], Y. Yamada[2], T. Yotani[2], T. Kubota[1].* 1) Epigenetic Medicine, University of Yamanashi, Chuo, Yamanashi, Japan; 2) Tsukuba Research Institute, Sekisui Medical Co., LTD.

DNA methylation is associated with gene regulation as a component of epigenetic mechanism. Abnormal methylation status alters gene expression, which results in a subset of congenital disorders including Prader-Willi syndrome (PWS) and Angelman syndrome (AS). Methylation-specific PCR (MSP) assay has currently been used for early diagnosis of PWS, which allows early start of growth hormone therapy to prevent obesity and diabetes mellitus. An HPLC column has been used to differentiate amino-acid sequences, and a new anion exchange HPLC column has recently been developed to differentiate DNA sequences, which is used for assessment of SNPs. Using the new column, we tried to differentiate DNA methylation status in PWS and AS patients using their sodium bisulfite-converted DNA samples which were PCR amplified with a primer set of *SNRPN* promoter region. As a result, control normal individuals showed a bimodal peak pattern (an early peak for the methylated allele and a late peak for the unmethylated allele), whereas PWS patients showed a single peak pattern (an early peak for the methylated allele only) and an AS patient showed a distinct single peak pattern (a late peak for the unmethylated allele only). These results were consistent with the results of MSP assay and bisulfite sequencing in the *SNRPN* region. Taken together, the new assay can clearly distinguish between 0, 50, and 100% methylation status. We are now investigating to define its sensitivity for methylation. Since the anion exchange HPLC column-based assay is a rapid and inexpensive method that can be automated on a standard type of HPLC devise, the assay will be a best high-throughput method for assessment of DNA methylation for a large number of patients with epigenetic-associated common diseases such as diabetes mellitus, and thus it will contribute to preemptive medicine of various diseases.

2518T

Development and verification of a Noonan genes Ion AmpliSeq™ panel. *M. Nelen[1], H. Yntema[1], F. Pantaleoni[2], J.C. Machado[3], C. Rossi[4], M. Tartaglia[2], J.L. Costa[3], K. Neveling[1].* 1) Dept. of Human Genetics., Radboud UMC, Nijmegen, The Netherlands; 2) Dept. of Hematology, Oncology and Molec. Medicine, Istituto Superiore di Sanità, Rome, Italy; 3) Ipatimup, Porto, Portugal; 4) Unità Operativa Genetica Medica, Policlinico S. Orsola-Malpighi, Bologna, Italy.

Noonan syndrome is a relatively common autosomal dominant congenital disorder with a high phenotypic variability. It is a clinically and genetically heterogeneous disorder that belongs to the group of Rasopathy diseases, caused by mutations in genes dysregulating the RAS/MAPK pathway. Currently, mutations in 14 genes have been described. Standard analysis using Sanger sequencing is expensive and time consuming. Prenatal analysis of all the genes is often not possible due to limited amount of material the small amount of available fetal DNA often allows the analysis of a maximum of five genes. Here we describe the development of a Noonan genes Ion AmpliSeq™ panel for Ion Torrent sequencing. It includes the coding regions of 14 known Noonan genes, plus 5 basepairs to cover exon-intron boundaries. In total, two pools including 269 amplicons were designed, allowing amplification of all the target regions with only 20 ng DNA. The amplicon sizes range between 125 and 275bps, covering 100% of the desired regions of interest. In a first pilot experiment, pre-and postnatal DNAs carrying known mutations have been sequenced. The average chip loading was 81%, and the average read length was 184bp. Coverage analysis revealed that 266 of the 269 amplicons were covered completely, whereas 3 amplicons show no coverage. The failing amplicons concern (parts of) exon 1 of the genes BRAF, CBL and MAP2K2. No pathogenic mutations have been described for these regions. Coverage depth for the other regions was good with 98% of all amplicons being covered above 100X. Mutation analysis was performed for all samples, and the described mutations (25 different mutations in 12 different genes) could be identified in all of them. The sequencing data of the prenatal samples is very much comparable to the postnatal samples. Further sequencing is ongoing, to reach a total cohort of 100 analyzed samples carrying known mutations, including both prenatal and postnatal DNAs. These preliminary results demonstrate that the Noonan Gene AmpliSeq™ panel may be used in the clinical research setting for prenatal as well as postnatal samples. The simultaneous analysis of all genes is not only attractive in case of limited DNA, but also offers a faster analysis at reduced costs.

2519S

Combined Approach of Targeted Exome Sequencing and Sanger Sequencing to Detect Pathogenic Mutations in Autosomal Dominant Polycystic Kidney Disease. *H. Park[1,2], N. Kim[3], S. Han[4], J. Jang[4], H. Kim[1], H. Jo[1], H. Ryu[1], K. Oh[1], H. Cheong[2,5], C. Ki[6], C. Ahn[1,4], W. Park[2,3,7], Y. Hwang[2,8].* 1) Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea; 2) Research Center for Rare Diseases, Seoul National University Hospital, Seoul, South Korea; 3) Samsung Genome Institute, Samsung Medical Center, Seoul, South Korea; 4) Transplantation Research Institute, Seoul National University, Seoul, South Korea; 5) Department of Pediatrics, Seoul National University Children's Hospital, Seoul, South Korea; 6) Department of Laboratory Medicine and Genetics, Samsung Medical Center, Seoul, South Korea; 7) Department of Molecular Cell Biology, School of Medicine, Sungkyunkwan University, Seoul, South Korea; 8) Department of Internal Medicine, Eulji General Hospital, Seoul, South Korea.

Background: Although genotype is the most important prognostic factor in autosomal dominant polycystic kidney disease (ADPKD), it has not been practical to perform Sanger sequencing due to high cost and technical difficulties. Recently, next generation sequencing method has been introduced in the genetic diagnosis and showed highly accurate and fast diagnostic potential. This study was performed to evaluate the efficacy of combined method of *PKD*1/2 targeted exome sequencing and Sanger sequencing in mutation detection of ADPKD. Method: We extracted DNA from whole blood of 210 ADPKD patients from 177 families. ADPKD was typically diagnosed by family history and imaging with age-dependent criteria. We used Agilent SureSelect Capture Array System for target enrichment and Illumina GAIIx for sequencing. The causal variants were identified and designated as follows: 1) stopgain SNVs, frameshift indels, or listed pathogenic variants in Mayo PKD database (https://pkdb.mayo.edu) as pathogenic, 2) non-frameshift indels or predicted as “damaging” using prediction tools as possibly causal, 3) likely neutral, 4) undetermined, and 5) no candidate. In order to validate mutations found in duplicated regions (exon 1-33 of *PKD*1 gene), we considered the alignments with mapping quality > 20 and further validated each mutation by Sanger sequencing. Results: Among 177 ADPKD families, 147 (83.1%) were found pathogenic or possibly causal variants by targeted exome sequencing alone. When we performed Sanger sequencing for 70 mutations found in the duplicated region of *PKD*1, 69 (98.6%) mutations were confirmed. Since exon 1 of *PKD*1 showed low capture efficiency (<50) due to high GC contents, we additionally performed exon 1 Sanger sequencing in the cases of undetermined significance (n =19) or no candidate (n = 14) and found 3 novel pathogenic mutations. Additional Sanger sequencing for exon 1 in *PKD*1 gene improved mutation detection rate as high as 84.2%. Among 126 pathogenic variants, novel pathogenic mutations comprised 76 (60.3%). Among pathogenic or possibly causal variants, *PKD*1: *PKD*2 were 121:28 (81.2% vs. 18.8%). Conclusion: Combined approach with customized capture-based targeted exome sequencing and additional *PKD*1 exon1 Sanger sequencing may improve mutation detection rate in ADPKD and make up for shortcomings of each method. * This study was supported by a grant of the Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (HI12C0014).

2520M

Development of panel for simultaneous screening of 98 mutations related with hearing loss. *M.C.C.M. Svidnicki, N.Z.P. Santos, S.M. Silva-Costa, G.M. Carvalho, A.M. Castilho, E.L. Sartorato.* State University of Campinas, Campinas, Brazil.

Recent advances in molecular genetics have allowed the determination of the genetic cause of non-syndromic hearing loss, but a large percentage of patients still remain with unidentified cause. It points an imminent need for new methodological strategies for the detection of an increased number of mutations in multiple genes. In this work we developed a panel of 98 mutations, previously identified, in 20 different genes for screening of hearing loss, using mass spectrometry system MassArray, Sequenom® and also evaluated the contribution of selected mutations in the etiology of deafness in Brazilian individuals. To minimize erroneous results, mainly caused by the interaction of primers in multiplex reactions, we tested and standardized the assays using 21 control samples with different mutations in GJB2, SLC26A4 and MT-RNRS1 genes. For every mutation that failed, primers were redesigned and the assays were grouped into different wells. After standardization, the values of sensitivity and specificity of the technique proved to be above 84%. A total 150 unrelated individuals affected with non-syndromic hearing loss were then screened using the developed mutation panel. The possibility of environmental factors causing HL was ruled out by historic information and medical examinations. Mutations were identified in 43 patients (28,7%) and the etiology could be concluded in 16% of the cases. All positive results were validated by other techniques. The c.35delG mutation in the GJB2 gene was the most prevalent, identified in a total of 24 individuals in at least one allele. Other mutations in the GJB2 were identified in 17 subjects, and the p.M34T mutation was the second most prevalent, detected in four cases. Mutations in SLC26A4 gene were found in five individuals, and p.V609G mutation was the most common of this gene, detected in three individuals. Mutations in MT-RNR1, MYO15A, OTOF and CDH23 genes were also identified in one patient each. The genotyping of mutations using mass spectrometry system has been shown to be faster and cheaper than Sanger sequencing and allowed to analyze a larger number of genes than those currently assessed to diagnose hearing loss. Thus, the panel developed presented promise for unraveling the etiology of hearing loss and for genetic counseling.

2521T

Genetic testing of targeted genes using next generation sequencing on PCR amplicons. *W. Yang, J. Yang, P.P.W. Lee, K.W. Chan, Y.L. Lau.* Paediatrics & Adolescent Med, Univ Hong Kong, Hong Kong, 10000, Hong Kong.

Sanger sequencing of PCR amplicons has been the predominant method for genetic testing for primary immunodeficiencies (PID). Despite largely successful endeavors in finding causal mutations for this group of diseases, the method is tedious and inefficient, and suffers from false negative detections under many circumstances. Next generation sequencing (NGS) is rapidly becoming applicable to genetic testing in clinical laboratories and has been shown to be advantageous in many aspects. In this study, we report two real cases in which whole exome sequencing (WES) successfully detected the causal mutations that have been missed by previous tests based on Sanger sequencing of PCR products. In one of the cases, a somatic mutation in the NLRP3 gene that only accounted for 12 percent of the total alleles from peripheral blood mononuclear cells in a patient suffering from chronic infantile neurological, cutaneous and articular syndrome was detected by WES. In another case, a large deletion in DCLRE1C gene missed previously by Sanger sequencing due to sample contamination was also detected by WES. We also tested a method of NGS on pooled PCR amplicons coupled with extensive multiplex identifier (MID) for molecular diagnosis and genetic screening of primary immunodeficiencies and other genetic diseases, which provides flexibility, cost effectiveness, and reasonable automation and throughput (contact yangwl@hku.hk).

2522S

Classification of Incidental Finding Variants in 6503 Participant's Exomes. *L. Amendola[1], P. Robertson[2], D. Nickerson[2], M. Dorschner[2], J. Salama[1], EVS. 6500 annotation consortium[1], G. Jarvik[1, 2].* 1) Divison of Medical Genetics, University of Washington, Seattle, WA; 2) Genome Sciences, University of Washington, Seattle, WA.

Two challenges in genomic medicine implementation are the classification of variant pathogenicity and the evaluation of incidental findings (IFs). The ACMG has proposed discovery and return of IFs for 56 genes from genomic tests. In order to explore expert crowd sourced variant classifications and estimate the rates of IFs, we classified all putatively pathogenic variants in the 6503 participants in the NHLBI exome sequencing project (ESP). The 6503 participants in the ESP included African (n=2,203) and European (n= 4,300, including 187 Ashkenazi) ancestry. We considered variants in 117 genes associated with medically actionable conditions which may be undiagnosed in adults. Briefly, the pathogenicity criteria included the requirement for published studies with multiple individuals and/or segregation or de novo mutation data or novel truncations, where that mechanism is known to cause disease. Using consistent criteria, 48 experts reviewed 628 variants that were classified as 'disease causing mutations' in the Human Genetic Mutation Database (HGMD) and had minor allele frequency < 0.005 for autosomal dominant and x-linked (612) and <0.1 for autosomal recessive (16) disorders. 67 stop and splice variants found in ESP but not in HGMD were also considered. Initially, a random 156 (25%) of the HGMD variants were blindly double-reviewed; of these 81 (52%) did not match and 53 mismatches involved pathogenic or likely pathogenic classification. We then blindly reclassified all missense variants classified as pathogenic or likely pathogenic: 56% changed classes (44/79) with 42 of 44 reclassifications moving from pathogenic or likely pathogenic to variant of uncertain significance The frequency of pathogenic, likely pathogenic, and expected pathogenic novel disruptive mutations were 0.6, 1.0, and 0.1% in those of European ancestry and 0.2, 0.5, and 0.4% in those of African ancestry. Relationships of classification to allele frequency and in silico scores were also explored. The deficit of findings in subjects of African ancestry may reflect the lack of relevant literature. 87% of HGMD 'disease-causing' mutations did not meeting our criteria for pathogenicity. Even with expert reviewers and clear criteria, there was substantial inconsistency in the classification of variants that was marked by over interpretation of variant pathogenicity. This suggests that crowd sourcing of variant annotation has limitations.

2523M

Detection, characterization, and validation of Hematological Malignancies with Archer(TM) FusionPlex(TM) Heme NGS assay. *B. Culver, B. Kudlow, J. Stahl, M. Callan, J. Amsbaugh, J. Myers.* Enzymatics, Boulder, USA.

Introduction The importance of understanding gene fusions in cancer biology is increasingly important, and may have profound impacts on next-generation drug discovery efforts. Historical testing methods such as fluorescence in-situ hybridization (FISH) are limited by scalability and subjective interpretation. Growing numbers of fusion gene partners to MLL, JAK2, and other oncogenes in hematological malignances have further complicated matters. Targeted sequencing technologies, such as Anchored Multiplex PCR (AMP), have the ability to simultaneously interrogate hundreds of targets and identify novel and known gene fusions from small amounts of input material in a single reaction. In this study, we will demonstrate the detection, breakpoint characterization, and concordance between current technologies and the targeted Archer(TM) FusionPlex(TM) Heme next-generation sequencing panel. **Materials and Methods:** Heme fusion samples were sourced from CAP-accredited tissue repository (Asterand) and through collaboration with Massachusetts General Hospital. Samples with known/unknown Hematological malignancies, as determined by existing technologies such as FISH, RT-PCR, and non-targeted sequencing, were provided for the study. The Archer FusionPlex Assay was performed according to Enzymatics package insert, and sequenced on the Illumina MiSeq platform. Sequencing analysis was performed via the Archer Analysis Pipeline, and reported fusion results were compared to current methods. **Results:** Concordance between current methods and the FusionPlex NGS results is high, with all positive samples also having a positive NGS result. In contrast to current methods, the Archer FusionPlex assay was able to characterize fusion partners previously unidentifiable via RT-PCR or other methods without prior knowledge of fusion partner or the laborious and high cost of RNA-seq screening. **Conclusions:** The Archer(TM) FusionPlex(TM) Heme panel is a powerful new tool for identifying disease relevant gene fusion events from small amounts of input material in Hematological cancers. Our initial results indicate that this targeted sequencing assay shows promise in matching the "gold standard" of FISH testing, while overcoming the obvious limitations of reflex testing. Furthermore the novel AMP-based sequencing method allows for cost-effective high-throughput identification of fusion breakpoints and fusion partners to the relevant oncogenes without prior knowledge of the partners.

2524T

Enabling Genomic Clinical Variant Assessment and Reduction of Variants of Unknown Significance through Intelligent Scoring. *C.L. Mead[1], M. Kallberg[1], T. Mann[2], S.S. Ajay[1], E. Ramos[1], D. Mote[1], S. Kruglyak[1], T.M. Hambuch[1].* 1) Illumina Clinical Services Laboratory, Illumina, San Diego, CA; 2) Progenity, 5230 S. State Road Ann Arbor, MI 49108.

Barriers to accurate human genome re-sequencing have largely been surmounted, enabling consideration of routine clinical applications. However, significant challenges remain for interpretation of whole genome data for diagnostic purposes. The shift from genetic to genomic scale testing significantly increases the interpretation burden, and yet there is a lack of cohesive information to facilitate genome scale variant assessment. Clinical assessment of whole genome sequencing (WGS) results based on literature alone produces numerous variants of unknown significance (VUS), which do not yield actionable information for clinicians. For Mendelian conditions, the relationship of disease prevalence to variant prevalence is well established and this relationship can be used to identify variants that can be confidently excluded from disease involvement. Our algorithm generates a statistical score specific to the characteristics of the variant and disease being assessed. We have established a clinical service for WGS and interpretation. The interpretation service is currently limited to exons +/- 15bp of 1600 genes whose clinical validity and involvement in 1211 monogenic conditions have been established. Here we describe our scoring algorithm that generates a statistical score for each variant to help rule out disease causality using information including disease penetrance, prevalence, inheritance mode, and variant frequency. We sequenced and interpreted 1600 genes in approximately 500 individuals and identified over 60,000 unique variants within exons +/- 15 nucleotides. Statistical analyses were performed to determine thresholds whereby variants were confidently ruled out from being associated with disease, thereby significantly reducing the number classified as VUS. WGS methods result in a significant variant assessment burden that must be addressed in order to enable effective integration into medical practice. While peer-reviewed literature sources provide the strongest evidence for variant assessment, this information is currently scarce on the whole genome level. In the context of variant assessment for monogenic diseases, where literature information is not available, our understanding of biology can enable the development of algorithms that are more refined and disease specific, which can improve interpretive reporting. A quantitative statistical score can help in the evaluation of the clinical variant assessment confidence.

2525S

Sensitive and Comprehensive Method to Detect Mutations in RB1 Gene Improves Care for Retinoblastoma Patients and Their Families. *W. Li, J. Buckley, P. Sanchez, D. Maglinte, L. Viduesky, T. Triche.* Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Los Angeles, CA.

Retinoblastoma is childhood eye cancer caused by inactivating mutations in both alleles of the tumor suppressor gene, RB1. RB1 is the only gene in which mutations are known to cause heritable predisposition to retinoblastoma. Retinoblastoma can lead to the loss of vision, eye(s), and sometimes even life. Molecular identification of a germline RB1 mutation in a timely manner is very important for the effective care and management of retinoblastoma patients and their families. However, current methods to detect disease causal mutations in RB1 are lengthy, usually taking weeks or months. In this study, we report a next-generation sequencing (NGS) based method that is capable of capturing pathogenic mutations in the entire 200 kb RB1 gene, both exons and introns. The process of DNA extraction, sequence data generation, and completion of data analysis requires only 3 days. In addition, the associated bioinformatic pipeline can take NGS data and accurately measure the copy number of the RB1 gene and MYCN gene, thus enabling the NGS method to simultaneously detect point mutations, small InDels and large deletion/duplications in the RB1 gene and gene amplification of the MYCN gene on a single test platform, which was impossible previously. Because of the deep sequencing coverage of the entire RB1 gene, this method can detect even a low percentage of mosaic mutations directly from blood samples, which is essential for genetic counseling of retinoblastoma patients and their families. To assess the functional impact of mutation and other RB anomalies like promoter hypermethylation, we are developing a novel targeted NGS assay to simultaneously assay both RNA and DNA, in order to measure the transcript level of expression and DNA mutational status in the same extract. This will be complemented by development of a targeted methylation assay to complete DNA, RNA, and epigenetic profiling of the RB1 gene. Taken together, we believe that these combined NGS based methods can lead to rapid, inexpensive, comprehensive, and sensitive methods for detecting RB1 mutations and aberrant expression. This in turn will enable detection of germ line carriers who are at lifetime increased risk of a second malignancy like osteosarcoma, a decidedly more aggressive and more often fatal disease compared to retinoblastoma.

2526M

A multi-platform amplicon sequencing method for fast and reliable variant detection. *M. Toloue, L. Matzat.* NGS, Bioo Scientific Corp, Austin, TX.

The application of NGS to clinical work and patient diagnoses has been limited by several factors including cost, time, and limitation of samples. Rapid benchtop sequencing as well as multiplexing reduce both cost and time limitations of sequencing; however, sample preparation for these platforms remains a rate limiting step. Whereas target capture methods require construction of a DNA-seq library and a target capture process, amplicon libraries can be constructed by simple PCR steps that sequentially build platform-specific and barcode sequences onto target sequences. Here we present an amplicon sequencing library construction method that produces libraries in highly multiplexed reactions and is compatible with multiple platforms. Our library preparation allows accurate targeting and even coverage across amplicons. Data will be presented demonstrating amplicon coverage and SNP detection as well as the flexibility for customized amplicon panel design.

2527T

A Comprehensive Profile of Hereditary Myopathies by Next Generation Sequencing on 43 Early-Onset Patients and Subsequent Development of a Biomarker Assay by Liquid Chromatography-SRM-Mass Spectrometry. *S. Hahn[1,2,3], V. Vasta[3], S. Jung[3], Q. Zhang[4], S. Eun[5], A. Cho[6], B. Lim[7], J. Chae[7].* 1) Department of Pediatrics, University of Washington School of Medicine, Seattle, WA; 2) Seattle Children's Hospital, Seattle, WA; 3) Center for Developmental Therapeutics, Seattle Children's Research Institute, Seattle, WA; 4) Fred Hutchinson Cancer Research Center, Seattle, WA; 5) Department of Pediatrics, Korea University College of Medicine, Seoul, Korea; 6) Department of Pediatrics, Ewha Woman's University School of Medicine, Seoul, Korea; 7) Department of Pediatrics, Seoul National University School of Medicine, Seoul, Korea.

Hereditary myopathies encompass a significant proportion of patients with chronic muscle disease. The diagnostic approach for hereditary myopathy is often difficult due to its heterogeneous genetic background and similar clinical presentations. Most patients require a muscle biopsy for the diagnosis, which is particularly challenging in infants or newborns. Unfortunately, many patients remain undiagnosed due to the lack of specific markers for many myopathies or due to ambiguous pathological results. Furthermore, Sanger sequencing of individual genes is challenging because one gene can cause a wide variety of clinical and/or pathological features, while similar clinical features can be caused by mutations in different genes. Here, we explored Next Generation Sequencing panel (~600 genes) on 43 patients with early-onset myopathy, targeting known pathogenic genes for hereditary myopathies or secondary muscle weakness. Fourteen novel/rare variants, fifteen known pathogenic variants and five VUS in 17 genes were identified in thirty two patients. Three had myasthenic myopathy (*DOK7, AGRN*, and *GFPT1*); two had peripheral neuropathy (*GARS, DYNC1H*). Incomplete penetrance was strongly suspected in four cases with variants in *COLA1, RYR1, CCDC78* and *MYBPC3* genes, indicating that careful interpretation is required in the context of clinical, laboratory and pathological findings when the rare variant is inherited from asymptomatic parents. As *COL6A* genes were most commonly affected in our cohort while no specific biomarkers are clinically available, we explored a Liquid Chromatography Selected Reaction Monitoring Mass Spectrometry (LC-SRM-MS) to identify and quantify the proteotypic peptides of Collagen VI in skin fibroblasts. We identified several candidate signature peptides for Collagen VI and are currently evaluating the method to see if the abundance in patient's fibroblasts is reduced or absent. This assay can open a valid experimental model to study the candidate drugs that are becoming available. Although our study indicates that simultaneous sequencing of multiple genes for various types of hereditary myopathies is clinically relevant, a substantial number of patients still remained without molecular diagnosis in our cohort implying that new causative genes for hereditary myopathy are to be identified. Further development of biomarker assay for hereditary myopathies is necessary and could be feasible with LC-SRM-MS in selected conditions.

2528S

Genetic Diagnosis of Duchenne Muscular Dystrophy by Clinical Exome, Whole Genome and Transcriptome Sequencing. *R.T. Wang[1,2], H. Lee[1,3], A. Eskin[1], V. Arboleda[1], K. Squire[1], J.L. Deignan[3], N. Khanlou[3], P.B. Shieh[4], S.F. Nelson[1,2,3].* 1) Human Genetics, University of California Los Angeles, Los Angeles, CA; 2) Center for Duchenne Muscular Dystrophy, University of California Los Angeles; 3) Department of Pathology and Laboratory Medicine, University of California Los Angeles; 4) Department of Neurology, University of California Los Angeles.

Duchenne muscular dystrophy (DMD; OMIM #310200) is the most common form of muscular dystrophy characterized by a progressive loss of skeletal muscle function. Roughly 1 in 5000 male births are affected worldwide. The X-linked *DMD* gene is one of the largest genes in the human genome and mutations within the 2.2 Mb gene are the unequivocal cause DMD. The mutational spectrum of DMD is heterogeneous with approximately 60-65% of affected individuals with large exonic deletions, 5-10% with large duplications and 25% with single base, small indels, or splice site mutations typically resulting in absent dystrophin expression. Although the gene responsible for DMD has been known for over 25 years, as many as 30% of affected boys have not received a molecular diagnosis in the United States, demonstrating some barriers to molecular diagnosis. Multiple ligation dependent probe amplification (MLPA) or array CGH, reliably identify exonic deletions and duplications through probe hybridization targeted to specific exons. When MLPA fails, Sanger sequencing of exons and splice sites is undertaken. The large size of the *DMD* gene, however, makes comprehensive assessment of all exonic mutations challenging and expensive, and a small portion of histologically diagnosed patients do not have a detected mutation. Here we report the use of whole exome sequencing as a clinical test to detect point mutations, small insertion and deletions, and single or multiple exonic deletions and duplications in the *DMD* gene simultaneously with high sensitivity and specificity. By sequencing to high depth of coverage, we reliably identified the disease causing point mutations, exonic deletions and duplications for 27 of 30 samples. We utilized paired end whole genome sequencing to identify duplicated breakpoints for two samples and further employed RNA-seq on biopsied muscle tissue coupled with whole genome sequencing in two refractory cases to uncover a novel genomic rearrangement and an unusual change in splicing with base-pair resolution.

2529M

Hi-Plex: a flexible, streamlined and cost-effective approach to targeted massively parallel sequencing. *T. Nguyen-Dumont[1], B.J. Pope[2,3], F. Hammet[1], M. Mahmoodi[1], H. Tsimiklis[1], M.C. Southey[1], D.J. Park[1].* 1) Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Parkville, VIC, Australia; 2) Victorian Life Sciences Computation Initiative, Carlton, VIC, Australia; 3) Department of Computing and Information Systems, The University of Melbourne, VIC, Australia.

Purpose: Massively parallel sequencing (MPS) has revolutionised biomedical research and offers enormous capacity for clinical application. We previously reported Hi-Plex, a streamlined highly-multiplexed PCR-based target-enrichment system for MPS. Here, we present its optimization to 1,000 amplicon-order multiplexing and gene panel screening of thousands of specimens.

Methodology: Hi-Plex library-building consists of a single-step, highly-multiplexed PCR, followed by size selection, using relatively inexpensive, readily available reagents in a simple half-day protocol. A given library can be sequenced with both the TruSeq and Ion Torrent chemistries. Hi-Plex does not require normalisation of barcoded libraries before pooling and sequencing. The Hi-Plex system is complemented by simple, automated primer design software that allows the uniform definition of library size, so that subsequent paired-end sequencing can achieve complete overlap of read pairs. Variant calling from Hi-Plex-derived datasets relies on the identification of variants appearing in both reads of read-pairs, permitting stringent filtering of sequencing chemistry-induced errors using our ROVER software. Comparable sequencing efficiency was achieved using material derived from lymphoblastoid cell lines and formalin-fixed paraffin-embedded tumours.

Results: Initially reported in a 60 amplicon multiplex (60-plex) setting, Hi-Plex's chemistry was improved to perform at the ~1,000-plex level. Using equal amounts of all primers, we observed that 95% amplicons were represented within 20-fold of the median coverage. Using the improved Hi-Plex chemistry and expanding on a published demonstration of Hi-Plex capabilities to perform accurate, high-throughput screening, we screened 1,285 breast cancer cases for mutations in the breast cancer susceptibility genes *PALB2* and *XRCC2*.

Conclusion: Hi-Plex is simple, accurate, low cost, modular and flexible in terms of target region design and sequencing platform. These features render the approach highly attractive for an extensive range of clinical and research applications.

2530T

GenoDENT: A targeted next-generation sequencing assay for the diagnosis and discovery of mutations in buccodental disorders. *M. Prasad[1], V. Geoffrey[1], V. Haushalter-Laugel[2], M. Paschaki[1,2], B. Gasse[3], C. Stoetzel[1], H. Dollfus[1,4], A. Bloch-Zupan[2,5,6].* 1) Laboratoire de Génétique Médicale, University of Strasbourg, Strasbourg, France; 2) Institute of Genetics and Molecular and Cellular Biology (IGBMC), Centre National de la Recherche Scientifique (UMR 7104), Institut National de la Santé et de la Recherche Médicale (U 964), University of Strasbourg, Illkirch, France; 3) Research group "Evolution et Développement du Squelette", UMR 7138-Evolution Paris-Seine, Institut de Biologie Paris-Seine, Université Pierre et Marie Curie, Paris, France; 4) Centre de référence pour les Affections Rares en Génétique Ophtalmologique (CARGO), Hôpital Civil, Strasbourg, France; Service de Génétique Médicale, Hôpital de Hautepierre, Strasbourg, France; 5) Faculty of Oral Medicine, University of Strasbourg, Strasbourg, France; 6) Reference Centre for Orodental Manifestations of Rare Diseases, Pôle de Médecine et Chirurgie Bucco-dentaires, Hôpitaux Universitaires de Strasbourg (HUS), Strasbourg, France.

Genetic disorders of the oral cavity and teeth, whether present in isolation or as part of a syndrome, encompass a wide range of disease phenotypes and numerous genetic loci. Furthermore, several of the constituent phenotypes show considerable locus heterogeneity, with evidence for the involvement of novel loci, thus rendering the molecular diagnosis of these phenotypes challenging. We have developed a targeted next-generation sequencing assay, GenoDENT, for the diagnosis and discovery of mutations in 560 known and candidate genes in orodental diseases. We demonstrate the utility of this assay for the molecular diagnosis of a wide variety of syndromic and non-syndromic dental disorders, such as amelogenesis imperfecta, dentinogenesis imperfecta, selective tooth agenesis, and ectodermal dysplasia. In eight patients with a known molecular diagnosis, GenoDENT was able to identify 100% of the mutations, including a ~3 Mb heterozygous deletion on the X chromosome. The target enrichment allowed a mean coverage of 371X, with 95% of targeted nucleotides being covered at ≥50X. Furthermore, in eight patients without a molecular diagnosis, we were able to identify known and/or novel mutations in known genes in 6/8 patients (75%). These included a frameshift mutation in *DSPP* (dentinogenesis imperfecta), missense mutations in *WNT10A* (selective tooth agenesis) and *EDA* (ectodermal dysplasia), and nonsense and missense mutations in *WDR72* (amelogenesis imperfecta). The mutations were validated by Sanger sequencing and their pathogenicity was established by segregation analysis in family members for whom DNA was available. Bioinformatic analyses predicted the novel missense mutations identified to be pathogenic. In the two patients in whom no mutations were identified in known genes, the GenoDENT pipeline allowed us to prioritize potentially novel pathogenic loci for further investigation. In conclusion, we have developed a reliable targeted sequencing assay for the genetic diagnosis of a wide variety of orodental disorders. This assay also provides the potential to identify novel loci underlying rare orodental disorders. This EU-funded project (ERDF) A27 "Oro-dental manifestations of rare diseases", is supported by the RMT-TMO Offensive Sciences initiative, INTERREG IV Upper Rhine program www.genosmile.eu.

2531S

A new approach to target capture for genetic testing: a novel DNA enrichment method for next-generation sequencing applied to clinically-relevant genes. *C. Richard[1], Y. Bei[1], B. Galvin[1], L. Apone[2].* 1) Directed Genomics, Ipswich, MA; 2) New England Biolabs, Ipswich, MA.

As next-generation sequencing expands from basic research into clinical applications, target enrichment protocols need to be simple, robust, and cost effective. To address this need, we have developed a new technique for target capture that offers significant advantages over traditional in-solution hybridization and multiplex PCR protocols. This method utilizes rapid hybridization of short probes to genomic DNA or cDNA fragments, to define the regions of interest, followed by enzymatic removal of off-target sequence, ligation of NGS platform specific adaptors, and PCR amplification. This method can individually tag DNA molecules with unique identifiers for duplicate filtering, is easily performed within one day, and is compatible with FFPE samples. To demonstrate this technique, we have designed a panel to capture over two hundred exons from oncogenes and tumor suppressor genes, covering several thousand COSMIC mutations. Analysis of the libraries sequenced on an Illumina MiSeq demonstrates the high specificity and uniformity of the target sequences generated by this approach.

2532M

Nijmegen breakage syndrome detected by newborn screening for T cell receptor excision circles (TRECs). *A.N. Adhikari[4], J. Patel[1], R.A. Gatti[2], C. Brown[2], U. Sunderam[3], K. Kundu[3], R. Srinivasan[3], S.E. Brenner[4], J.M. Puck[5], J.A. Church[6].* 1) Department of Pediatrics, Children's Hospital of Los Angeles, Los Angeles, CA; 2) Departments of Human Genetics and Pathology & Laboratory Medicine, David Geffen University of California Los Angeles School of Medicine, Los Angeles, CA; 3) Innovations Labs, Tata Consulting Services, Hyderabad, AP, India; 4) University of California, Berkeley, CA; 5) Department of Pediatrics and Institute for Human Genetics, University of California, San Francisco, CA; 6) Department of Pediatrics, Keck School of Medicine, University of Southern California and Children's Hospital, Los Angeles, CA.

Severe combined immunodeficiency (SCID) is a group of disorders characterized by reduced or absent T-cell number or function. Newborn screening for SCID utilizes quantification of T cell receptor excision circles (TREC). These circular DNA byproducts of thymic T cell receptor rearrangements correlate with numbers of circulating T cells. Although TREC screening was developed to identify patients with SCID, it has identified T lymphopenia in newborns with disorders that may not otherwise present until later in life. We present a case of an infant patient with Nijmegen breakage syndrome, an autosomal recessive disorder characterized by microcephaly, a distinct facial appearance, short stature, immunodeficiency, radiation sensitivity, and a strong predisposition to lymphoid malignancy. Low newborn TREC led to confirmatory testing that demonstrated persistently low T cells from birth. Whole exome sequencing (WES) was performed on the infant and his healthy parents to look for causative genetic variant(s) potentially associated with the observed phenotypes in the infant.

The sequencing data were processed through our analysis pipeline. The first several steps employed standard tools for mapping and variant calling, but integrated two different carefully-tuned callers to yield high quality sets of variants. Comprehensive variant annotation was performed to flag the compound heterozygous, de novo, uniploidal disomy, and X-linked recessive variants with potential clinical importance. After filtering the variants and prioritizing genes based on an immunological disorder associated gene list, we found two previously unreported compound heterozygous nonsense mutations in the gene for nibrin (*NBN*) (c.1140C>T;p.Gln344*, c.952T>G;p.Leu281*), predicted to cause nonsense-mediated mRNA decay of transcripts from both alleles in the infant patient. The identified *NBN* variants were confirmed experimentally using Sanger sequencing in a CLIA-approved Medical Genetics laboratory. Immunoblotting showed absence of nibrin protein and a colony survival assay demonstrated radiosensitivity comparable to patients with ataxia telangiectasia. Thus, WES after abnormal SCID newborn screening made the early diagnosis of Nijmegen breakage syndrome.

2533T

Establishing distinctive criteria for reporting genomic sequencing results in healthy versus ill newborns: The BabySeq Project. *O. Ceyhan-Birsoy[1, 2, 3], S.S. Kalia[4], P.J. Park[3], T.W. Yu[2, 3, 5], P.B. Agrawal[2, 3, 6], R.B. Parad[7], I.A. Holm[2, 3], A. McGuire[8], R.C. Green[1, 4], A.H. Beggs[2, 3], H.L. Rehm[1, 9].* 1) Laboratory for Molecular Medicine, Partners Center for Personalized Medicine, Cambridge, MA; 2) Division of Genetics and Genomics, Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 3) Department of Pediatrics, Harvard Medical School, Boston, MA; 4) Department of Medicine, Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 5) Department of Neurology, Boston Children's Hospital, Boston, MA; 6) Division of Newborn Medicine, Boston Children's Hospital, Boston, MA; 7) Department of Pediatric Newborn Medicine, Brigham and Women's Hospital, Boston, MA; 8) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 9) Department for Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

Newborn genomic sequencing may enable detection of a wide range of conditions for which knowledge can improve health outcomes. At the same time, to mitigate the risks of ambiguous information, distinct criteria need to be defined for return of results in healthy versus ill newborns. We developed different genome analysis protocols and reporting strategies for these two populations in the BabySeq project, a randomized controlled trial exploring the use of newborn genomic sequencing. 240 healthy newborns and 240 sick newborns in the NICU will be enrolled and randomized to receive exome sequencing. All newborns in the sequencing group will receive a Genomic Newborn Sequencing Report (GNSR) with variants in genes that have strong evidence to cause a highly penetrant childhood-onset disorder. To determine which genes meet these criteria, over 1,000 genes have been curated and categorized with respect to multiple attributes, including strength of evidence for disease association, mode of inheritance, age of onset, and penetrance. To minimize risks associated with reporting difficult-to-interpret variants, only pathogenic and likely pathogenic variants will be returned on the GNSR. In addition to the GNSR, newborns in the NICU will also receive an Indication Based Genome Report (IGBR), which will focus on genes with strong or moderate evidence for the indication. We have created lists of genes associated with 20 of the most common presentations in the NICU, including hypoglycemia, seizures, and hypotonia. If results from these predefined lists are non-explanatory, genes with novel loss-of-function or biallelic rare variants will also be considered. Consistent with a diagnostic testing approach, the IBGR will return pathogenic, likely pathogenic, and uncertain significance variants in genes with a clear causal relationship to the indication. Multiple IBGRs may be requested as the clinical scenario evolves or if new indications arise. Finally, standardized report formats consisting of a one-page results summary followed by detailed explanation of variants and diseases have been generated to communicate results efficiently to medical geneticists as well as non-geneticist providers. We will monitor the understanding of our reports by patients and physicians, and their impact on clinical outcomes, healthcare utilization, and safety. These assessments will contribute to elucidating the potential risks and benefits of genomic sequencing in newborns.

2534S

Detection of Low Level Mixed Chimerism Using High Throughput SNP Genotyping. *A.A. Nakorchevsky[1], E. Flores[1], X. Li[2], T. Hong[2], A.O.H. Nygren[1].* 1) Agena Bioscience, San Diego, CA, USA; 2) Hackensack University Medical Center, Hackensack, NJ, USA.

Patients diagnosed with blood malignancies often receive allogeneic bone marrow transplant or stem cell transplants following the regimen of chemotherapy. Status of the post-transplant patients has to be monitored carefully to allow for early diagnosis of such post-transplant adverse effects as transplant rejection, graft vs. host disease or hematologic relapse. Triaging of the transplant recipients in clinical and research settings is achieved by monitoring the Minimal Residual Disease (MRD) or measuring the amount of mixed chimerism in peripheral blood lymphocytes (PBL) of the patients. Mixed chimerism is a phenomenon where in addition to the donor cells the recipient white blood cells are detected post allogeneic transplant indicating that the possibly malignant recipient cells have evaded cytoreductive treatment and are undergoing hematopoiesis. While MRD molecular techniques target disease-specific markers such as mutations and translocations, mixed chimerism is detected via PCR-based typing techniques such as Short Tandem Repeat (STR) and Variable Number Tandem Repeat (VNTR) analysis. We created a single nucleotide polymorphism genotyping method to detect mixed chimerism in PBL and circulating cell free DNA. The panel targets 92 independent SNPs with a minor allele frequency 0.45-0.55 via multiplex PCR followed by single base extension into the SNP site and detection with MALDI-TOF mass spectrometry. Identification of low percentage mixed chimerism is achieved by analyzing the cumulative skew in genotyping data across a cohort of 92 markers. The feasibility of the panel and algorithm was verified using artificially created mixed chimeric samples from both related and unrelated individuals. Using this model system we achieved a 1% limit of detection with an analytical sensitivity and specificity of 0.95 and 0.9 respectively. Finally we evaluated a set of pre-validated clinical chimeric samples and obtained 100% concordance with an orthogonal technology. The advantages of this SNP-based method compared to already established methodologies is that it has higher sensitivity than STR-based methods and it does not require disease specific markers or prior knowledge of either donor or recipient genotypes. Results are obtained within 8 hours with no library preparation and the experimental workflow can be multiplexed to increase sample throughput. The panel and the analysis algorithm can also be used with other genotyping or sequencing technologies.

2535M

Silver-Russell syndrome and segmental UPD(7q) detected by array CGH. *P. Tavares[1], A. Vaglio[2], R. Lemos[1], C. Ventura[1], A. Pereira[1], A. Sousa[1], J. Sá[1], J. Pinto Basto[1], R. Quadrelli[2], P. Rendeiro[1].* 1) CGC Genetics, Porto, Portugal (www.cgcgenetics.com); 2) Instituto de Genética Médica, Hospital Italiano, Montevideo, Uruguai.

Introduction: 20 to 60% of cases of Silver-Russell syndrome (SRS) are caused by the epigenetic changes of DNA hypomethylation at the telomeric imprinting control region (ICR1) on chromosome 11p15, involving the H19 and IGF2 genes. About 10% of cases are due to maternal uniparental disomy (UPD) of chromosome 7 [Penaherrera et al 2010]. A few cases are due to segmental UPD7 [Hannula et al 2001 and Eggermann 2008]. Objective: We report a SRS clinical case due to a maternal segmental UPD7 and will compare the clinical and molecular data with the published segmental UPD7 SRS clinical cases. Results: The clinical diagnosis was confirmed by the identification of a run of homozygosity at 7q11.22 to 7q31 with 38.7 Mbp (Cytoscan 750k, Affymetrix) with abnormal methylation pattern on genes GRB10 and MEST (MS-MLPA ME032-A1, MRC Holland), confirming a maternal UPD7. Conclusions: This clinical case supports the association between segmental UPD7 and SRS. Despite the fact that segmental mUPD7 is restricted to 7q arm, imprinted genes on 7p arm are also involved, enlightening the complexity of this disease. This clinical case can contribute for a better understanding of the molecular mechanisms of SRS.

2536T

CILIOME RESEQUENCING: A LIFELINE FOR MOLECULAR DIAGNOSIS IN LEBER CONGENITAL AMAUROSIS. *S. HANEIN[1,2], I. PERRAULT[3], M. NICOULEAU[3], S. SAUNIER[4], C. BOLE[5], P. NITSCHKE[6], F. METMER[3], O. XERRI[3], N. DELPHIN[1], A. MUNNICH[1,2], J. KAPLAN[3], J.-M. ROZET[3].* 1) Department of Genetics, Hopital Necker, Paris, France; 2) Translational Genomics Plateform, INSERM UMR1163, U Paris Descartes - Sorbonne Paris Cité. Imagine-Institue of genetic diseases, Paris, France; 3) Laboratory of Genetics in Ophthalmology, INSERM UMR1163, Univ. Paris Descartes - Sorbonne Paris Cité. Imagine-Institue of genetic diseases, Paris, France; 4) Molecular bases of hereditary kidney diseases: nephronopthisis and hypodysplasia. INSERM UMR1163, Univ. Paris Descartes - Sorbonne Paris Cité. Imagine-Institue of genetic diseases, Paris, France; 5) Genomics Plateform, INSERM UMR1163, Univ. Paris Descartes - Sorbonne Paris Cité. Imagine-Institue of genetic diseases, Paris, France; 6) Bioninformatics Plateform, INSERM UMR1163, Univ. Paris Descartes - Sorbonne Paris Cité. Imagine-Institue of genetic diseases, Paris, France.

Leber Congenital Amaurosis (LCA) is the most common cause of incurable childhood blindness. Non syndromic LCA is characterized by a clinical, genetic and physiopathological heterogeneity. Hitherto 18 genes are identified which mutations cause a stationary and dramatically severe cone-rod dystrophy with very poor visual performances or a progressive, yet severe, rod-cone dystrophy with low but measurable visual acuity in the first two decades of life. LCA genes are involved in variable retinal functions; cilia dysfunctions are the leading cause of the disease (7 cilia genes; ca. 30% of the cases). Patients with LCA occasionally develop renal, neurologic and or skeletal symptoms defining a series of syndromes which hallmark is cilia dysfunction. So far, 19 cilia genes are identified which cause syndromic type 1 or 2 LCA. With the exception of CEP290 and IQCB1 which mutations cause non syndromic or syndromic LCA, the identification of disease causing mutations in an infant with LCA is of great help to make a prognosis with respect to the retinal and extraocular outcome. Considering the prevalence of cilia dysfunction in syndromic and non syndromic LCA, we used ciliome sequencing as a molecular diagnosis and progonosis tool and a research tool to identify novel LCA genes. The DNA of 60 unrelated children with LCA was subjected to ciliome sequencing using a custom 5.3 Mb Agilent SureSelect Target Enrichment library which captures 32,146 exons of 1,666 genes selected form cilia databases. Biallelic disease-causing mutations in known genes were identified in 17/60 patients: CEP290 (n = 4), CRB1 (n = 4), RPGRIP1 (n = 2), LCA5 (n = 1), IQCB1 (n = 2), IFT140 (n = 2), AHI1 (n = 1), ALMS1 (n = 1). Single heterozygote disease-causing mutations were detected in 9/60 individuals but Sanger sequencing failed to detect a mutation on the second allele. Finally, 3/60 patients harbored biallelic mutations in three novel genes. In summary, ciliome resequencing allowed identifying the disease-causing gene in 33 % of the patients addressed for LCA making targeted sequencing an interesting alternative to expensive whole exome sequencing. The identification of mutations in genes responsible for Mainzer-Saldino syndrome (IFT140), Joubert syndrome (AHI1) and Alström syndrome (ALMS1) before extraocular expression become obvious demonstrates the importance of NGS-based molecular diagnosis to set-up a rational and efficient follow-up of patients.

2537S

Somatic mosaicism for a novel PDHA1 mutation in a male with severe pyruvate dehydrogenase complex deficiency: a case report and literature review. *S. Zhang[1], K. Deeb[1], J. Bedoyan[2], R. Wang[3], L. Sremba[3], M. Schroeder[1], G. Grahame[2], M. Boyer[3], S. McCandless[2], D. Kerr[2].* 1) Dept. of Pathology, University Hospitals Case Medical Center, Cleveland , OH.,; 2) Center for Inherited Disorders of Energy Metabolism, University Hospitals Case Medical Center; 3) Division of Metabolic Disorders, CHOC Children's Hospital, Orange, CA.

Mutations in the X-linked PDHA1 gene encoding the pyruvate dehydrogenase complex (PDC) PDH-E1α subunit are mostly responsible for PDC deficiency. Males harboring hemizygous PDHA1 mutations are clinically more severely affected, while females often display variable clinical manifestations due to skewed X-inactivation. Mosaic PDHA1 mutations are less frequently reported and may manifest in milder phenotypes. Here, we report a novel mosaic mutation in a male patient with a severe phenotype. The patient is a 4 ½ year old boy with a clinical history of congenital microcephaly, significant brain abnormalities including agenesis of the corpus callosum, colpocephaly, hypoplasia of the cerebellar vermis and Dandy-Walker variant, persistent seizures, profound developmental delay, and failure to thrive. He exhibited persistently elevated blood lactate, pyruvate, and alanine with normal lactate to pyruvate ratio, suggestive of PDC deficiency. PDC activities in cultured skin fibroblasts (SF) from this patient (0.63 and 0.76 nmol/min/mg protein) were below the reference range (mean 2.42; range 1.26-4.42, n = 329), corresponding to 26% and 31% of the mean, respectively. The patient was found to be mosaic for a novel, missense mutation, c.523G>A (p.A175T), confirming the clinical and biochemical findings of PDC deficiency. Molecular analysis of the PDHA1 gene from peripheral blood of the proband's mother did not reveal the same change, consistent with a de novo event in this patient. This is the seventh reported case of male somatic mosaicism for a PDHA1 mutation, but notably with a significantly severe clinical phenotype.

2538M

Performance evaluation of inherited disease panels for Korean patients with Limb-girdle muscular dystrophies (LGMD). *H. Kim[1], N.J. Kwon[1], J. Kim[1], M.W. Seong[2], H. Park[2], S.H. Seo[2], D.D. Seol[2], S.I. Cho[2], K.S. Yang[1], S.S. Park[2], J.S. Seo[1,3].* 1) Macrogen Inc., Seoul, Korea; 2) Department of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea; 3) Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul, Korea.

The genomic data produced by next-generation sequencing (NGS) are widely used to better understand disease development and drug response. Recently, numerous companies have developed and commercialized targeted sequencing panels that focus on specific disease of interest. In this study, the performance of three commercialized inherited disease panels were evaluated on 16 Korean patients, suspected to have Limb-girdle muscular dystrophies (LGMD). The TruSight One sequencing panel and TruSight Inherited Disease (Illumina), which utilizes prbobes to capture the target regions, and Ion AmpliSeq Inherited Disease Panel (Thermo-Lifetechnologies), which utilizes the amplicon system, were evaluated. The TruSight One sequencing panel provides comprehensive coverage of more than 4800 clinical relevant genes and focuses on the exonic regions, which harbor the disease-causing mutations. On the other hand, TruSight Inherited Disease and Ion AmpliSeq Inherited panel are more inherited disease specific. TruSight Inherited Disease sequencing panel focuses on the coding region and intron-exonic boundaries of 552 genes of pediatric-onset Mendelian disorders, while Ion AmpliSeq Inherited Disease sequencing panel focuses on the coding exons of over 300 genes associated with over 700 unique inherited diseases, including neuromuscular, cardiovascular, developmental, and metabolic diseases. The sequencing data were produced using MiSeq, HiSeq2500 (Illumina), and Ion Proton (Thermo-Lifetechnologies). The basic analysis for sequencing quality and disease associated variants of each panels were performed through our internal analysis pipeline. We analyzed common targeted regions of the three panels, and all identified mutations were confirmed using Sanger Sequencing. Additionally, the data produced by HiSeq2500 and MiSeq were compared to identify mutation that were found at high depth of HiSeq2500 data but not present in MiSeq Data. We evaluated the accuracy of each panels to validate NGS technologies for clinical diagnosis. We show that inherited disease panels are able to detect rare mutation in the disease samples, and our data provides good foundation for further clinical studies.

2539T

Expanding the phenotype of genes encoding the components of the PI3K pathway. *M.A. Walkiewicz[1], A. Dang Do[2], N.L. de Macena Sobreira[2], J. Beuten[1], J. Zhang[1], Z. Niu[1], F. Xia[1], J.N. Bodurtha[2], B.H. Graham[1,4], F. Scaglia[1,4], D.M. Muzny[3], A.L. Beaudet[1], R.A. Gibbs[1,3], Y. Yang[1], C.M. Eng[1].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Texas Children's Hospital, Houston, TX 77030.

Mutations in genes encoding the components in the PI3K pathway, namely *AKT3*, *PIK3R2*, and *PI3KCA*, have been recently reported in individuals with megalencephaly-capillary malformation (MCAP - MIM 602501), megalencephaly-polymicrogyria-polydactyly-hydrocephalus (MPPH - MIM 603387), CLOVES (MIM 612918), and Cowden (MIM 615108) syndromes. The first proband is a 22 month old female who presented with poor feeding since birth, stiffness of extremities, megalencephaly, communicating hydrocephalus, and electroclinical seizure by EEG. The initial course was complicated by sepsis leading to hyperlactatemia, hypoglycemia, metabolic acidosis, and thrombocytopenia. Brain MRI revealed multiple foci of T1 hyperintensity in the cerebral white matter thought to represent heterotopias. By WES we identified a *de novo*, heterozygous c.1393C>T (p.R465W) mutation in the *AKT3* gene. The same mutation was previously reported in a patient with congenital megalencephaly, somatic asymmetry, connective tissue dysplasia, and umbilical hemangioma. The second proband is a 2.6 year old male born prematurely at 36 weeks. He has global developmental delay, macrocephaly, dolichocephaly, and hepatomegaly. MRI of the brain demonstrated perisylvian polymicrogyria. By WES we identified a *de novo*, heterozygous c.1669G>T (p.D557Y) variant in the *PIK3R2* gene. The third proband is a 24 month old female who presented with global developmental delay, macrocephaly, asymmetric facies, nevus flamus, atrial tachycardia, omphalocele, left hydronephrosis, multiple other dysmorphic features including 2-3 toe syndactyly, and recent accelerated weight gain. Brain MRI revealed small posterior fossa with crowding of the cerebellum, ventriculomegaly, and mild parietal periventricular T2 hyperintensity. By WES we identified a *de novo*, mosaic (mutant 27: wild type 185) c.1093G>A (p.E365K) mutation in the *PIK3CA* gene. This change was previously reported in a patient with megalencephaly. In summary, we report three additional individuals with mutations in the components in the PIK3-pathway. Although all three patients had megalencephaly, abnormal brain imaging, and global developmental delay, other clinical findings confounded a definitive clinical diagnosis. This further demonstrates the utility of WES in confirming clinical diagnosis and expanding disease phenotype.

2540S

CLINICAL ADVANTAGES OF HIGH COVERAGE COMPREHENSIVE NGS PANELS IN THE MOLECULAR DIAGNOSIS OF HEREDITARY CANCER MUTATIONS. *H. DAI, G. DOUGLAS, E. GORMAN, S. CHEN, X. TIAN, Y. FENG, F. LI, J. WANG, V. ZHANG, L. WONG.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX., United States.

Introduction: Accurate assessment of hereditary cancer risk by molecular diagnosis of predisposing mutations is essential for disease management and genetic counseling. When ordering such genetic tests, it may be challenging to single out relevant genes while still aiming to obtain evaluations as comprehensive as possible. High coverage capture-based Next Generation Sequencing (NGS) panel analysis provides a promising solution to meet clinical demands in a cost-efficient way. Methods: 61 hereditary cancer-associated genes, grouped into 12 disease-specific panels and one comprehensive panel, were targeted with in solution capture followed by next generation sequencing. Exons with coverage <20X, or with possible pseudogene(s), are complemented by amplicon/Sanger sequencing to reach 100% coverage. Mutations/unclassified variants detected by NGS were further confirmed by Sanger sequencing. Potential copy number variation (CNV) suggested by our NGS assay are confirmed by either array CGH or MLPA methods. 10 phase I samples (commercially available plus internal control gDNA from different types of tissues) and 13 phase II samples (with known mutations) were used for validation. Results: The mean coverage is 1255X/750X for phase I and II validation, respectively. Phase I samples have an average of 97 variants in the comprehensive panel, which are 100% concordant with amplicon/Sanger sequencing results. In phase II samples, all known point/indel mutations in BRCA1, MEN1, MUTYH, VHL, APC and PTEN were successfully detected. Furthermore, pathogenic exonic CNVs in VHL, MLH1, TP53, PTEN, EPCAM and a whole gene deletion in APC gene were also identified successfully and confirmed. As an example of highlighting the advantage of our NGS assay, we were able to detect both a point mutation (c.65C>A) and an exon 14 deletion of PMS2 gene in one sample simultaneously, despite the well-known PMS2 pseudogene interference. Conclusions: This capture based deep coverage NGS panel provide accurate and comprehensive analysis for hereditary cancers genes. The combination strategies of both deep coverage of all coding regions and Sanger confirmation/complementation allows accurate identification of a wide spectrum of mutation types, including point mutations, small indels, mosaicism. In addition, CNVs at the exon level can be analyzed simultaneously. Our data underscore the important clinical utility of NGS-based analysis in the molecular diagnosis of hereditary cancer.

2541M

Mutations in *STK11* identified exclusively in individuals with clinical histories suggestive of Peutz-Jeghers syndrome. *J.S. Dolinsky[1], S.B. Keiles[1], T.F. Pesaran[1], C. Horton[1], M. Umali[1], K. McGoldrick[1], S. Li[1], E. Chao[1,2].* 1) Ambry Genetics, Aliso Viejo, CA., USA; 2) University of California, Irvine, School of Medicine, Irvine, CA.

Pathogenic mutations in *STK11* are detected in an estimated 80% of individuals clinically diagnosed with Peutz Jeghers syndrome (PJS). PJS is characterized by gastrointestinal PJS-type hamartomatous polyposis, mucocutaneous pigmentation, and cancer predisposition including breast cancer, which can occur at early ages in PJS with up to a 57% breast cancer risk by age 70. Due to the risk of early onset breast cancer, it has been proposed that family members could develop breast cancer prior to polyposis symptomatology. Thus, *STK11* has been included on a number of commercially available and academic research hereditary breast cancer (HBC) multigene panel tests. In this study, we sought to identify individuals with a *STK11* mutation detected on HBC multigene panels offered by Ambry Genetics from March 2012 through May 2014. We hypothesized that some individuals with HBC in the absence of other PJS features will harbor a *STK11* pathogenic mutation. 12,928 individuals underwent panel testing with BRCAplus (*BRCA1*, *BRCA2*, *CDH1*, *PTEN*, *STK11*, *TP53*) or BreastNext (BRCAplus genes and *ATM*, *BARD1*, *BRIP1*, *CHEK2*, *MRE11A*, *MUTYH*, *NBN*, *NF1*, *PALB2*, *RAD50*, *RAD51C*, *RAD51D*). Clinical data submitted via test requisition form and subsequently verified by the clinician was assessed for individuals with a *STK11* mutation. Clinical data from an additional 73 individuals with mutations in *STK11* identified through single gene testing or another multigene panel was also assessed. Several probands had histories consistent with HBC, however those probands met clinical criteria for PJS. Remaining individuals with a *STK11* pathogenic mutation had a clinical diagnosis of PJS or features consistent with PJS, including one individual identified through a HBC panel. The remaining 12,927 individuals referred for HBC multigene panel testing did not harbor a *STK11* pathogenic mutation. While variants of unknown significance (VUS) are not common in *STK11*, having been identified in approximately 0.44% of the HBC multigene cohort, there are significantly more VUS than mutations identified in *STK11* on HBC panels ($p<0.001$). This analysis does not support our hypothesis, but rather suggests that inclusion of *STK11* may not be appropriate on multigene HBC panels, introducing more uncertain results without improving diagnostic yield. *STK11* testing should be reserved for broad spectrum hereditary cancer panels, polyposis panels or single gene analysis in individuals with a history suggestive of PJS.

2542T

Risks associated with utilizing molecular data to guide tumor surveillance in BWS. *C. Shuman[1,2,3], S. Choufani[2], L. Steele[3,4], N. Parkinson[3,4], J. Lauzon[5], J. Tagoe[6], P. Ray[3,4], R. Weksberg[1,2,3,7,8].* 1) Clinical & Metabolic Genetics, Hosp for Sick Children, Toronto, ON., Canada; 2) Genetics and Genome Biology Program, Research Institute, The Hospital for Sick Children; 3) Department of Molecular Genetics, University of Toronto; 4) Paediatric Lab Medicine, Molecular Genetics, The Hospital for Sick Children; 5) Department of Medical Genetics, Alberta Childrens Hospital; 6) Genetic Services, Alberta Health Services; 7) Institute of Medical Science, University of Toronto; 8) Department of Paediatrics, University of Toronto.

In general, it is well accepted that tumor risk assessment can be significantly improved by incorporating epigenomic/genomic data. In this regard, recent publications (Brioude et al. 2013; Scott et al. 2006) recommend tailoring tumor surveillance for children with Beckwith-Wiedemann syndrome (BWS) based on the specific molecular etiology, when identified. However, technical limitations of the test in determining somatic epigenetic mosaicism can lead to negative health outcomes. BWS is an etiologically heterogeneous, pediatric overgrowth syndrome associated with an increased risk (~ 4 - 21%) for embryonal tumor development. This risk led to tumor surveillance recommendations including 3 monthly abdominal ultrasounds. (Epi)genetic alterations involving the telomeric imprinting centre, IC1 (i.e. gain of methylation [GoM]at IC1or paternal uniparental disomy [UPD] of 11p15.5) have the highest risks for tumor development. Those with alterations involving the centromeric imprinting centre, IC2 (i.e. loss of methylation [LoM] at IC2 or mutations in CDKN1C) have significantly lower risks and no cases of Wilms tumor (WT) have been reported. These findings led to new molecular group specific recommendations to discontinue ultrasound screening for WT in children with demonstrated IC2 LoM. We present 3 cases of children with BWS and IC2 alterations on initial MS-MLPA testing who subsequently were found to have renal tumors/nephrogenic rests detected by abdominal ultrasound. Pt 1: Nephrogenic rests (stable) were detected in a 16 month male with IC2 LoM in DNA from saliva and blood, on repeat testing. Pt 2: WT was found in a 15 month old female with LoM in DNA from blood. Retesting on a new blood sample revealed borderline findings of IC2 LoM and IC1 GoM consistent with UPD. Pt 3: Unilateral WT was detected in a 9 month old female. Repeat testing on DNA obtained from blood and skin showed borderline IC2 LoM. No evidence of UPD was found. Pyrosequencing these cases in the research lab demonstrated methylation indices most consistent with low level mosaicism for UPD. Therefore, we would advocate caution in revising medical management guidelines especially when the etiology involves mosaicism and/or the testing technology generates data that are not definitive. Technological advances may resolve some of these issues to support stratification of tumor risk assessment and surveillance by molecular test results.

2543S

Universal Lynch syndrome screening in an integrated health care system: Assessment of patient perspectives on benefits and barriers. *J.E. Hunter[1], J. Zepp[1], M. Gilmore[1], J. Davis[1], E. Esterberg[1], K.R. Muessig[1], S.K. Peterson[2], S. Syngal[3], L. Acheson[4], G. Wiesner[5], K.A.B. Goddard[1], J. Reiss[1].* 1) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 2) Department of Behavioral Science, University of Texas MD Anderson Cancer Center, Houston, TX; 3) Dana-Farber Cancer Institute, Boston, MA; 4) Departments of Family Medicine and Community Health, Reproductive Biology, and Oncology, Case Western Reserve University, Cleveland, OH; 5) Vanderbilt Hereditary Cancer Program, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN.

Lynch syndrome (LS) is the most common form of hereditary colorectal cancer (CRC) and accounts for 2-3% of all CRC cases. A diagnosis of LS has significant impact on clinical management of patients and their at-risk family members. Universal screening for LS using either microsatellite instability (MSI) or immunohistochemistry (IHC) testing among CRC patients is increasingly implemented in clinical practice. The aim of this study was to assess knowledge of and attitudes toward MSI screening among newly diagnosed CRC patients. Participants were recruited through Kaiser Permanente Northwest and were administered a phone survey regarding MSI screening. The study population consisted of 74 men and 50 women age 39 to 87 years. Only 3% of participants had any prior knowledge of MSI screening, and 8% anticipated having an abnormal screening result. However, most participants expressed interest in MSI screening: 92% wanted to know their risk of hereditary CRC and 93% agreed that MSI screening should be available to anyone. Overall, participants endorsed the benefits of screening for themselves and their families: 87% agreed they would be glad that their doctor ordered this test, 94% would like information that could benefit their family, and 99% responded they would share their screening results with their family. Older participants were less likely to be worried about their family's risk of CRC and lower income participants were more likely to report that MSI screening would not help them or their family. The most commonly reported barrier to MSI screening was worry regarding the cost of additional testing and surveillance, endorsed by 72% of participants. In particular, younger and lower income participants were more likely to be worried about additional costs associated with an abnormal screening result. Another potential barrier to screening was a concern about privacy and confidentiality of the test result, endorsed by 30% of participants. Few participants (10%) indicated concern about losing their health insurance; however, younger and lower income participants were more likely to be concerned. Although CRC patients in our study had little a priori knowledge about MSI screening, they responded positively to undergoing screening and recognize the potential benefits to themselves and their families. However, concern regarding additional out-of-pocket healthcare costs may be a barrier to MSI screening, particularly among younger and lower income patients.

2544M

Diagnostic Exome Sequencing as the Foundation of Building Pharmacogenomics Based Therapeutic Models for the Treatment of Ion Channel Epilepsy. *S.K. Gandomi, K. Waller, K.D. Farwell Gonzalez, L. Shahmirzadi, R. Baxter, B. Tippin-Davis, S. Tang.* Clinical Genomics, Ambry Genetics, Aliso Viejo, CA., USA.

Diagnostic exome sequencing is currently the most comprehensive clinical genetic testing option available to patients with neurological disorders. Historically, individuals presenting with features suggestive of an underlying genetic syndrome had limited options in the diagnostic process, especially in regards to their subsequent management. With the recent advancement of Next Generation sequencing techniques, and the rapid growth in clinical utility of diagnostic exome sequencing (DES) in the clinical neurology setting, the diagnostic and treatment options available for epilepsy patients is progressively changing. Here we present our institution's DES data to show the effectiveness of DES in diagnosing ion channel related epilepsy, which therefore lays the foundation for potential pharmacogenomic treatment protocols in the near future. In the first 496 cases sent to our institution for DES, 131 (27.93%) patients were diagnosed with epilepsy as a major clinical feature upon referral. Approximately 43 (32.8%) of these patients were found to have positive, pathogenic alterations, and an additional 11 (8.4%) patients were found to have "likely positive" alterations in a variety of genes. In both of these cohorts combined (n=54), 1 patient had an alteration in a calcium-activated chloride channel gene (ANO3[610110]), 3 patients had alterations in sodium channel genes (SCN3A [MIM 182391], SCN1A[MIM 182389]), and 3 additional patients had alterations in potassium channel genes (KCNQ2 [602235], KCNC3 [176264]). As the current literature suggests, there are rapidly increasing therapeutic possibilities for epilepsy, especially for disorders involving major ion channels such as SCN1A (MIM 182389). We anticipate the continued and increased use of DES in identification of pathogenic alterations for patients with ion channel-related epilepsy, and recommend the integration of DES technology into research and clinical settings for the long-term development of targeted therapies.

2545T

Next-generation sequencing diagnostics for inherited arrhythmogenic cardiac disorders. *C. Marschall, I. Vogl, HG. Klein.* Center for Human Genetics, 82152 Martinsried, Bavaria, Germany.

Inherited arrhythmogenic cardiac disorders are composed of primary arrhythmia syndromes such as long QT syndrome (LQTS), catecholaminergic polymorphic ventricular tachycardia (CPVT) and Brugada syndrome (BrS) and the cardiomyopathies with structural changes leading to an arrhythmia risk such as hypertrophic cardiomyopathy (HCM), dilatative cardiomyopathy (DCM) and arrhythmogenic right ventricular cardiomyopathy (ARVC). All these diseases are associated with significant morbidity and mortality and are a known risk factor for sudden cardiac death. The presymptomatic diagnosis is essential as for most of the cases a therapy is available. Mutation detection in the index cases followed by cascade screening of the family members could be life saving. As some cardiomyopathies may present with only little structural symptoms, the differential diagnosis of these arrhythmogenic disorders may be difficult. Unfortunately some arrhythmogenic disorders are extremely heterogeneous. A Nextera sequence capture in solution target enrichment assay was designed for the capture of coding regions including splice sites of 61 genes known to be associated with arrhythmogenic disorders. Target enrichment followed by re-sequencing on the Illumina MiSeq was used for mutation detection. More than 40 clinical samples with a suspected inherited arrhythmogenic disorder were analyzed with a minimum coverage of 30-fold. Data analysis was performed with the CLC Genomic Workbench software. Potentially disease causing variants and regions with an insufficient coverage were re-analyzed with Sanger sequencing. Disease causing mutations were detected in over 40% of the cases and less than 20% showed ambiguous variants.

2546S

Genetic testing leads clinical care in neonatal diabetes: a new paradigm. *E. De Franco[1], S.E. Flanagan[1], J.A.L. Houghton[1], H. Lango Allen[1], D.J.G. Mackay[2,3], I.K. Temple[3], S. Ellard[1], A.T. Hattersley[1].* 1) University of Exeter Medical School, Exeter, United Kingdom; 2) Wessex Regional Genetics Laboratory, Salisbury Health Care Trust, Salisbury, UK; 3) Human Genetics and Genomic Medicine, Faculty of Medicine, University of Southampton, Southampton, UK.

Traditional genetic testing focusses on analysis of one or a few genes according to clinical features; this testing paradigm will now change as next-generation sequencing allows simultaneous analysis of multiple genes. Neonatal diabetes is the presenting feature of many discrete clinical phenotypes defined by different genetic aetiologies. The genetic subtypes differ in treatment requirements and associated clinical features. We investigated the impact of early, comprehensive testing of all known genetic causes of neonatal diabetes. We studied 1020 patients with neonatal diabetes diagnosed before 6 months of age referred from 79 countries between 2000 and 2013. Mutations were identified by comprehensive genetic testing including targeted next-generation sequencing of 21 neonatal diabetes genes. A genetic diagnosis was obtained in 840 patients (82%). The most common cause were potassium channel mutations (n=390), which were less common in consanguineous families (46% vs 12%, p<0.0001). Transfer from insulin to sulfonylurea therapy with improved glycaemic control was observed in most of these patients. Median time from diabetes diagnosis to referral for genetic testing decreased from over 4 years before 2005 to less than 3 months (2011-2013). Early referral altered the clinical phenotype at time of genetic testing. In patients with a genetic diagnosis of transient neonatal diabetes remittance was seen only in 10% with early referral (less than 3 months from diagnosis) compared to 100% with late referral (more than 4 years from diagnosis, p<0.001). Similarly in patients with genetically diagnosed Wolcott-Rallison syndrome 11% with early referral had syndromic diabetes compared to 82% (p<0.001) with late referral when skeletal/liver involvement was common. Patients are now referred for genetic testing close to presentation with neonatal diabetes. Comprehensive testing of all aetiologies identifies causal mutations in over 80% of cases. The genetic result predicts optimal diabetes treatment and development of related features. This represents a new paradigm for clinical care with genetic diagnosis preceding development of clinical features and guiding clinical management.

2547M

Identification and characterization of *CFTR* deletion and duplication mutations among nonwhite patients with cystic fibrosis. *I. Schrijver[1,2], L. Pique[1], S. Graham[3], M. Kharrazi[4].* 1) Department of Pathology, Stanford Medical Center, CA; 2) Department of Pediatrics, Stanford Medical Center, CA; 3) Sequoia Foundation, La Jolla, CA; 4) California Department of Public Health, Richmond, CA.

Even with the implementation of newborn screening programs across the United States, there remains an inequitable identification of cystic fibrosis (CF [OMIM #219700]) in nonwhite groups. Diagnosis can be improved by determining the type and estimated frequency of individual mutations in nonwhite populations. We specifically investigated the prevalence of deletion and duplication mutations, which are not detectable by Sanger sequencing, in CF patients in the U.S. with African, Native American, Asian, East Indian or Middle Eastern backgrounds. Of the approximately 500 eligible nonwhite CF patients in the Cystic Fibrosis Foundation (CFF) Patient Registry database with zero or one *CFTR* [OMIM *602421] mutation(s) instead of the expected two, 142 probands were enrolled into our study. Of the enrolled probands, 89 were fully genotyped by direct sequencing of *CFTR*. However, 26 had only a single mutation and 27 had no mutations detected. These patients were subsequently analyzed by Multiplex Ligation-dependent Probe Amplification (MLPA) and novel deletion/duplication breakpoints were further characterized by a walking PCR technique. MLPA testing of the 52/53 probands for whom DNA remained available identified 15 del/dup mutations in 13 probands: two were identified in two probands each, one of whom already had a pathogenic mutation identified by sequencing (c.1521_1523delCTT, p.Phe508del). Thus, only 14/15 of these mutations accounted for unidentified alleles (14/78 = 18%). Six of the rearrangements were novel and included one simple deletion, three complex deletions containing short (2-12 bp) insertions, one duplication and a multigenic deletion of exons 19-24 of *CFTR* that extends into the neighboring *CTTNBP2* gene on chromosome 7. Two previously reported rearrangements - a deletion of exons 2-3 and a complex deletion of exons 17a-17b that includes a 62 bp insert - were detected in four probands each (~5% of unidentified alleles). Deletions and duplications in the *CFTR* gene appear to be relatively common in nonwhite CF patients and account for ~18% of unidentified alleles after Sanger sequencing. MLPA should be routinely applied in diagnostic *CFTR* testing, and is an especially valuable part of the testing algorithm in nonwhite individuals. Inclusion of such mutations facilitates early diagnosis and can help improve the quality of life of nonwhite and mixed ethnicity individuals affected with this condition.

2548T

Predicted management implications of CytoScan Dx® Assay in a consecutive cohort of patients suspected of Developmental Delay, Intellectual Disability, and/or Congenital Anomalies. *A. Chaubey[1], R. Pfundt[2], K. Kwiatkowski[3], A. Roter[3], B. DuPont[1], E. Thorland[4], R. Hockett[5], A. Shukla[3], E.T. Fung[3].* 1) Cytogenetics, Greenwood Gen Ctr, Greenwood, SC; 2) Human Genetics, Radboud University, Nijmegen, The Netherlands; 3) Clinical Applications, Affymetrix, Santa Clara, CA; 4) Cytogenetics, Mayo Clinic, Rochester, MN; 5) Cytogenetics, CombiMatrix, Irvine, CA.

Background: The prevalence of developmental disability is 13.87% and occurs across all racial, ethnic, and socioeconomic groups. Establishing an underlying diagnosis early has the potential to reduce health care costs, provide physicians and families with knowledge of which disorder is affecting the child, prognosis, and comorbidity information, all of which have implications beyond only medical treatment. Although microarrays are now common first tier tests in this patient population and supported by the medical guidelines, payer reimbursement for testing is inconsistent. Methods: In a consecutive and unselected cohort of 960 patients who had previously undergone regular genetic testing for suspicion of a genetic disorder related to developmental delay, intellectual disability, congenital anomalies, and/or dysmorphic features, we assessed the diagnostic yield of the CytoScan® Dx Assay as compared to historical patient reports. To then understand how this relates to management implications we assessed the clinical utility of CytoScan Dx Assay by applying the Rigg's criteria of actionability to the chromosomal abnormalities identified by CytoScan Dx Assay. Results: 61.7% of all samples were male with an average age of 7.8 ± 11.3. Overall, 86% of the subjects were assessed using a microarray as part of routine patient care. Abnormal or pathogenic findings were reported similarly for RPC (13.3% or 128/960) and CytoScan Dx Assay (13.8% or 132/960). 33% or 43/132 of the abnormal patients (comprising 51 CNVs) identified by CytoScan Dx Assay had predicted clinical management implications. Conclusions: This is the first study to assess the clinical performance of CytoScan Dx Assay within a consecutive cohort of patients with developmental delay, intellectual disability, and/or congenital anomalies that were referred for whole genome chromosomal testing. CytoScan Dx Assay diagnostic yields are similar to previous studies and one-third of the patients with pathogenic chromosomal identified by CytoScan Dx Assay are predicted to have clinical management implications which may improve health outcomes.

2549S

Comprehensive evaluation of the FBN1, LTBP2 and ADAMTSL4 genes in 667 patients with ectopia lentis. *B. Callewaert[1], L. De Potter[1], S. Van Nuffel[1], E. Debals[1], J. De Backer[1], O. Vanakker[1], H. Van Esch[2], M. Biervliet[3], J. van den Ende[4], J. Désir[5], A. De Paepe[1], P. Coucke[1].* 1) Center for Medical Genetics, Ghent University Hospital, Ghent, Ghent, Belgium; 2) Center for Medical Genetics, University Hospital of Leuven; 3) Center for Medical Genetics, University Hospital of Brussels; 4) Center for Medical Genetics, University Hospital of Antwerp; 5) Department of Medical Genetics, Hôpital Erasme, Université Libre de Bruxelles.

Introduction: Ectopia lentis (EL), a displacement of the lens of the eye, has an estimated prevalence of 6.4 per 100 000 live born children. EL can present as an isolated phenomenon or within a syndromal constellation with multisystemic involvement. The most prevalent congenital cause of EL is Marfan Syndrome (MFS), which requires a lifelong cardiovascular follow-up. Because of the evolving phenotype, the diagnosis of MFS is difficult to make at young age and *FBN1* testing is often requested. When no detectable *FBN1* mutation is found, the absence of a definite diagnosis may cause psychological distress and sometimes unnecessary examinations. Recently, mutations in *LTBP2* have been implicated in congenital glaucoma and microspherophakia with EL. *ADAMTSL4* mutations have been implicated in EL and EL et pupillae. This study evaluates the contribution of mutations in *FBN1*, *LTBP2* and *ADAMTSL4* mutations in isolated EL in a large study group. Methodology: Over a 22 month time period a total of 667 probands with EL were referred for *FBN1* analysis. We analyzed the *FBN1*, *ADAMTSL4* and *LTBP2* genes in these patients using a step-wise next-generation approach. Mutations were confirmed with Sanger sequencing. Results: 616 out of 667 probands harbored an *FBN1* mutation (of which 44 did not fulfill the 2010 Ghent revised criteria for MFS) and 51 did not (of which 19 patients did fulfill clinical criteria for MFS) resulting in a detection rate of 92.4% (616/667). In the remaining 51 patients, homozygous or compound heterozygous mutations in *ADAMTSL4* were found in four patients. (c.767_786del; c.2237G>A, c.2021_2022delCT, c.2977C>T, c.963dup). One proband showed a homozygous mutation in *LTBP2* (c.4964A>G). In another proband only one mutation (c.3850C>T, p.Arg1284Cys) was found. Conclusions: *FBN1* is by far the primary gene to screen upon the diagnosis of EL. However, our results show that the clinical spectrum resulting from *LTBP2* and *ADAMTSL4* mutations includes isolated EL. Therefore, if no *FBN1* mutation is found, screening of *ADAMTSL4* and *LTBP2* is advised as this may reveal a final diagnosis for 12% of the remaining patients and even up to 20% if the MFS criteria are not met. In addition, our data indicate that *LTBP2* mutations are no likely cause of *FBN1* negative MFS with EL.

2550M

Barriers to Cytogenetic Testing in Adults with Congenital Heart Disease. *J. Vengoechea Barrios[1], L. Carpenter[1], R.T. Collins[2].* 1) Division of Genetics, University of Arkansas for Medical Sciences, Little Rock, AR; 2) Department of Medicine and Department of Pediatrics, University of Arkansas for Medical Science.

The University of Arkansas for Medical Sciences adult congenital heart disease (ACHD) interdisciplinary clinic began in August of 2012. The clinic includes an ACHD cardiologist, an internist/geneticist, and a genetic counselor. A recent retrospective study of cytogenetic testing in children with syndromic and non-syndromic CHD reported a 15% rate of clinically significant results (1). We sought to determine the rate of clinically significant cytogenetic findings in our ACHD population. A retrospective review was performed of all records of ACHD patients seen in our clinic by both the cardiologist and geneticist. 51 patients were identified for the study as of April 1, 2014. The majority of patients (64.7%; 33/51) were female, as many patients came to care due to pregnancy. The most common diagnoses were bicuspid aortic valve (with either coarctation or aneurysm of the aorta) and Tetralogy of Fallot, each seen in 10/51 patients (19.6%). Other diagnoses included pulmonary stenosis (4/51), isolated coarctation of the aorta (4/51), transposition of the great arteries (4/51) and AV canal defect (4/51). One patient had a previously established diagnosis of Turner syndrome. A cytogenetic test was offered to 37 (74%) of the remaining 50 patients (SNP microarray in 35 patients, a high-resolution karyotype in one patient and a 22q11 FISH in 1 patient), 8/37 patients did not consent to genetic testing (21.6%). Of the remaining 29 patients, testing was completed in 12 cases (41.3%) of which 8 were women (66.7%). The results of the 12 patients included: 1 22q11 deletion, 1 incidental finding of Klinefelter, 1 6q27 gain of uncertain significance, 1 benign variant, and 6 normal microarrays. Two results are pending at this time. Of those patients who gave consent for genetic testing, 17/29 (59%) did not receive testing: 4 patients had no insurance; 3 had insurance that denied coverage; 6 declined due to out-of-pocket cost; and 4 patients did not follow-up after consent was obtained. The present study demonstrates the difficulties in performing cytogenetic testing in ACHD patients. Barriers to access for genetic testing include lack of insurance, restrictive insurance policies with genetic testing exclusions, out-of-pocket costs and difficulties in follow-up after insurance has approved testing. Reference 1.Connor, J. A., et al. (2014), Genetic Testing Practices in Infants with Congenital Heart Disease. Congenital Heart Disease, 9: 158-167.

2551T

Optimized workflow for single cell copy number profiling using high resolution oligo CGH arrays. *S. Basehore, P. Costa, N. Novoradovskaya, A. De Witte, S. Fulmer-Smentek.* Agilent Technologies Genomics, Santa Clara, California.

Microarray-based comparative genomic hybridization (aCGH) is a powerful technique used in cytogenetic and cancer research for genome-wide detection of DNA copy number (CN) changes. As a consequence of cell and tissue heterogeneity, there has been an increasing focus in single cell analysis in tumors and micrometastases characterization, and in pre-implantation and noninvasive prenatal (fetal cells circulating in maternal blood) investigation. Traditional FISH and PCR based techniques, and more recently BAC arrays, have been used to provide insights into a single cell's genome, with limited resolution. Here we describe a same day, cost-effective, analysis workflow that combines whole genome amplification (WGA), with CN profiling using high-resolution oligo CGH microarrays for single cell research. To assess the accuracy of aberration detection, a single cell model system was built from a set of normal and aberrant cell lines with varying sizes of known CN changes. As references, lymphocytes isolated from normal male and female individuals were used. Due to the minute amounts of genetic material contained within each cell, all samples and references were subjected to a multiple displacement amplification-based WGA to increase the amount of DNA while maintaining its genomic representation. Amplified samples were then differentially labeled and combined in 7 pairs of test vs. test samples and 1 pair of reference vs. reference per 8-pack microarray slide. This eliminates the use of a reference for every array and allows the processing of 14 samples on one slide. Combined samples were hybridized for 2-6 hours to an 8x60K CGH microarray containing probes optimized for single cell WGA and analysis. Following hybridization, microarray slides were washed and scanned. The data were extracted and analyzed for CN alterations using algorithms and a single cell specific analysis method implemented in Agilent CytoGenomics 2.9. Using an optimized workflow for sample processing and analysis in 7-11 hours, CN changes were accurately assessed in the single cell model system. Aberrations affecting whole and portions of chromosomes, as small as 6 Mb, were confidently identified and confirmed by gDNA CN profiling. This short workflow enables researchers to obtain reliable results at a high resolution while remaining cost-effective.

2552S

Performance comparison between low coverage semiconductor sequencing and array comparative genomic hybridization to analyze copy number variation. *B. Min, M. Seo, J. Kim.* Seoul National University, Seoul, South Korea.

Copy number variations are alterations of the DNA and have an important role in the pathogenesis of human disease including cancer and genetic disorders. Array comparative genomic hybridization is the most widely used analytical method for genetic identification of copy number. Next generation sequencing technologies provide an opportunity to detect copy number by comparing the number of sequence reads between patients and control samples. To evaluate the use of semiconductor sequencing for copy number variation diagnosis, Detection of copy number using the Agilent 4 X 180k aCGH platform was compared with low coverage semiconductor sequencing on the patients with genetic disorders. Detected changes in copy number using low coverage sequencing were validated by using aCGH results data and Sanger sequencing. Most copy number variants detected among patients using the aCGH method were successfully detected by low coverage sequencing. In the diagnostic setting, A balance of low coverage sequencing cost and diagnostic sensitivity and specificity should be considered. Turnaround time of experiment also should be considered to challenge array comparative genomic hybridization method. semiconductor sequencing method is alternative sequencing platform with a reduced turnaround time and cost of sequencing. We propose that low coverage semiconductor sequencing can be used as a diagnostic procedure for the patients with CNVs.

2553M

Effects of clinician guided genomic risk assessments: HelloGene. *H. Jin[1], T. Kim[1], K. Ahn[1], J. Bhak[2], H. Joo[1], S. Dong[3].* 1) TheragenEtex, Suwon, Gyeonggi, South Korea; 2) Personal Genomics Institute, Genome Research Foundation, Suwon, South of Korea; 3) Kyung Hee University School of Medicine, Seoul, South of Korea.

Genome wide association studies (GWASs) on common diseases have been reported increasingly since 2005. These results were applied for the personalized genomic risk assessments tests by several corporations, such as 23andMe, Navigenics, DeCODEme and so on. These companies offer direct-to-consumer (DTC) tests without the obligatory involvement of the health care provider. Since DTC is illegal in Korea, the test results of personalized genomic risk assessments test; HelloGene by Therabio Inc., should be delivered and explained by clinician. HelloGene service was launched in January, 2013 in Korea. However, genomic risk assessments test were not popular in the healthcare facilities, and social awareness about risk assessment test was limited among clinicians. To overcome these circumstances, Therabio Inc. has initiated the clinical trial with 18 university hospitals via academic society, the Korean Society of Health Screening and Promotion, in January, 2014. The trial consists of two survey questionnaires which are given before and after the HelloGene service. The first questionnaire consists of the ten questions regarding basic understanding and personal impression about the test such as previous experience, knowledge and so on. After the test, the subjects were surveyed with the second questionnaire, which have 20 questions mainly about the satisfaction and change in perspective about the genomic risk assessments test. Up to now, 55 subjects were completed both questionnaires. Interestingly, 54 out of 55 subjects were responded that it was helpful to have genetic counseling by clinician. In addition, 54 out of 55 subjects showed willingness for changing their lifestyle according to the results to prevent diseases. Even though the trial is on-going, it is already producing important information regarding general populations' impression and perspective towards genomic risk assessment tests and it is predicted to showing positive aspects for genomic risk assessment tests in Korea.

2554T

Rare finding of non-transient congenital extra and intra-medullary acute myeloid leukemia with Beckwith-Wiedemann syndrome. *G.A. Jervis, H. Monforte, R. Ortega, L. Harris, M.J. Sutcliffe.* Department of Pathology and Laboratory Medicine, All Children's Hospital, John Hopkins Medicine, St. Petersburg, FL.

A male neonate presented with macrosomia, hepatosplenomegaly, right ventricular hypertrophy, macroglossia, neutropenia, and "blueberry muffin rash". Extramedullary myelomonocytic leukemia cutis was diagnosed by pathology from skin biopsy with a concomitant involvement of acute myeloid leukemia (AML), FAB subtype M5 in the bone marrow. Constitutional SNP HD array revealed a terminal 4.8 Mb gain of 11p15.5, a terminal 3.7 Mb deletion of 8p23.3, and a mosaic 10.5 Mb deletion between 17q21.31-q22. Cytogenetics and FISH confirmed an unbalanced der(8)t(8;11)(p23.3;p15.4) with duplication of the Beckwith Wiedemann syndrome (BWS) region; an overgrowth and tumor predisposition syndrome. MS-MLPA detected an imprinting abnormality with hypermethylation of the *H19* locus and normal methylation of the *KCNQ1OT1* locus, consistent with the diagnosis of BWS, likely due to a paternal duplication. This mechanism, found in <1%; of BWS cases, may be associated with an even higher risk for tumor development (Wilms tumor/ hepatoblastoma) and developmental delay. Cytogenetic bone marrow analysis confirmed the der(8)t(8;11)© but also revealed an acquired balanced t(5;17)(q35;q22) in 7/20 cells. Unexpectedly, the 17q© deletion was not detected in the bone marrow (cytogenetic and array negative). Possible explanation for this discrepancy may be tissue specific mosaicism. The clinical effects of this mosaic deletion are unknown. The acquired translocation is interesting because: 1) the *RARA* (retinoic acid receptor alpha) gene had normal signal pattern by FISH, differing from the t(5;17) described in AML M3 that disrupts this gene, 2) 5q35 region is part of a childhood AML translocation, t(5;11)q35;q15.5), that contains the *NPM1* gene, with mutations in AML, and 3) 17q22 region was involved in both the constitutional deletion and acquired translocation, suggesting regional instability. Although there are reports of leukemias (ALL, AML) in older children secondary to BWS, literary search found no references of congenital leukemia with this condition. This may be the first reported case of neonatal AML in a patient with BWS, adding to the complexities of defining this condition.

2555S

Screening Results From 79424 Patients Tested by CFnxt, a 147 Mutation Cystic Fibrosis Screening Assay Built on the Illumina BeadXpress® Platform. *C. Holland, R. Tinawi-Aljundi, M. Weindel, J. Stoerker.* Progenity, Ann Arbor, MI.

Introduction: In March 2011, the American College of Obstetricians and Gynecologists (ACOG) updated its screening recommendations for cystic fibrosis to include offering CF screening to all women of reproductive age. The current ACOG mutation panel offers a comparatively limited risk-reduction for racial minority patients. We developed a CF screening assay that interrogates 147 mutations and 4 variants and provides greater risk reduction across all ethnicities. Methods: The assay consists of 12 multiplex PCR reactions, treatment of the product with a SAP/Exo I mixture, 12 multiplex allele-specific primer extension (ASPE) reactions with biotinylated CTPs, consolidation of ASPE reactions, hybridization onto the VeraCode® Bead set (Illumina), binding of the streptavidin-Alexa Fluor® (Invitrogen) conjugate, and scanning of the beads by the BeadXpress®. 96 samples and controls are run per batch in 384-well plates. Sample transfer is assisted by automated liquid handling. All putative positive samples are repeated for confirmation. The run-time for CFnxt is approximately 12 hours. Results: Through 2 June 2014, Progenity has completed testing on 79,424 patients. A total of 2484 heterozygotes and 17 compound heterozygotes were confirmed, resulting in an overall confirmed positive rate of 3.15%. Of the heterozygotes detected, 1944 were part of the ACOG panel (77.7%), and 557 were non-ACOG panel mutations (22.3%). A total of 22 out of 23 ACOG mutations were identified, along with 67 non-ACOG mutations. This includes 557 patients that would not have been otherwise identified. The most common mutations observed were deltaF508 (1368), R117H (237), W1282X (62), G542X (61), D1152H (60), G551D (36) and 3849+10kbC>T (27). In addition, out of the top 20 mutations identified, 5 were non-ACOG mutations. A total of 251 samples resulted in failed analysis (0.32%). Conclusions: The CFnxt assay provides robust screening across a broad range of CF mutations, thereby delivering significantly increased carrier detection for all ethnicities, compared to the ACOG panel, and an improvement in identification over other expanded panels.

2556M

A novel StripAssay identifies genetic variants modifying beta-thalassemia disease severity. *C. Oberkanins, B. Rauscher, H. Puehringer.* ViennaLab Diagnostics, Vienna, Austria.

Background: The clinical phenotype of patients with beta-hemoglobinopathies is extremely heterogenous, ranging from nearly asymptomatic forms of thalassemia intermedia to severe transfusion-dependent thalassemia major. The wide phenotypical variability is associated with the type of beta-globin mutation, the co-inheritance of alpha-thalassemia and the ability for persistent production of fetal hemoglobin (HbF) in adult life. For the latter, three different quantitative trait loci, accounting for 20-50% of HbF variation, have been identified by now. Single nucleotide polymorphisms (SNPs) in the gamma-globin gene promoter (HBG2), in the BCL11A gene and the HBS1L-MYB intergenic region lead to increased residual HbF levels in adults. Methods: A teststrip-based reverse-hybridisation assay was developed for the simultaneous detection of SNPs in the HBG2 (g.-158 C>T), BCL11A (rs1447407, rs10189857), HBS1L-MYB (rs28384513, rs9399137) genes. Results: The new StripAssay enables the concomitant identification of genetic variants known to influence beta-thalassemia disease severity. Based on the presence of positively modifying alleles, and combined with alpha- and beta-globin genotyping, it allows the prediction of patients likely to display less severe phenotypes. Favourable properties, such as the rapid DNA extraction protocol, ready-to-use reagents and teststrips, as well as the potential for automation of the hybridisation/detection and interpretation steps, make the StripAssay convenient and easy to perform within less than six hours. Conclusions: Testing for genetic modifiers influencing disease severity will lead to more specific and effective treatment, and support clinical decisions regarding the beginning of transfusion therapy in beta-thalassemia patients. Furthermore, the knowledge about prognostic markers has implications for genetic counselling and prenatal diagnosis.

2557T

A multidisciplinary approach to the evaluation and care of patients with inherited lung disease. *B. Raby[1,2], D. Toledo[3], L.M. Yonker[4], E. Henske[1], T.B. Kinane[4].* 1) Division of Pulmonary and Critical Care Medicine, Brigham & Women's Hospital, Boston, MA; 2) Channing Div. Network Medicine, Brigham & Women's Hospital, Boston, MA; 3) Partners Center for Personalized Medicine, Boston, MA; 4) Massachusetts General Hospital, Boston MA.

Like other medical disciplines, clinical pulmonary medicine is entering the post-genome era, where increased accessibility to genetic testing, converging with greater public awareness of the role of genetics in health, poses new challenges to pulmonologists. Moreover, most clinical training programs provide little exposure to monogenic conditions other than alpha-1-antitrypsin deficiency and cystic fibrosis. To address this, we established the Brigham and Women's Hospital Pulmonary Genetics Center, a multidisciplinary program that centralizes the diagnosing, managing, and counseling of patients and at-risk family members with genetic lung diseases. All patients are evaluated by a pulmonary geneticist in conjunction with a certified genetic counselor. Patients are triaged for appropriate genetic testing, and for subsequent evaluations by specific subspecialists, including pediatric pulmonologists, adult medical geneticists, a diverse team of pulmonary subspecialists with expertise in rare lung diseases, and the lung transplant team. Clinical fellows from both the pulmonary and medical genetics fellowship programs formally attend the clinic as part of their training. Partnering with the Laboratory for Molecular Medicine (LMM) of the Partner's Center for Personalized Medicine, we developed a CLIA-certified next-generation sequencing panel capable of evaluating more than 60 genes related to diverse inherited lung diseases, including familial forms of cystic disease, pulmonary fibrosis, bronchiectasis, and pulmonary hypertension. All patients are counseled both prior to and following testing. Here we describe the first 75 referrals to the center, outlining the spectrum of patients evaluated, their experiences undergoing genetic testing, and clinical outcomes resulting from this testing. The most common reasons for referral were for evaluation of patients with fibrosis or cystic lung disease with a family history of similar, and for young adults presenting with unusual constellations of clinical findings suggesting rare genetic disease. Pathogenic variants have been identified for the majority of patients tested. The most common diagnosis, made in 10% of referred patients, has been fibrosis caused by short telomere syndrome. The initial case series also includes families with Birt-Hogg-Dube Syndrome, Blau Syndrome and Primary Ciliary Dyskinesia. The implications of confirmed diagnoses are described, highlighting the utility of this approach.

2558S

Identification of truncating mutations in the *CHD8* gene in patients with autism spectrum disorders by clinical whole exome sequencing (WES). *J. Zhang[1], M.S. Leduc[1], Z. Niu[1], R.E. Person[1], W. Alcaraz[1], E.E. Roeder[3], T. Moss[4], J.M. Nguyen[5], M.R. Wagle[6], P.A. Ward[1], A.A. Braxton[1], T.R. Vaughn[1], D.M. Muzny[2], A.L. Beaudet[1], R.A. Gibbs[1,2], C.M. Eng[1], F. Xia[1], Y. Yang[1].* 1) Dept Molec & Human Gen, Baylor Col Med, Houston, TX; 2) The Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Department of CHSA Pedi-Genetics, Texas Children's Hospital, TX; 4) Genomic Medicine Institute, Cleveland Clinic, OH; 5) Department of Pediatrics, University of Texas at Houston, TX; 6) Department of Pediatrics-Neurology, Texas Children's Hospital, TX.

Autism spectrum disorder (ASD) is a complex neurodevelopmental disability characterized by impaired social interactions, communication difficulties and restricted and repetitive stereotyped behaviors, with an overall prevalence of 62 cases per 10,000 people. Genome-wide studies such as chromosomal microarray and whole exome sequencing (WES) have identified *de novo* changes in numerous genes in patients with ASD. While most of those genes have yet to be validated as candidate autism genes, a few such as *CHD8* and *ADNP* have been better defined. *CHD8* encodes chromodomain helicase DNA binding protein 8 and was previously linked to autism in nine patients in one report in 2012 (PMID 23160955). Here we report four additional cases referred for clinical WES who were found to harbor novel truncating mutations in *CHD8*. Patient 1 was a 5.5 year old female with clinical phenotypes including expressive language delay, dyspraxia, motor delay, hypotonia, dysmorphic features, tall habitus, macrocephaly, obesity, strabismus, and scoliosis. She was found to carry a heterozygous *de novo* c.4611dupA (p.V1538fs) mutation in the *CHD8* gene. Patient 2 was a 14 year old male presented with autism, intellectual delay with regression, attention deficit disorder, hypotonia, ataxia, tremors, tics, dysmorphic features, nausea, constipation, seizures, retinal disease and scoliosis. This patient had a heterozygous *de novo* c.2565del (p.N855fs) mutation in the *CHD8* gene. Patient 3 was a 4.5 year old female with autism, global developmental delay, speech delay, hypotonia, dysmorphic features and macrocephaly and a heterozygous c.4468dupT (p.C1490fs) mutation in the *CHD8* gene. The parental samples were not available for study. Patient 4 was a 5 year old female with autism, severe intellectual disability, speech delay, hypotonia, dysmorphic features, macrocephaly, and hyperextensibility. WES revealed a heterozygous c.6518C>A (p.S2173X) in the *CHD8* gene, Sanger sequencing showed that the mother is negative for this mutation; the father was not available for study. Overall, the four cases with truncating *CHD8* mutations identified by WES further expand the phenotypic spectrum of *CHD8*-related disorders. The occurrence of *CHD8* mutations in our clinical WES samples is four out of a total of 1800 unrelated patients referred for autism spectrum disorder and intellectual disability with or without other organ system involvement.

2559M

Use of a Patient-Centered Conceptual Framework to inform the development of the ClinGen Resource. *M.S. Williams[1], S. Aronson[2], H.L. Rehm[3], S. Goehringer[1], A. Milosavljevic[4], E.M. Ramos[5], Clinical Genome Resource Consortium.* 1) Genomic Med Inst, Geisinger Health System, Danville, PA; 2) Executive Director of Information Technology, Partners HealthCare Personalized Medicine Boston, MA; 3) Dept. of Pathology, Brigham & Women's Hospital and Harvard Medical School Director, Laboratory for Molecular Medicine, Partners Personalized Medicine Boston, MA; 4) Department of Molecular and Human Genetics Baylor College of Medicine Houston, TX; 5) Division of Genomic Medicine, National Human Genome Research Institute, NIH Bethesda, MD.

Background Healthcare systems represent a complex 'system of systems' resulting in challenges in coordination, access, communication and integration that need to be addressed in order to improve the quality and decrease the cost of care. One such system is the Clinical Genome Resource, a product of the NIH-funded Clinical Genome Resource consortium (ClinGen). The ClinGen resource will provide information about actionable genomic information to clinicians at the point of care. As the integration of the resource into the healthcare system was being contemplated, the consortium agreed to explore a patient-centered conceptual framework. The key insight in this framework is that the only constant in the care delivery system is the patient and the patient's caregivers. Following this framework, we reimagined the ClinGen resource information delivery process from the patient/caregiver perspective—so-called patient-centered care. **Methods** The patient-centered framework created by the UK National Health System Heart Improvement Programme was used. The interaction of the patient/caregivers with their providers within this framework is termed the clinical microsystem. A workgroup of ClinGen was convened to use this framework to conceptualize the organization and relationship of the elements that would interact with the ClinGen resource either directly or indirectly. **Results** The workgroup used an iterative approach to develop a draft 'system map'. This was presented to the ClinGen Consortium steering committee for comment and discussion which were incorporated into a final draft. This draft defined the critical applications in the ecosystem such as Electronic Health Records and Laboratory Information Systems, dbGaP, ClinVar, and other external data sources and identified key interactions and data flows that needed to occur to achieve optimal utility. The final design was approved by the steering committee. **Conclusion** The approved system map will be used to construct and implement the ClinGen resource. This means that the project is committed to effectively delivering information to the clinical microsystem such that the goal of improving health through the use of evidenced-based genomic information can be realized. Currently the various ClinGen workgroups are taking ownership of the different interactions and data flows to determine the specifications and standards that will be needed to optimize each interaction and system dependency.

2560T

Mutation analysis of *CYP21A2* and correlation between genotype - phenotype in 163 Vietnamese patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *V. Dung[1], K. Tran[2], P. Le[2], K. Nguyen[1], H. Le[1], M. Fukami[3], V. Ta[2].* 1) Medical Genetics and Metabolism, National Hospital of Ped, Hanoi, Viet Nam; 2) Hanoi Medical University, Research Center of Gene and Protein, Hanoi, Vietnam; 3) National Research Institute for Child Health and Development, Department of Molecular Endocrinology, Tokyo, Japan.

Background: Congenital adrenal hyperplasia (CAH) is one of the most common inherited metabolic disorders. The most common form of CAH (95%) is caused by mutations in *CYP21A2*, the gene encoding the adrenal steroid 21-hydroxylase enzyme (21-OHD; OMIM 201910). Objective: To identify the mutations in the *CYP21A2* gene in Vietnamese patients with CAH and attempt a genotype-phenotype correlation. Methods: Molecular analysis was performed using PCR, multiplex ligation dependent probe amplification and direct sequencing of PCR products of the *CYP21A2* gene in 163 CAH patients. Correlation between phenotype and genotype was evaluated based on identified mutations and clinical manifestations. Results: Mutations were identified in 321 alleles. Twelve different causative mutations with 29 different genotype were identified in *CYP21A2* including two novel mutations. The most frequent genetic defect was c.293-13A/C>G (115 alleles; 35.8%) mutation, followed by Large deletion (88 alleles; 27.4%); c.1066C>T (p.R356W) (49 alleles; 15.3%); c.515T>A (p.I172N) (33 alleles; 10.3%). The rarer mutations were c.952C>T (p.Q318X) (9 alleles; 2.8%); c.1276C>T (p.R426C) (7 alleles; 2.2%); c.920_921insT (L307fs) (7 alleles; 2.2%); c.1447_1448insC (p.R483fsX541) (6 alleles; 1.9%); c.737delA (E246fs) (1 allele; 0.3%); c.3G>A (p.M1I) (1 allele; 0.3%); and novel mutations p.I112X (1 allele; 0.3%) and p.S125X (4 alleles; 1.2%). The majority of patients (127 cases; 77.9%) were homozygotes. Sixteen cases were compound heterozygous. Genotype accurately predicted phenotype in 97.3 and 95.3% of patients with salt-wasting (SW) and simple virilizing (SV), respectively. 44 cases with homozygous mutation c.293-13A/C>G are associated with SW phenotype and 3 patients present with SV form. Conclusions: The spectrum of mutations of the *CYP21A2* gene in Vietnamese patients is comparable to the reported in other populations. Large deletion accounts for nearly one-third of the genetic defects. Therefore, laboratory should include methods for detecting point mutations as well as large deletions. Genotype-Phenotype correlation was high in the studied patients.

2561S

Preventive effects of α-tocopherol on telomere shortening in human buccal cells. *S. Yabuta, Y. Shidoji, M. Masaki.* Graduate School of Human Health Science, University of Nagasaki, Nagasaki, Nagayo, Japan.

[Background and Objectives] Telomere shortening is considered as a solid index for senescence of human cells. However, telomere length may be affected by both physical aging and oxidative stress. In as much as telomere shortening is caused by oxidative stress, it should be preventable with meal intervention (Blackburn, E et al., JAMA 2010;303:250-257). According to the previous results, oxidative stress may cause genomic DNA damages as well as suppression of telomerase activity, which is also active in certain somatic cells. In this context, habitual intake of antioxidant nutrients such as β-carotene and vitamin E may be important to prevent telomere shortening. The present study addresses interactions between the relative telomere length (RTL) in buccal cells and anti-oxidant vitamin-related genes including BCMO1, SR-B1, and ISX genes. These three gene products are involved in intestinal uptake and metabolism offat-soluble vitamins. Finally, we analyzed a relationship between buccal RTL and daily intake of antioxidant nutrients. [Methods] The subjects consist of Japanese men and women (n=70, aged 20-59y) voluntarily recruited with a written informed consent. Telomere length was measured by monochrome multiplex quantitative PCR method with buccal cells. The telomere length was expressed as a relative telomere length (RTL) comparing with a reference sample. BCMO1 (rs6564851), SR-B1 (rs2278986), and ISX (rs362090) weretypedby real-time PCR method with TaqMan probes. The data of nutrient intake were collected by food frequency questionnaire(FFQ). [Results and Conclusions] Allele frequencies of G and T in rs6564851 were 0.836 and 0.146, respectively and these values fit into those of Japanese population uploaded in SNP database. In association of RTL with the gene polymorphisms, there was a positive and significant correlation between buccal RTL and daily intake of α-tocopherol in men of BCMO1 T-carriers ($p<.05$). When BCMO1 T-carriers were further classified by ISX AA type and ISXG-carrier type, only the ISX AA type showed a positive correlation between buccal RTL and α-tocopherol intake ($p<.05$). Further in vitro experiment is under way to validate a protective effect of α-tocopherol on the telomere shortening, where the role of hydrogen peroxide as oxidative stress may be modified by combined genotype of BCMO1 T-carrier and ISX AA homozygote in a human cell line.

2562M

Expanded Screening for Pathogenic Mutations in the Ashkenazi Jewish Population. *B. Baskovich[1], S. Hiraki[2], K. Upadhyay[2], P. Meyer[2], S. Carmi[3], N. Barzilai[2], A. Darvasi[4], K. Offit[5], S. Bressman[2], L. Ozelius[6], I. Peter[6], J. Cho[6], G. Atzmon[2], L. Clark[3], T. Lencz[2], I. Pe'er[3], H. Ostrer[1,2], C. Oddoux[1,2].* 1) Montefiore Medical Center, Bronx, NY; 2) Albert Einstein College of Medicine, Bronx, NY; 3) Columbia University, Manhattan, NY; 4) Hebrew University of Jerusalem, Givat Ram, Jerusalem; 5) Memorial Sloan Kettering Cancer Center, Manhattan, NY; 6) Mount Sinai School of Medicine, Manhattan, NY.

Introduction: A number of founder and recurrent mutations are prevalent in the Ashkenazi Jewish population. Screening for these mutations has reduced the incidence of many autosomal recessive diseases. Yet, only 22 conditions are available on current Ashkenazi Jewish screening panels. The purpose of this study is to expand the knowledge base of pathogenic mutations in this population and develop an expanded screening panel. **Methods:** Whole-genome sequencing was performed by Complete Genomics on 128 disease-free Ashkenazi Jews, which identified 13,768,157 variants. These variants were passed through a pipeline which queried the Online Mendelian Inheritance in Man (OMIM) and ClinVar databases. The pipeline yielded 201 variants that were called pathogenic in these databases. These variants were manually curated through literature review and scored for pathogenicity. The literature was also reviewed for known Ashkenazi Jewish mutations. The results were combined and weighted by clinical utility and allele frequency. **Results:** In addition to the 22 conditions that are currently screened for, an additional 69 conditions that have pathogenic alleles were identified. In total, 168 significant pathogenic mutations were established for 52 autosomal recessive, 34 autosomal dominant and 5 X-linked disorders. The diseases include those that inform risk to offspring, which are generally autosomal recessive, and range from very severe, such as glycine encephalopathy and Leigh syndrome, to milder, such as hyperoxaluria and ichthyosis. Some of these disorders are informative for adult-onset conditions that may convey risk of future cancer (*BRCA1/2* and mismatch repair testing), obesity, or macular degeneration or may have atypical presentations. Some mutations convey different phenotypes in the monoallelic and biallelic states; thus, testing would offer information pertaining to both personal risk as well as risk to offspring. **Conclusions:** As screening for a larger number of mutations becomes possible, we need to reframe our goals of population screening and genetic counseling. An expanded panel can further reduce the incidence of autosomal recessive conditions, reduce morbidity and mortality by offering earlier diagnosis, and satisfy a need for information and enable planning for future life events. The number of mutations identified could be feasibly screened for on a population-wide basis and are individually selectable following genetic counseling.

2563T

Clinical exome sequencing: Initial experience in a pediatric and biochemical genetics clinic. *J.A. Bernat[1], A. Ahmad[1], M.C. Hannibal[1], C.E. Keegan[1,2], B.C. O'Connor[1], K.M. Owens[1], S.C. Quinonez[1], J.L. Schuette[1,2], A.H. Seeley[1], J.G. Thoene[1], B.M. Yashar[2], D.M. Martin[1,2], J.W. Innis[1,2].* 1) Division of Pediatric Genetics, Metabolism, and Genomics, Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Technological advancements in next-generation sequencing have allowed whole exome sequencing to become feasible as a clinical test in cases of suspected genetic disease when initial tests have failed to yield a molecular diagnosis. Although many single case reports exist, there is limited data regarding molecular diagnostic success rates of exome sequencing in a medical genetics clinic setting. Initial estimates have suggested a diagnostic yield of 15-30% in select cases. Recent publications relevant to incidental findings (pathogenic variants in genes unrelated to the patient's phenotype) have also shown that this test has a broad and complex potential impact. We quantified the diagnostic yield and incidental findings in exome sequencing cases in the University of Michigan Pediatric and Biochemical Genetics Clinics, from October 2012 to the present. Cases were identified in which clinical exome sequencing tests were sent by UM Pediatric Genetics faculty. Clinic notes were reviewed for phenotype information, previously-ordered studies, and family history. Research-based testing and next-generation sequencing-based focused gene panel testing were excluded. To date, results have returned for 39 tests. In 12 cases (31%), pathogenic variants in disease-causing genes related to the patients' phenotypes were identified. New diagnoses included Bainbridge-Ropers syndrome (MIM 615485), Kabuki syndrome (MIM 147920), Myhre syndrome (MIM 139210), Opitz-Kaveggia syndrome (MIM 305450), triosephosphate isomerase deficiency (MIM 615512), and Van Maldergem syndrome (MIM 615546). 23 of the 27 remaining cases had variants of uncertain significance (VUSs) reported. No patients or families opted out of receiving a report of incidental findings. 5 cases (13%) had incidental findings in one of the 56 genes recommended to be reported by the ACMG, with pathogenic variants identified in *BRCA1*, *BRCA2*, *MYBPC3*, and *MYH7*. Early results suggest that clinical exome sequencing is a promising new technology that can lead to a molecular diagnosis in a subset of select cases. When applied properly, such testing may lead to a diagnosis more quickly and likely at less cost in both monetary and psychological terms. An understanding of the limitations of this testing and the possibility of incidental findings is necessary for its proper application. Additional work will be required to determine the molecular etiologies for cases in which exome sequencing is negative.

2564S

Implications of Massively Parallel Sequencing in screening for Autosomal Recessive conditions: the risk of being a "genetic wallflower". *L. Burnett[1,2,3], D. Chesher[1], L. Nguyen[1], R.M. Lew[4], A. Proos[1], L.C. Ding[1], L. Koe[1].* 1) NSW Health Pathology North, Sydney, NSW, Australia; 2) Sydney Medical School, University of Sydney, NSW, Australia; 3) School of Information Technologies, University of Sydney, NSW, Australia; 4) Department of Obstetrics and Gynaecology, QEII Research Institute for Mothers and Infants, University of Sydney, NSW, Australia.

Introduction Genetic screening for autosomal recessive (AR) genetic carriers is available in many communities "at risk" due to high prevalence of pathogenic variants. Examples include Tay-Sachs disease, thalassaemia and cystic fibrosis. The range of tests included in screening programs is broadening, and will accelerate with the introduction of Massively Parallel Sequencing. As more genes are included in screening programs, a risk emerges of being a "genetic wallflower", rejected by every suitor because of the AR genes they carry. We recently modelled the frequency of autosomal dominant (AD) genetic conditions arising as Incidental Findings (IF) in Whole Genome Sequencing. We found that the proportion of tested individuals with significant IFs plateaus to a limit even as the number of genes tested increases beyond those in the ACMG IF Recommendations.

Aim To model the rate of increase in AR genetic carriers with increasing numbers of AR genes in test menus. We also explore the likelihood of genetic wallflowers emerging.

Methods A mathematical model based on binomial distribution was used to predict the number of AR carriers based on prevalence of pathogenic variants and number of tested conditions. Monte Carlo simulation was used to calculate the probability of any two individuals in this screened population being carriers of exactly the same AR genetic conditions.

Results The model was validated using known rates of AR genetic carriers found in Australian Ashkenazi Jewish community screening programs. We have found that the numbers of AR genetic carriers will rise as the number of tested conditions is increased. We also report the likelihood of genetic wallflowers emerging in AR genetic carrier screening programs.

Conclusions As the number of AR conditions included in testing panels is increased, the number of AR genetic carriers identified in a population will increase. This behaviour is in marked contrast to that for AD conditions, where the proportion of a population with a reportable IF will plateau. These findings have significant implications for health economic evaluation and planning. Finally, we report on the ethical and social implications of genetic wallflowers emerging.

2565M

Analyses of *TSC* genes among Brazilian tuberous sclerosis complex (TSC) patients. *L.G. Dufner-Almeida[1], J.P.G Almeida[2], L. Masulk[3], M. Richartz[3], M. Miranda[3], S.A. Antoniuk[3], S. Rosemberg[2], L.A. Haddad[1].* 1) Department of Genetics and Evolutionary Biology, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil; 2) Department of Pediatrics, Division of Child Neurology, Faculdade de Medicina da Santa Casa de São Paulo, São Paulo, Brazil; 3) Department of Pediatrics, Universidade Federal do Paraná, Curitiba, Brazil.

Tuberous sclerosis complex (TSC) is a multisystem disorder, with variable expression and autosomal dominant inheritance. Clinically it is due to hamartia and hamartoma development in different tissues, notably in the brain, kidneys, heart, skin and lungs, causing organ dysfunction. Mutations in either tumor suppressor genes *TSC1* or *TSC2* are responsible for TSC. TSC1/TSC2 protein heterodimer inhibits the mammalian target of rapamycin (mTOR) complex 1, controlling cell growth and proliferation. Although TSC diagnosis is basically clinical, since the 2012 specialist panel review, the finding of a *TSC1* or *TSC2* pathogenic mutation has been considered sufficient for the definite diagnosis of the disease. In addition, mTOR inhibitors have been clinically approved to treat under specific guidelines three hamartomas that afflict TSC patients. To set up the basis for molecular diagnosis of TSC and functional analysis of *TSC1* and *TSC2* mutations in São Paulo, Brazil, an ongoing project aims to establish and compare at long term Sanger and massively parallel technologies for *TSC1* and *TSC2* DNA sequencing. As a preliminary step, we conducted Sanger sequencing of the *TSC1* gene, encompassing its full coding sequence, an average of 132 bp of intronic segments next to exon boundaries, in addition to the gene core promoter of 28 Brazilian patients with definite TSC diagnosis. Seven patients (25%) displayed *TSC1* nonsense/frameshift mutations. Among 31 other DNA variants identified, 27 were DNA polymorphisms. Two and one additional DNA point variants from the same patient flanked a putative, specific transcription factor binding site, 5' to the *TSC1* core promoter. In addition, a novel DNA variant residing in the *TSC1* noncoding exon 2 was predicted to change the sequence potential to behave as a splicing enhancer. In summary, we describe 25% of TSC patients with pathogenic mutations in the *TSC1* coding sequence. Moreover, our data disclose four novel DNA variants in *TSC1* potentially regulatory regions that are likely to unravel novel pathogenic mutations, and thus need to be experimentally tested.

2566T

The Implementing Genomics in Practice (IGNITE) Network: A Coordinated Effort to Study and Improve Implementation of Genomics in Clinical Practice. *S.E. Kimmel[1], G.S. Ginsburg[2], E.P. Bottinger[3], J.C. Denny[4], D.A. Flockhart[5], T.I. Pollin[6], E.B. Madden[8], K. Weitzel[7], M. Alexander[1], C. Horowitz[3], L.A. Orlando[2], N.S. Calman[3], P.R. Dexter[5], M. Levy[4], A.R. Shuldiner[6], H.A. Junkins[8], J.A. Johnson[7], IGNITE Network.* 1) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA., United States; 2) Duke University Medical Center, Duke University, Durham, NC; 3) Icahn School of Medicine at Mount Sinai, Mount Sinai School of Medicine, New York, NY; 4) Vanderbilt University School of Medicine, Vanderbilt University, Nashville, TN; 5) Indiana University School of Medicine, Indiana University, Indianapolis, IN; 6) University of Maryland School of Medicine, University of Maryland, Baltimore, MD; 7) College of Pharmacy, University of Florida, Gainesville, FL; 8) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Purpose: Despite great progress in understanding the impact of genomics on patients' health and outcomes from therapy, implementation of genomics in practice remains challenging. There is an urgent need to better understand the methods of incorporating patients' genomic information into clinical care, evaluating implementation efforts, and measuring outcomes of genomics implementation. Methods: The aims of the Implementing Genomics in Practice (IGNITE) Network, funded by the National Human Genome Research Institute and the National Cancer Institute of the NIH, are to assess methods of implementation of genomic information in practice within specific projects, to rigorously evaluate these methods across projects in order to create generalizable knowledge, and to provide proofs of concept of the value of genomics in medical care. Special interest has been placed on partnering between established and less experienced groups; inclusion of diverse settings and underserved populations; and optimal utilization of electronic medical records (EMR) and clinical decision support (CDS) tools. Demonstration projects and a Coordinating Center have been funded to incorporate genomic information (genetics, pharmacogenetics, and family history) into the EMR and provide CDS for implementation of appropriate interventions or clinical advice across multiple and diverse sites. IGNITE projects include incorporation of structured family history tools, genotyping protocols for more accurate diagnosis and customized treatment of disease, and germline and somatic pharmacogenomics. With the assistance of a Coordinating Center, the projects are using established frameworks for implementation research to define barriers to implementation, develop and disseminate solutions to these barriers, and assess outcomes, especially those most important to ensuring rapid and sustained adoption, such as patient and provider satisfaction, clinical outcomes, and cost effectiveness. Summary: The IGNITE Network's goal is to determine what works and what does not in genomic medicine implementation in practice and to provide generalizable knowledge that can be used to foster similar implementation in other settings, for other diseases, and for the myriad of potentially useful genetic-based testing and histories. Details of the IGNITE projects and the challenges and lessons learned from implementation can benefit the scientific and clinical community.

2567S

Clinical exome sequencing identifies a novel gene, LINS, associated with intellectual disability, failure to thrive, seizures, dysmorphology, and language regression. *I. Lu[1], L. Shahmirzadi[1], R. Baxter[1], S. Tang[1], K. Gonzalez[1], E. Rosebrough[2], O. Abdul-Rahman[2].* 1) Ambry Genetics, Aliso Viejo, CA; 2) University of Mississippi Medical Center, Jackson, MS.

Clinical Whole Exome Sequence analysis of an 8 year-old Caucasian female affected with intellectual disability (ID), failure to thrive, myopia, microcephaly, seizures, dysmorphic features, and language regression with aphonia identified two compound heterozygous alterations in LINS, the human homolog of the Drosophila segmentation lins protein, which is involved in the Wnt signaling pathway. The paternally-inherited c.2020dupA frameshift alteration and maternally-inherited c.1394+1G>T splice site alteration are both expected to result in a deleterious effect on the LINS gene. The patient evaded diagnosis through clinical evaluation and extensive genetic testing over many years including negative chromosome analysis, FISH for Smith Magenis syndrome, CGH/SNP microarray, Angelman syndrome methylation analysis, MECP2 and CDKL5 analysis for Rett syndrome, mitochondrial DNA array and gene sequencing panel, as well as uninformative biochemical results. The two LINS gene alterations we identified are likely to provide an explanation for the patient's symptoms and no other likely candidate gene alterations were identified to explain the patient's phenotype. LINS is a newly characterized gene involved in human cognition and brain development. Mutations in LINS have been described in only two cases of autosomal recessive ID and microcephaly. In one previously reported case, a homozygous frameshift mutation was identified in an Iranian family with four children affected with moderate ID and microcephaly. In a separate case, a 5-nucleotide homozygous deletion affecting a donor splice site was identified in a Yemeni male and female sibship affected with early onset ID, developmental delay and head nodding behavior. Consanguinity was present in both of the previously reported cases. Phenotypic overlap with the current case include microcephaly and intellectual disability; however, there appears to be a spectrum of clinical symptoms associated with LINS mutations. Diagnostic exome sequence analysis is an effective method for contributing to the mutational and phenotypic spectrum of LINS-related diseases.

2568M

Genetics of beta thalassemia in Iran: Still requires consideration? *N. Mahdieh[1], B. Rabbani[2], R. Shiri Heris[3].* 1) Deputy of Research and Technology, Ministry of Health and Medical Education, Tehran, Iran; 2) Medical Genetic Group, Qazvin University of Medical Sciences, Qazvin, Iran; 3) Faculty of Paramedicine, Ilam University of Medical Sciences, Ilam, Iran.

Beta thalassemia is observed in many countries. Prevalence of thalassemia-causing mutations is different among various parts of the world and even within different ethnic groups living in Iran. In addition to clinical, biochemical and hematologic diagnostic methods, molecular diagnostics is used for prenatal diagnosis of thalassemia. We analyzed all published data about the frequency of beta globin mutations among Iranian subjects. Mutation frequency was calculated in more than 5000 at risk couples studied in pre-marriage counseling of screening strategies in Iran. Furthermore, the frequency of HBB gene mutation was determined in 2003 major beta thalassemia. On the basis of the present study, more than ninety mutations have been reported to cause beta thalassemia the country. Ten following mutations were common: IVS I-5 (G>C), IVS II-1 (G>A), FSC36/37 (-T), IVS I-110 (G>A), FSC 8/9 (+G), IVS I-1 (G>A), FSC8 (-AA), IVSI-25 bp, CD30 (G>C) and FSC44 (-C), contributing 80 % of all mutations. IVS II-1 (G>A) is common across the country but other mutations might be common among specific ethnic groups. Our results showed the current prevention strategy is an efficient way to reduce the allele frequency of disease.

2569T

Diagnostic Yield of Genetic Testing in the Children's Hospital of Colorado Autism Genetics Specialty Clinic. *N.J.L Meeks[1,2], K. Brown[1,2], B. Miller[1].* 1) Children's Hospital Colorado, Aurora, CO; 2) University of Colorado, Anschutz Medical Campus, Aurora, CO.

Background: Autism spectrum disorders (ASDs) are a group of neurodevelopmental disorders characterized by deficits in social communication, restricted interests, repetitive patterns of behavior and symptom onset early in development. Several well-defined genetic disorders have been identified in patients with ASDs; however, clinical genetic testing is diagnostic in only a minority of cases. The purpose of this study was to determine the diagnostic yield of genetic testing in a population of children with ASDs seen at the Autism Specialty Genetics Clinic, part of the Autism Treatment Network, at the Children's Hospital Colorado. Methods: A retrospective chart review was performed of patients seen in the Autism Genetics Clinic. Patients who met inclusion criteria had the following: an initial or follow-up genetics evaluation within the past six years; a diagnosis of an ASD through formal evaluation; and at least one molecular genetic test completed and resulted. Results: Of 130 charts reviewed, 107 patients met study criteria. The average number of molecular tests ordered per person was 3.05 [95% CI: 2.74-3.35]. Eighteen patients (16.82%) had molecular genetic testing that was diagnostic. Copy number changes found on chromosomal microarray (CMA) accounted for 67% of all pathogenic findings with a yield of 13.9%. Other pathogenic findings included mutations in the PTEN, MECP2 and PTPN11 genes, positive FISH testing for a familial mutation, and two patients with abnormalities detected on karyotype. Biochemical screening for metabolic conditions, FMR1 testing, and multi-gene intellectual disability panels evaluating for more than one genetic condition did not yield any pathogenic findings. Of all molecular testing completed, 21.5% identified a variant of uncertain significance. Discussion: The results of this chart review are comparable to previously published studies on genetic evaluations of children with ASDs with an overall diagnostic yield of genetic testing at 16.82%. Unlike previously published studies, almost every patient (95%) had a chromosomal microarray as part of their evaluation. These results continue to support the use of CMA as a first line evaluation in children with ASDs. Also, 1 in 5 molecular or cytogenetic tests resulted in a variant of uncertain significance. Incorporating this information in clinic visits will improve pre-test genetic counseling in this population.

2570S

Whole exome sequencing as a diagnostic tool for complex neurological disorders. *G.R. Monroe[1], G. Frederix[2], S. Savelberg[1], K. Duran[1], M.M. van Haelst[1], G. Visser[1], P.M van Hasselt[1], P. Terhal[1], H. Kroes[1], J. van der Smagt[1], A. Hövels[2], G. van Haaften[1].* 1) UMC Utrecht, Utrecht, Netherlands; 2) Utrecht University Pharmacoepidemiology and Clinical Pharmacology, Utrecht, Netherlands.

Whole exome sequencing (WES) provides a straightforward detection of rare variation including de novo mutations in parent-child trios, but a systematic interpretation of the diagnostic yield and an assessment of the costs associated with implementing WES in the clinic is currently lacking. We determine the increased diagnostic yield and provide a thorough cost analysis of the clinical care with and without the use of WES. This study will serve as a guide for centers intending to implement WES and provide a clear diagnostic benefit versus the associated costs of obtaining this increased diagnostic yield. Twenty children of healthy, unrelated parents were randomly selected from the 2011 patient population at the Sylvia Toth Center (STC) in Utrecht, the Netherlands. The STC is a specialized center for children with complex neurological disorders and diverse phenotypes. DNA libraries were prepared using Kapa Biosystems reagents, enriched using Agilent Sureselect All exon V5 with a custom pooling protocol, and sequenced on the Illumina HiSeq 2500. A list of exonic variants was obtained by filtering against public and our in-house database according to the expected *de novo* inheritance model as well as for recessive and compound heterozygote variants. In parallel, the clinical records of patients were obtained and a cost summary of medical treatments, hospital visits, care, and all other resource use was compiled per patient. This cost was then compared to the cost of care using WES, assessed retrospectively on each patient. Comparing the diagnosis and costs with and without the use of WES gives a clear picture of the economic feasibility of putting WES into standard diagnostic practice at the STC and similar centers. The diagnostic yield from the 13 patients sequenced thus far is 23%, confirming past studies diagnostic yields on intellectual disability cohorts. The three variants found are in genes recently associated with intellectual disability (*ANKRD11, CTNNB1, ANDP*), and the variants are all frame-shift deletions resulting in protein truncation. On average these patients have had numerous visits to the hospital, overnight stays and different diagnostics to unravel the genetic cause of their neurological disorder. It is therefore deemed very plausible that the total cost of the current diagnostic pathway is many times higher compared to WES, indicating the economic feasibility of implementing WES early in the diagnostic pathway.

2571M

Genetic investigation of cystic fibrosis transmembrane regulator mutations in a cohort of consecutive patients candidate for assisted reproductive techniques. *F. Papa[1], M. Rongioletti[1], MB. Majolini[1], C. Vaccarella[1], I. Simonelli[2], V. Mazzucchi[1], A. Luciano[1], P. Pasqualetti[2], G. Liumbruno[1].* 1) Clinical Pathology Department, San Giovanni Calibita Hospital, Rome, Italy; 2) Medical Statistics & Information Technology Fatebenefratelli Association for the Research, Rome, Italy.

Introduction. The present study, investigated the frequency of mutations in the CFTR gene, in a group of consecutive patients candidate for assisted reproductive techniques with the aim of identify subjects carriers of the most severe ones. Methods. 22,416 alleles were screened for 56 CFTR gene mutations utilizing the CFTR INNO-LiPA Results. CFTR mutations were detected in 6.2% of the screened alleles. In the large group of alleles analyzed 93.4% were wt, 4.4% were characterized by mild mutations, and 1.7% by severe or severe/mild mutations. Indeed, the most common severe mutation was ΔF508/N observed in 192/22,416 (0.86%) of all alleles analyzed, followed by the N1303K mutation with the frequency of 36/22,416 (0.16%). Whereas regarding mild mutations, the most frequent was the 5T polymorphism present in 916/22,416 (4.1%). Conclusions.Our results together with previous studies, reinforce the importance of an accurate determination of mutations in the CFTR gene, including the 5T polymorphism, in order to inform the couple of their carrier risk and the possibility on having affected child. Moreover, our findings highlight the potential of genetic screening as a tool to identify possible compound heterozygous subjects without CF-like symptoms.

2572T

Next-generation sequencing for the diagnosis of autism spectrum disorders using genomic capture targeting multiple candidate genes. *L. Rodriguez-Revenga[1, 2, 3], MI. Alvarez-Mora[1, 3], I. Madrigal[1, 2, 3], R. Calvo[4], O. Puig[4], M. Mila[1, 2, 3].* 1) Biochemistry and Molecular Genetics Department, Hospital Clinic, Barcelona, Spain; 2) CIBER de Enfermidades Raras (CIBERER), Barcelona, Spain; 3) IDIBAPS (Institut d'Investigacions Biomèdiques August Pi I Sunyer), Barcelona, Spain; 4) Psychiatry and Child and Adolescent Psychology Service, Hospital Clinic, Barcelona, Spain.

Autistic spectrum disorders (ASD) is one of the most common neurodevelopmental disabilities, with an average estimated global prevalence of 62 cases per 10,000 children and an approximate 4:1 male to female ratio. Over the past few years, the genetic basis of ASDs has been pursued aggressively using all kind of high-throughput genomic analysis technologies: SNP-array, CGH-array, next generation sequencing based techniques and genome-wide association studies. Despite the progress in the identification of several candidate genes and causative genomic copy number variations (CNVs), the vast majority of ASD cases still remain unexplained. An obstacle for molecular diagnosis of ASD patients has been clinical and genetic heterogeneity of patient cohorts in combination with a recently pointed out multihit model of the disease. This model is based on the observation that most of the abnormalities identified have been associated with highly variable phenotypes and seems insufficient to cause ASDs on their own, supporting the hypothesis that CNVs contribute to ASDs in association with other CNVs or point variants located elsewhere in the genome. In this study, we developed a next-generation sequencing-based screening based on the capture of a panel of genes involved, or suspected to be involved in ASD, on pooling of indexed DNA and on paired-end sequencing in an Illumina MiSeq platform, followed by confirmation by Sanger sequencing. A cohort of 44 ASD patients with negative result for arrayCGH was screened to evaluate this strategy in terms of sensibility, specificity, practicability and cost. In silico analysis was performed using DNAnexus® software. Sequencing data provided, on average, 99.75% coverage of the 44 genes selected at more than 100-fold mean depth of coverage. Disease causing mutations were identified in 10 patients resulting in a molecular diagnosis rate of 22.7%. These results demonstrate the efficiency of NGS in performing molecular diagnosis of ASD. The failure of finding recurrent mutations highlights the genetic heterogeneity of ASD. Acknowledgments: This study was supported by Fundación Alicia Koplowitz (AKOPLOWITZ11_006). The CIBER de Enfermedades Raras is an initiative of the ISCIII.

2573S

Prevalence of ACMG Incidental Findings in The Cancer Genome Atlas Germline Samples. *S. Sanga, A. Vladimirova, R.D. Goold, T.M. Klingler.* Station X, Inc., San Francisco, CA.

Our genomes are going to be an integral part of our future health care. Already, exome and genome sequencing are rapidly being integrated into the practice of medicine presenting opportunities in characterizing rare diseases, managing patient-specific treatment in cancer, prenatal screening for disease risk, and identifying novel pharmacogenomic biomarkers. Such sequencing has the potential to reveal medically-valid or useful findings that might be unrelated to the indication under investigation, In recognition of this, the American College of Medical Genetics and Genomics recently published their recommendations for reporting incidental findings in clinical exome and genome sequencing [Green et al., 2013]. The ACMG has recommended that laboratories performing germline clinical sequencing seek and report mutations of the specified classes or types in a select set of genes. They also encourage dialog among stakeholders and the creation an ongoing process for updating these recommendations over time as their use clinical practice provides more information on their utility. Here, in an effort to help assess the clinical utility of the ACMG guidelines for reporting on incidental findings, we apply the recommended guidelines to more than 4,000 germline exomes and genomes currently available as part of The Cancer Genome Atlas, and report our findings. In order to accomplish this analysis rapidly across thousands of samples, we developed a structured query based on the ACMG guidelines and have made it available for general use in GenePool, a genomics software platform designed for cohort-scale genomics data management, analysis, and collaboration.

2574M

Fabry disease diagnosed through family screening. *G. Sarca[1], C. Dragomir[2], E. Severin[3].* 1) Professor C.Angelescu Hospital, Bucharest, Romania; 2) Genetic Lab, Bucharest, Romania; 3) Carol Davila University of Medicine and Pharmacy, Bucharest, Romania.

Background. Early recognition of Fabry disease (FD) is difficult due to the clinical variability, the non-specific signs and symptoms and the rarity of the disease. In the same time, early recognition involves early initiation of enzyme replacement therapy (ERT) known to delay the time and prevent the disease progression. Our study describes the screening for FD using genetic testing after a diagnosis of Fabry disease in a male family member.-Objectives. To detect the disease causing mutation and to clarify the genetic status of asymptomatic family members.Subjects and methods. All 7 family members (both males and females) were genetically tested. Genetic testing included isolated DNA from blood samples and sequence analysis of all coding exons and all intron-exon boundaries of the GLA gene.Enzyme activity levels were evaluated too. Results. The disease causing mutation in the family was the pathogenic GLA mutation (c.485G>A) present in exon 3. Two males were found hemizygotes having one copy of GLA gene mutation and no enzyme alpha GAL activity. All five females were found carrier of the same mutation and had alpha-Gal A activity in the normal range.Younger male received early ERT and expressed a mild disease phenotype (no cardiac and renal involvement) comparing with his older male relative. Conclusion. Measurement of alpha-Gal A enzyme activity is unreliable for carrier detection. Early genetic testing should be considered in younger persons with a family history of Fabry's disease.

2575T

Translating allelic heterogeneity of GJB2 gene to clinical practice in Romanian population with congenital isolated hearing-loss. *E. Severin[1], C. Dragomir[2], A. Stan[2], D.T. Stefanescu[2], L. Savu[2].* 1) Carol Davila University of Medicine and Pharmacy, Bucharest, Romania; 2) Genetic Lab, Bucharest, Romania.

Background. Different alleles within the same gene can cause a similar variant phenotype. Previously published studies showed the allelic heterogeneity of GJB2 gene as main genetic cause of isolated congenital hearing-loss phenotype. The proportional distribution of the different mutations within GJB2 gene varies in different ethnic groups. Objectives. The aim of the present study was to provide a complete and updated spectrum of mutations in GJB2 gene and to identify the most prevalent mutations in Romanian population. Subjects and Methods. To overcome our aims, we used consecutive sample selection in order to obtain a better representation of entire Romanian population. All available persons with congenital hearing-loss were included based on the inclusion criteria (clinical records of congenital sensorineural hearing loss, non-syndromic hearing loss, no related findings, patients with affected siblings but unaffected parents). Testing protocols included ARMS-PCR and DNA sequencing techniques for detection of known mutations or identification of mutations within two genes associated with hearing loss, GJB2 and GJB6. Results. Most prevalent mutation was c.35delG (40.0%) in both homozygotic and heterozygotic forms. The second mutant allele was W24X (8.75%) also founded in homo- or heterozygotic forms, followed by c.-23+1G>A originally named IVS1+1G>A and R127W mutations with lower frequencies. Conclusions. The study reveals: c.35delG mutation as most common one in Romanian population; absence of GJB6 gene deletions; genetic background of congenital hearing-loss in local population; supports improvement of testing protocols (newborn screening, carrier screening) and treatment choices (cochlear implant or other hearing device). In addition, genetic information is used for more precise genetic counselling services.

2576S

Where's the Benefit? Views on Genetic Testing for ASD. *K.B. Shutske[1], H. Starks[1,2,3,4], H.K. Tabor[1,2,3,4], W. Burke[1,2].* 1) Institute for Public Health Genetics, University of Washington, Seattle, WA; 2) Bioethics and Humanities, University of Washington, Seattle, WA; 3) Department of Pediatrics, University of Washington, Seattle, WA; 4) Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA.

Background: Autism Spectrum Disorders (ASD) comprise a clinically heterogeneous group of disorders with a strong genetic component. Current guidelines from ACMG and others recommend genetic testing for all individuals diagnosed with ASD, using aCGH (array Comparative Genomic Hybridization) testing as the first tier test, yet genetic testing for ASD is not consistently implemented in the clinical setting. **Methods**: We conducted semi-structured interviews with 13 parents of children diagnosed with ASD, and 15 providers from multiple specialties involved in the care of patients with ASD to identify how parents and providers define the benefits of and barriers to genetic testing for ASD. **Results**: Although parents were open to testing, it was often a low priority compared to other aspects of their child's care. Some parents expressed confusion about the test and reasons for conducting it, but they were willing to have their child tested, based primarily on their provider's recommendation, to ensure that they were doing "all they could" for their affected child. Wanting a causal explanation for ASD or information for reproductive decision making were motivators for some parents, but for the majority the motivator was "crossing it off the list" of things tried. While some providers were supportive of testing, many providers were ambivalent. Providers often made the decision to offer testing on a case by case basis. Barriers to testing mentioned by both providers and families included difficulty in conducting the blood draw, concern that information might create guilt or blame, and the cost of the test, especially when insurance would not cover it. Additional concerns from providers were stigma, variants of uncertain clinical significance, incidental findings, the added burden to the family with possible waste of time and energy, and the fact that the testing is too complex for many families to understand. The biggest barrier for families was not being aware of it. **Discussion**: Currently there is variability in provider attitude and practices toward genetic testing for ASD, as well as parental uncertainty about the value of testing. When providers recommend testing, parents generally follow their advice but they do so primarily to ensure that they are providing all recommended care. Further discussion among stakeholders is needed to clarify the purpose of genetic testing and its appropriate use in ASD care.

2577M

Pediatric neurodevelopmental disabilities refractory to traditional diagnosis: Diagnostic rate, cost and change-in-care of whole genome versus exome sequencing. *S.F. Kingsmore[1], S.E. Soden[1], C.J. Saunders[1], E.G. Farrow[1], L.K. Willig[1], L.D. Smith[1], J.B. LePichon[1], N.A. Miller[1], D.L. Dinwiddie[2], G. Twist[1], A. Noll[1], B.A. Heese[1], L. Zellmer[1], A.M. Atherton[1], A.T. Abdelmoity[1], J.E. Petrikin[1], N. Safina[1], S.S. Nyp[1], B. Zuccarelli[1], I.A. Larson[1], A. Modrcin[1], S. Herd[1], M. Creed[1], Y. Zhaohui[3], Y. Zuan[3], R.A. Brodsky[3].* 1) Center for Pediatric Genomic Medicine, Department of Pediatrics, and Department of Pathology, Children's Mercy Hospital, Kansas City, KS 84108; 2) University of New Mexico Health Science Center, Albuquerque, NM 87131; 3) Department of Medicine, Johns Hopkins University, Baltimore, MD 21205.

Neurodevelopmental disabilities (NDD), including global developmental delay, intellectual disability, epilepsy and neurobehavioral disorders, affect more than 3% of children and are attributable to single gene mutations at over 1,000 loci. Traditional methods yield molecular diagnoses in less than one half of children with NDD. Whole genome sequencing (WGS) and whole exome sequencing (exome sequencing) can enable diagnosis of NDD, but their clinical and cost effectiveness are not known. In 100 families, with 119 affected children, 41% received molecular diagnoses by exome sequencing or rapid WGS. 11 of 15 (73%) families with critically ill infants were diagnosed by WGS. 30 of 85 families (35%) that were refractory to traditional diagnosis, received diagnoses by exome sequencing. Prior to enrollment, the latter were symptomatic for an average of six years, and their families had expended an average of $18,504 in negative diagnostic tests. Referring physicians reported a change in clinical management and/or clinical impression of the pathophysiology in 46% of newly diagnosed families. If performed at symptom onset, exome sequencing/WGS would have reduced the time to molecular diagnosis by an average of 64 months. It is suggested that trio WGS should be part of the initial diagnostic work-up of NDD.

2578T

Whole exome sequencing identified a *RP2* mutation in a large Turkish family. *E. Koparir[1], O.F. Karatas[1,2], A. Sarac[3], B. Yuksel[3], M.S. Sagiroglu[4], B. Ozer[4], M. Solak[5], M. Ozen[1,6].* 1) Department of Medical Genetics, Istanbul University, Cerrahpasa Medical School, Istanbul, Turkey; 2) Molecular Biology and Genetics Department, Erzurum Technical University, Erzurum, Turkey; 3) Genetic Engineering and Biotechnology Institute, TUBITAK Marmara Research Center, Kocaeli, Turkey; 4) Information Technologies Institute, TUBITAK Marmara Research Center, Kocaeli, Turkey; 5) Department of Medical Genetics, Afyon Kocatepe University, Afyon, Turkey; 6) Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX, 77030, USA.

X-linked retinitis pigmentosa (RP) is a clinically and genetically heterogeneous disorder accompanied by abnormalities of the rod and cone photoreceptors or the retinal pigment epithelium of the retina leading progressive vision loss. Affected individuals suffer from night blindness, constriction of peripheral visual fields and later loss of central vision. All 16 patients have vision loss, however, the disease onset time and severity for each patient are variable. WES has been commonly used as a diagnostic tool to identify the molecular basis of this genetically heterogeneous disorder. We performed WES in a three-generation large Turkish family with 16 affected individuals to screen all retinitis pigmentosa-related genes and to look for novel candidate genes. R282W missense mutation was detected in exon 3 of *RP2* gene and confirmed by conventional Sanger sequencing. Although R282W was initially presented as a polymorphic variant, recent studies suggested that this missense variant might have a pathological potential and affects RP2 expression, its interactions and intracellular localization. Our findings support this hypothesis and strengthen the pathogenic potential of R282W mutation albeit the fact that additional mutations in other genes might be necessary for compete penetrance of retinal degeneration phenotype.

2579S

Whole exome sequencing as a tool to enhance patient care: Experience in a midsize academic clinical genetic setting. *A. Pandya, K. Withrow, T. Causey, V. Pallante, J. Propst, M. Jaworski, R. Lewandowski.* Dept Human & Molecular Genetics, Med Col Virginia, VCU, Richmond, VA.

Whole exome sequencing (WES) is expected to revolutionize the practice of clinical genetics with a paradigm shift in how clinicians reach a diagnosis in patients with an unknown etiology. It holds tremendous potential thereby raising expectations of physicians and families to provide an etiologic diagnosis. We report on our experience in a small subset of patients offered WES to identify an etiology. WES was preauthorized and offered to fifteen patients in the past 18 months since its clinical availability. All families were eager for testing and some initiated conversation about WES. A rigorous consent process followed regarding pros and cons of WES and the potential for incidental findings. All the patients had negative single gene molecular testing and SNP microarray. Results from twelve patients are presented (three pending). A pathogenic variant was identified in five patients (41% yield). Two have changes in an established disease gene: ATRX, associated with alpha thalassemia X-linked MR syndrome and ARID1A, associated with Coffin-Siris syndrome, although the phenotype was not evident clinically. In an adult with ataxia, a variant in the KCND3 gene confirmed a diagnosis of SCA19. One patient had a change in a newly described gene DYNC1H1, allowing a better genotype -phenotype correlation. Although the initial report in another patient suggested no clear pathogenic changes, discussion between the clinician and laboratory personnel resulted in attributing clinical significance to a variant of unknown significance (VOUS), with diagnosis of a rare type of congenital disorder of glycosylation. In two patients a de novo change in novel genes was identified but lack of supporting data or functional studies prevents attribution of clinical significance. In the remaining six patients four have a negative result and two have a VOUS. This report demonstrates some of the challenges and benefits of WES in a clinical setting and offers a learning opportunity to enable wider utility. The yield for an etiology in our sample is high (41%), although far from 100% as expected by families. We have not had an incidental finding reported to date, but the families are open to receiving this information. Despite the paradigm shift brought about by WES with test results guiding the phenotypic delineation, we highlight the importance of a thorough clinical evaluation which guides interpretation of sequence variants of unknown significance.

2580M

Diagnostic exome sequencing establishes molecular diagnoses among patients with gastrointestinal disease. *L. Shahmirzadi[1], K.F. Gonzalez[1], D. El-Khechen[1], Z. Powis[1], C. Alamillo[1], C. Mroske[1], D. Shinde[1], K. Radtke[1], R.M. Baxter[1], B. Tippin-Davis[1], S.K. Gandomi[1], E.C. Chao[1,2], S. Tang[1].* 1) Ambry Genetics, Aliso Viejo, CA, 92656; 2) Department of Pediatrics, University of California, Irvine, Irvine, CA 92697.

Diagnostic exome sequencing (DES) has proven to be an effective and efficient diagnostic tool for patients with undiagnosed genetic diseases. This test often indicated when a patient has a suspected genetic diagnosis and where prior genetic testing has been uninformative, when there are multiple genes/diseases in the differential, or when the phenotype is not consistent with a known genetic disorder. DES has been instrumental in providing a diagnosis for patients among a broad range of indications. The five most common indications for testing in our cohort have included patients with neurological, musculoskeletal, craniofacial, ophthalmologic, and gastrointestinal phenotypes. In a retrospective analysis of the first 500 reported patients who underwent DES at one laboratory, we identified 125 patients who presented with syndromic gastrointestinal (GI) disease. Among these patients, 8 presented with GI symptoms as the main feature, while GI issues were among other indications for testing in the remaining patients. Some of the most common GI symptoms included irritable bowel syndrome (IBS), gastroesophageal reflux disease, constipation, diarrhea, intestinal hypomobility, abdominal distension, eosinophilic esophagitis or colitis, omphalocele and colon/rectal lesions. A positive or likely positive findings was identified in 35/125 patients, providing a 28% diagnostic rate for syndromic gastrointestinal indications. 30 molecular diagnoses were established within well-characterized genes, while 5 cases identified pathogenic mutations in novel genes. These novel genes included: *IL21R, MYH10, CACBA1E, ITSN1*, and *PURA*. The positive findings among characterized genes included the following genes: *TRPS1, ANKRD11* (2 cases), *TUBB4A* (2 cases), *ALG1, PTCHD1, SHANK3, STXBP1, ACTG2* (2 cases), *ARHGEF9, SCN1A, KAT6B, PKD1, FOXG1, SYNGAP1, NDUFA1, PCNT, SMC3, TMEM231, OFD1, ANO3, COL3A1, AMPD2, GARS, COL4A1, HMBS, and NGLY1.* Although GI disease is not the most common indication for DES, these findings highlight the clinical utility of DES for patients with potentially syndromic disease involving the GI tract and GI disease alone.

2581T

Assessment of a next generation sequencing panel to detect mutations in 40 genes causing renal tubulopathies. *E. Ashton[1], D. Bockenhauer[2], W. van't Hoff[2], R. Kleta[2], N. Lench[1].* 1) Regional Genetics, Great Ormond Street Hospital NHS Foundation Trust, London, United Kingdom; 2) Department of Nephrology, Great Ormond Street Hospital NHS Foundation Trust, London, United Kingdom.

The renal tubule regulates blood pressure, as well as water, acid-base and electrolyte homeostasis and there are currently mutations described in around 40 different genes associated with these disorders. We describe here evaluation of a 40 gene panel (including autosomal dominant, autosomal recessive and X-linked genes) that was developed within the EURenOmics consortium (www. Eurenomics.eu) in 92 patients by next generation sequencing. The Multiplicom MASTR ™ technology (Multiplex Amplification of Specific Targets and Resequencing) was used for library preparation. Nine multiplex PCR reactions were performed for each patient followed by pooling of products to produce amplicons for all regions of interest in the 40 genes in the panel. This was followed by a second round of PCR to add patient-specific barcodes and adaptors for sequencing. Sequencing was carried out using the Illumina MiSeq system using the version 3 2 x 300 bp sequencing chemistry (generating between 13.2 and 15 Gb of data per sequencing run). Data analysis was performed with an in-house bioinformatics pipeline using BWA for alignment, VarScan for variant calling and Variant Effect Predictor (VEP) for variant annotation (Ensembl). All mutations identified by next generation sequencing were confirmed by Sanger sequencing. Included in the patient cohort were 10 positive controls with previously known mutations in *AVPR2* (MIM 300538), *CLCN5* (MIM 300008), *CLDN19* (MIM 610036), *HSD11B2* (MIM 614232), *KCNJ1* (MIM 600359), *KCNJ10* (MIM 602208), *OCRL* (MIM 300535), *SLC12A1* (MIM 600839) and *SLC4A1* (MIM 109270) to assess the sensitivity of the methodology.

We have simultaneously screened a group of patients diagnosed with renal tubulopathies for mutations in 40 different genes using a Multiplicom MASTR ™ panel. Most of these genes are currently not available for testing in UK diagnostic laboratories. The testing described here allows screening of multiple genes simultaneously to provide a faster diagnosis for patients at reduced cost, with mutation-specific testing then available to other family members. The flexibility of the Multiplicom MASTR™ method means that any other genes identified to be causative of renal tubulopathy could be included in the design in future.

2582S

Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *T. Zemojtel[1,2], S. Köhler[1], L. Mackenroth[1], M. Jäger[1], J. Hecht[1], P. Krawitz[1], L. Graul-Neumann[1], S. Doelken[1], N. Emke[1], M. Spielmann[1], N.C. Oeien[1,5], M.R. Schweiger[1], U. Krüger[1], G. Frommer[6], B. Fischer[1], U. Kornak[1], A. Ardeshirdavani[7], Y. Moreau[7], S.E. Lewis[8], M. Haendel[9], D. Smedley[10], D. Horn[1], S. Mundlos[1,3,4], P.N. Robinson[1,3,4].* 1) Insitue for Medical Genetics and Human Genetics, University Clinics Charité, Berlin, Berlin, Germany; 2) Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland; 3) Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany; 4) Berlin Brandenburg Center for Regenerative Therapies (BCRT), Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany; 5) Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13125 Berlin, Germany; 6) Agilent Technologies, Hewlett-Packard-Straβe 8, 76337 Waldbronn, Germany; 7) Department of Electrical Engineering, STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven, Belgium; 8) Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA; 9) University Library and Department of Medical Informatics and Epidemiology, Oregon Health & Sciences University, Portland, OR, USA; 10) Mouse Informatics group, Wellcome Trust Sanger Institute, Hinxton, UK.

Less than half of patients presenting with suspected genetic disease receive an etiological diagnosis. Though next-generation sequencing (NGS), and in particular whole exome sequencing (WES), has enabled an unprecedented acceleration in the pace of Mendelian disease gene discovery, extensive challenges remain to integrate NGS, bioinformatics, and clinical data into effective diagnostic workflows. Because only variants in the ~2800 established Mendelian disease genes can be interpreted in a clinical context, we have established an approach that targets variants in these causative genes, which we will refer to as the disease-associated genome (DAG), developing a targeted enrichment DAG panel (7.1 Mb), for which we achieved a coverage of 20-fold or better for 98% of bases. Furthermore, we established a computational method termed Phenotypic Interpretation of eXomes (PhenIX) that evaluates and ranks variants based on variant pathogenicity and semantic similarity of patients' phenotype profiles described by Human Phenotype Ontology (HPO) terms to those of 3991 Mendelian diseases associated with the ~2800 disease genes. In computer simulations, ranking genes based only on the variant score put the true gene in first place less than 5% of the time; in contrast PhenIX placed the correct gene in first place over 86% of the time. We retrospectively tested PhenIX on 52 patients with previously identified mutations and known diagnoses, achieving a mean rank of 2.1 for the correct gene. In a prospective study on a group of 80 individuals who had remained without a diagnosis despite intensive clinical workup and targeted genetic testing, PhenIX analysis enabled a diagnosis in 23 cases (~29% at a mean rank of ~2.4). An experienced clinical geneticist needed an average of ~1.5 hours per patient to evaluate the ranked list produced by PhenIX. Thus, the combination of targeted NGS investigation of the DAG followed by phenotype-driven bioinformatic analysis provides the means for quick and effective differential diagnostics in medical genetics.

2583M

Genomic diagnosis in children with developmental delay/intellectual disability. *K. Bowling[1], K. East[1], M. Amaral[1], C. Penfold[1], J. Whittle[1], C. Henegar[1], S. Simmons[2,3], E.M. Bebin[2,3], E.J. Lose[2,3], K. Brothers[4], G.S. Barsh[1], R.M. Myers[1], G.M. Cooper[1].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) North Alabama Children's Specialist, Huntsville, AL; 3) University of Alabama at Birmingham, Birmingham, AL; 4) University of Louisville, Louisville, KY.

Developmental delay, intellectual disability, and related congenital defects (DD/ID) affect ~1-2% of children who are born worldwide. DD/ID phenotypes greatly impact affected individuals and their families, and constitute a costly public health burden. Genetic diagnosis of DD/ID-affected children comprises an ideal exploratory arena to address the challenges and unanswered questions facing the widespread use of DNA sequencing in clinical contexts. While cytogenetic and microarray-based platforms have been used extensively for DD/ID diagnostics, causal genetic variation remains undetermined in a large proportion of DD/ID-affected individuals. Therefore, as part of the Clinical Sequencing Exploratory Research Consortium, we are conducting whole exome sequencing in ~450 parent-offspring trios over a period of 4 years to establish a genetic diagnosis for unexplained cases of DD/ID. Primary genetic results (DD/ID causative) as well as secondary findings (pathogenic variation detected in parents) are being returned to probands and their parents in a clinic setting, and follow-up studies are underway to determine how genomic test results affect the well-being of participants. To date, we have enrolled and consented 99 parent-offspring trios (some with DD/ID-affected siblings), and have completed exome sequencing and analysis for 29 of these trios. Among children with no diagnosis after standard clinical testing (e.g., clinical microarrays), we have successfully identified DD/ID pathogenic variants in 6 probands (21% of our total) including mutations in *TCF4*, *SCN2A*, *MECP2*, *ARX* and *GRIN2B*. We have also found variants, that while not definitively causal, are plausible candidates; for example, compound heterozygosity for impactful variants in *PCNXL2* shared by affected siblings, and de novo variants in *EBF3*, a conserved transcription factor known to be important in brain development. Further, we have identified potentially pathogenic variation in 15 of the 58 exome-sequenced parents (26%). Secondary findings include variation in *RYR1* (malignant hyperthermia susceptibility) and *DSP* (arrhythmogenic right ventricular cardiomyopathy), as well as carrier status for *HEXA* (Tay-Sachs disease) and *CFTR* (cystic fibrosis). The results of our clinical sequencing project demonstrate, thus far, that large-scale DNA sequencing improves the rate of definitive diagnoses in DD/ID-affected children and thereby meets a valuable clinical need.

2584T

Autosomal Dominant Hypertrophic Cardiomyopathy (HCM) is an important modifier of the cardiomyopathy of Fabry Disease (FD): Implications for α-Galactosidase A replacement therapy. *D. Doheny[1], R. Desnick[1], D. Macaya[2], C. Antolik[2], H. Rehm[3], A. Alfares[3], T. Callis[4], J. DaRe[4], S. Garman[5].* 1) Dept Gen & Genomic Scienc, Icahn School of Medicine at Mount Sinai, New York, NY; 2) GeneDx, 207 Perry Pkwy, Gaithersburg, MD; 3) Harvard Medical School, Laboratory for Molecular Medicine, 65 Lansdowne Street, Cambridge, MA; 4) Transgenomic, 5 Science Park, New Haven, CT; 5) Department of Biochemistry and Molecular Biology, University of Massachusetts, 710 North Pleasant street, Amhurst, MA.

Hypertrophic Cardiomyopathy (HCM) is a common autosomal dominant disorder of the myocardium that affects ~1 in 500 individuals world-wide and is the most common cause of sudden cardiac death in individuals under 35 years. To date, variants in >20 genes have been identified that cause HCM, most encoding sarcomere proteins. The cardiac features of HCM are indistinguishable from the cardiomyopathy of FD due to α-galactosidase gene (GLA) mutations. Screening panels that sequence the most common HCM-causing genes also include GLA, detecting pathogenic (P) and likely pathogenic (LP) variants. We hypothesized that previously unrecognized FD patients, particularly those with the "later-onset" phenotype, would be detected by these panels. In 6486 HCM patients referred for panel sequencing, GLA variants were identified and stratified by sex and phenotype, including "classic" or "later-onset" phenotypes, benign variants (BV), and variants of unknown significance (VUS). The presence of HCM variants was also recorded. Of the 6486 HCM patients tested, 96 (1.5%) were identified with a GLA variant. Of these 96, 27% also had a P or LP HCM variant, most being in the β-myosin heavy chain (MYH7) and the myosin binding protein C (MYBPC3) genes. When stratified by Fabry phenotype and sex: 1) 13 patients had the "classic" phenotype, but only one, a heterozygote, had a P/LP HCM variant, indicating that the cardiomyopathy in most was due to FD; 2) among the other Fabry phenotypes, a higher percentage of patients had a P/LP HCM variant suggesting that the cardiomyopathy was due primarily to HCM variants: among 32 patients with the "later-onset" phenotype, 25% had a HCM variant; among 43 with GLA BVs, 35% had a HCM variant; and among 8 with GLA VUSs, 25% (both females) had a HCM variant. Of those with GLA BVs, 79% had the pseudodeficient D313Y lesion of which 38% had a HCM variant. These findings are the first to indicate that variants in other genes may modify or underlie the pathogenesis of the major life-threatening manifestations of FD. Clearly, all patients with FD and left ventricular hypertrophy or HCM should be tested for HCM-causing genes, and those with HCM found to have GLA mutations: 1) should be evaluated by a FD specialist, 2) at-risk family members should be tested, and 3) affected males considered for early intervention by enzyme replacement therapy. Additionally, identification of P/LP HCM variants in FD patients may also have implications for treatment options.

2585S

Characterization of Malaysian children with Beckwith-Wiedemann syndrome and Silver-Russell syndrome using methylation specific - multiplex ligation-dependent probe amplification. *M. Thong[1], K. Thurga[2], R.Y.Y. Poh[2], I. Taufik[1], H.B. Chew[3], G.S. Ch'ng[3], W.T. Keng[3].* 1) Department of Paediatrics, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia; 2) Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia; 3) Department of Genetic, Hospital Kuala Lumpur, 50586 Kuala Lumpur, Malaysia.

Introduction: Beckwith-Wiedemann syndrome (BWS [MIM 130650]) and Silver-Russell syndrome (SRS [MIM 180860]) cause gigantism and growth retardation, respectively. Imprinting centre (IC) 1 controls the expression of IGF2 and H19 whereas IC2 controls the expression of CDKN1C and KCNQ1OT1. Methodology: Blood samples were collected from 13 BWS and 18 SRS patients based on clinical features. Genomic DNA was extracted. The DNA of patients and controls were subjected to multiplex ligation-dependent probe amplification (MLPA) to detect the copy number and methylation status in imprinted genes of chromosome 11 for BWS and SRS. The resulting PCR products were sent for fragment analyses. Results and discussion: Out of 13 BWS patients, three patients had IC2 hypomethylation with normal IC1 methylation. The IC2 hypomethylation may be implicated in KCNQ1OT1 activation and CDKN1C inhibition, leading to the gigantism. Six patients had IC2 hypomethylation and IC1 hypermethylation simultaneously which indicated paternal UPD at chromosome 11p15.5. One patient had both IC1 and IC2 hypomethylation and two patients had IC1 hypermethylation with normal IC2. One mutation was identified in CDKN1C at position 78 in one BWS patient. One had normal methylation at both IC's. Among the 18 SRS patients, five showed normal methylation at both IC's. One showed hypomethylation at both IC1 and IC2, and two showed IC1 hypomethylation with IC2 hypermethylation. Hypomethylation of IC1 causes the biallelic expression of H19 and biallelic silencing of IGF2, resulting in growth restriction. However, one had normal IC1 and IC2 hypermethylation and one had normal IC1 with IC2 hypomethylation. Three had hypermethylation at both IC1 and IC2 and five had IC1 hypermethylation with normal IC2. None of our SRS samples showed UPD in chromosome 7. Conclusion: The patients in this study presented with gain and loss of methylation, duplication and deletion in both BWS and in SRS, in different imprinted genes, emphasizing the heterogeneity of these two growth disorderd syndromes.

2586M

An age-based categorical framework to guide informed decision-making about next generation sequencing results in newborn screening. *L.V. Milko, J. O'Daniel, K. Foreman, C. Turcott, J. Booker, L. Boshe, M. Gucsavas-Calikoglu, A. Aylsworth, J. Muenzer, D. Frazier, B. Powell, M. Roche, N. Strande, N. Vora, C. Powell, J.S. Berg.* University of North Carolina Chapel Hill, School of Medicine, Chapel Hill, NC.

A tenet of newborn screening (NBS) is the potential for early detection and prevention of diseases prior to clinical manifestation of symptoms. Although next generation sequencing (NGS) offers great promise in increasing the number of conditions that can be presymptomatically screened, it also presents significant ethical complexity for conditions where efficacious treatment may not be available, or where avoiding the "diagnostic odyssey" might be the only benefit. The NC NEXUS (North Carolina Newborn Exome Sequencing for Universal Screening) project is developing a modified "binning" classification framework to characterize different categories of information that parents might be interested in learning, and to study the psychosocial impacts of their choices. The framework is divided into four quadrants that balance a semi-quantitative scale of "clinical actionability" (a measure of the likelihood and severity of disease outcomes, and the efficacy and potential harms of interventions), against the age of onset of disease or age of implementation of preventative intervention. The core panel ("NGS-NBS") is composed of clinically actionable genes that are implicated in childhood-onset disorders, including conditions on standard newborn screens. All parents of screened newborns will receive results of the core NGS-NBS panel. Given the lack of empirical data on how to present other types of genetic information, parents will be randomized into either a control arm receiving only results of the NGS-NBS panel, or an experimental arm tasked with deciding whether to learn additional information from three categories: 1. clinically actionable conditions with typical disease onset or interventions beginning in adulthood (eg. Lynch syndrome); 2. childhood-onset, non-medically actionable disorders with no known effective interventions (eg. Rett syndrome); and 3. carrier status for recessive disorders. Information about adult-onset disorders with no known interventions (eg. Alzheimer disease) will not be returned. We hypothesize that classifying the types of results that parents may opt to learn into clear categories will assist physicians and parents in engaging in informed decision-making in the complex setting of genome-scale sequencing of healthy newborns. It will also be useful in understanding how and why parents choose whether to participate in sequencing of their child's DNA, what results they elect to receive, and the impacts of their decisions.

2587T

Targeted resequencing in intellectual disability and epilepsy in routine diagnosis, preliminary results. *D. Lederer, V. Benoit, Ch. Verellen-Dumoulin, I. Maystadt.* Centre for Human Genetics, IPG, Gosselies, Belgium.

Epilepsy is a common neurological disorder with a lifetime incidence rate of 3%. Intellectual deficiency affects 1-3% of children. Microarray CGH detects a chromosomal abnormality in 10-15% of individuals with intellectual disability and in 8% of patients with epileptic encephalopathy. Extensive metabolic study and brain imaging techniques lead to a diagnositc in respectively 5% and 0,8% of cases. There are many conditions that associates epilepsy and mental retardation, caused by mutations in more than 100 genes, making genetic diagnosis expensive and challenging in the absence of other congenital anomalies. With the recent advances in genetics, it is becoming easier to analyse panel of genes responsible for a condition. In order to simplify the genetic diagnosis in epilepsy and intellectual deficiency, we have gathered in one analysis 150 genes. The coding regions are amplified with a customed Ampliseq protocol and we sequence the fragments on an Ion Proton®. Mutation calling is done with Life technologies variant caller and Cartagenia. Here, we present the preliminary results. To date, we have tested more than 120 patients. The inclusion criteria to this study are intellectuel deficiency or developmental delay and epilepsy. We have confirmed mutations in few genes: STXBP1, FOXP1, KCNQ2, SCN2A, RFT1, ATRX, GLDC, NRXN1, NLGN3, SHANK3, SYNGAP1, MAGI2, SLC9A6, KDM5C. Overall, we are expecting a causal mutation rate of 15-25%. We will present the clinical and metabolic data of patients with causal mutations. Moreover, other substitutions detected were already described in the litterature in patients with specific epilepsy with or without intellectual disabilty and were sometimes inherited from a healthy parent (SCN1A, KCNQ3, SCN8A, SCN1B). They could act as modifier genes inducing epilepsy in patients affected by a second mutation responsible for the intellectual phenotype. We need more data to confirm this hypothesis. In conclusion, gene panel analysis could be a first line diagnostic test in patients with intellectual deficiency and epilepsy in view of low cost and high mutation pick up rate.

2588S

Should the ACMG expand the required reportable disorders or findings on Exome Sequencing? Reporting a recent experience. *R.M. Zambrano, Y. Lacassie.* Department of Pediatrics, LSUHSC, New Orleans, LA. and Children's Hospital of New Orleans.

Exome and Genome sequencing are becoming the standard in the Genetic clinics for the evaluation of patients with rare disorders. These technologies are allowing the identification of the molecular bases of many syndromes and also uncover incidental or secondary findings that could be of medical value for the patient and their family. The ACMG has published a statement listing the recommendations for reporting incidental findings in Clinical Exome and Genome Sequencing. Based in a recent experience in which the use of incidental clinical Exome data supported the clinical diagnosis of severe primary hyperparathyroidism in a neonate, we propose to include disorders that can cause severe neonatal diseases and also recommend that the laboratories performing these tests do report all the recessive genes in which both parents are heterozygotes. We report a 5-day-old Caucasian female born to a 29-year-old father and a 32-year-old consanguineous G5P2 mother. The 2nd and 3rd pregnancies were miscarriages. Their 4th pregnancy resulted in a child presenting congenital scoliosis and mild developmental delay. Because of consanguinity, Exome sequencing was requested. It didn't find an expected autosomal recessive disorder but identified a previously unreported de novo mutation in the TRPS1 gene (AJMG 2014, in press). Their last pregnancy was detected early. Only complications included polyhydramnios and premature labor requiring bed rest for 8 weeks. The baby was born NSVD at 374/7 WGA. The newborn was discharged home at DOL 2 but had to be re-admitted after 2 days due to hyperbilirubinemia, poor feeding, weight loss, hypotonia and lethargy. Laboratory work-up showed severe hypercalcemia with a serum calcium level of 33.7 mg/dL. Repeated was 36.8 mg/dL and PTH was elevated at 867 pg/mL (nl 15-65 pg/mL). We suspected neonatal severe hyperparathyroidism (NSHPT) caused by mutations in the CASR gene. As the parents had been sequenced, we contacted the laboratory which quickly confirmed that both parents were heterozygous for a mutation in the CASR gene: c.206G>A (p.R96H). This allowed us to further support our clinical diagnosis of NSHPT. Sanger sequencing confirmation on the proband is pending. Reporting biparental heterozygosity for genes causing severe neonatal disease or disorders with available treatment is recommended. This would allow genetic counseling, prenatal testing, and the prompt therapy after birth or later.

2589M

Exploration of the benefit of risk-stratified colorectal cancer screening based on common genetic variants - current status and future potential. *S.K. Naber[1], S. Kundu[2], K.M. Kuntz[3], Z.K. Stadler[4], R.W. Burt[5], M.S. Williams[6,7], N. Calonge[8], D.T. Zallen[9], T.G. Ganiats[10], E.P. Whitlock[11], E. Webber[11], K.A.B. Goddard[11], N.B. Henrikson[12], M. van Ballegooijen[1], A.C.J.W. Janssens[2], A.G. Zauber[4], I. Lansdorp-Vogelaar[1].* 1) Department of Public Health, Erasmus Medical Center Rotterdam, Rotterdam, Netherlands; 2) Department of Epidemiology, Emory University, Atlanta, GA, United States; 3) Department of Health Policy & Management, University of Minnesota, Minneapolis, MN, United States; 4) Memorial Sloan-Kettering Cancer Center, New York, NY, United States; 5) Department of Medicine, University of Utah, Salt Lake City, UT, United States; 6) Intermountain Healthcare, Salt Lake City, UT, United States; 7) Genomic Medicine Institute, Geisinger Health System, Danville, PA, United States; 8) The Colorado Trust, Denver, CO, United States; 9) Department of Science and Technology in Society, VirginiaTech, Blacksburg, VA, United States; 10) Department of Family and Preventive Medicine, University of California, San Diego, CA, United States; 11) Center for Health Research, Kaiser Permanente Northwest, Portland, OR, United States; 12) Group Health Research Institute, Seattle, WA, United States.

Background. Common genetic variants contribute to colorectal cancer (CRC) risk and can be used to stratify the population into CRC risk categories. However, the discriminatory performance of such risk-stratification algorithms is currently limited. In this study, we investigate the current and potential future benefits of using risk-stratified colonoscopy screening, based on common genetic variants, versus uniform colonoscopy screening at ages 50, 60 and 70. **Methods.** We used the MISCAN-Colon microsimulation model to determine cost-effective colonoscopy screening strategies for people with a relative risk (RR) for CRC of 0.1, 0.2, ..., 9.8, 9.9 and 10. The costs and effects of risk-stratified screening in the population were determined based on the current discriminatory performance of common genetic variants (area under the ROC curve (AUC) of approximately 0.6) compared to uniform screening at ages 50, 60 and 70 for all. Because it is expected that the discriminatory performance of risk-stratification based on common genetic variants will increase in the future, we also estimated costs and effects for risk-stratified screening based on hypothetical common genetic variants with higher levels of discriminatory performance (AUC of 0.65, 0.70, ..., 0.90). **Results.** With current discriminatory performance, the optimal colonoscopy screening strategy ranged from no screening for people with a RR of 0.1 to screening every 3 years from age 40 until age 85 for people with a RR of 5.4 - 10. Screening at ages 50, 60 and 70 was optimal for people with a RR between 0.9 and 1.3. This stratification resulted in 1% more life years gained than uniform screening (less than 1 life year per 1,000 individuals) for the same overall costs. With increased discriminatory performance, the gain in life years increased from almost 4% for an AUC level of 0.65 to more than 18% for an AUC level 0.90. **Conclusions.** Given the very modest discriminatory performance of common genetic variants in risk-stratification for CRC, the current benefits of risk-stratified CRC screening based on these variants are limited. New variant discoveries are needed to yield a substantial improvement in discriminatory performance, and are necessary for risk-stratified screening to become clinically significant.

2590T

HCV infection and interferon-based treatment induce p53 gene transcription in chronic hepatitis C patients. *J. Nowak[1], B. Swiatek-Koscielna[1], E. Kaluzna[1], J. Rembowska[1], I. Mozer-Lisewska[2], I. Bereszynska[2], J. Wysocka-Leszczynska[1], A. Kowala-Piaskowska[2], J. Wysocki[1], D. Januszkiewicz-Lewandowska[1, 2, 3].* 1) Inst Human Gen, Polish Academy Sci, Poznan, Poland; 2) University of Medical Sciences, Poznan, Poland; 3) Department of Medical Diagnostics, Poznan, Poland.

It is suggested that tumor suppressor p53 gene is implicated in the interferon (IFN)-mediated innate immunity against viruses. This study aims at examining transcriptional response of p53 gene to HCV infection and IFN-based therapy in chronic hepatitis C (CHC) patients. p53 gene expression was quantified by TaqMan real-time PCR in peripheral blood mononuclear cells (PBMCs) of 63 CHC patients (HCV genotype 1 and 4) treated with pegylated IFN-α and ribavirin and of 51 healthy individuals. Analyses were performed before, at week 4 and 12 of treatment. p53 gene expression was significantly upregulated in CHC patients before and at week 4 of anti-HCV therapy. No significant differences in p53 mRNA expression between rapid virological responders (n=23), complete early virological responders (n=17) and non-responders (n=23) were found. Moreover, no significant correlation between p53 gene expression and HCV-RNA viral load was observed. The results obtained indicate that HCV infection and IFN-based treatment induce p53 gene transcription in PBMCs. It can be suggested that p53 gene may play a role in HCV infection but is not directly involved in treatment-induced HCV elimination. Moreover, variations in p53 gene expression not determine on-treatment anti-HCV response in CHC patients with 'difficult-to-treat' HCV genotype 1 and 4.

2591S

Assisted reproductive treatment is not a risk factor for chromosomal abnormalities in spontaneous abortion. *s. Shim.* Dept of Ob and Gyn, CHA Gangnam Medical Center, seoul, South Korea, MD.

Background: The causes of miscarriage include fetal and maternal factors. More than half first-trimester pregnancy losses are attributed to fetal chromosomal abnormalities, in which autosomal trisomy is known to be the most frequent cause (about 60%). Maternal risk factors for first-trimester pregnancy losses are maternal obesity, old age, and prior history of pregnancy loss. Recently, ART is more frequently utilized with an increase in infertility rate. However, there is a paucity of data about the frequency of embryonic chromosomal abnormalities in first-trimester miscarriage after assisted reproductive treatment (ART). We examined whether or not embryonic chromosome aberration in first-trimester pregnancy losses is more frequent in pregnant women with ART than in those with natural pregnancy. Methods: We conducted a retrospective study in CHA Medical Center. Study population consisted of 511 pregnant women (natural pregnancy group 31.31% <160/511> vs. ART pregnancy group 68.49% <350/511>) who underwent either spontaneous abortion or surgical dilatation and curettage between 2011 and 2012 with the diagnosis of missed abortion during first trimester (gestational age < 14wks). We compared the frequency of embryonic chromosome aberration according to natural pregnancy group or ART pregnancy group, BMI >= 25 or BMI < 25, the presence or absence of prior history of pregnancy loss, and maternal age >= 35 or < 35. Results: The overall rate of aneuploidy was 51.6%, and mean maternal age 35.0 years. There was no significant difference in abnormal karyotypes between ART pregnant group and natural pregnancy group (P=0.154). However, abnormal karyotypes were significantly more frequent in patients with maternal age >= 35 years than those with maternal age < 35 years, in patients with BMI >= 25 than in those with BMI < 25, and in pregnant women with prior history of pregnancy loss than in those without prior history of pregnancy loss (each for P<.05). Conclusions: ART treatment is not a risk factor for chromosomal abnormalities in the first trimester miscarriage. However, maternal age more than 35 years, maternal obesity (BMI >= 25) and prior history of pregnancy loss are associated with abnormal karyotyping in first trimester miscarriage.

2592M

A workflow based information system infrastructure to support translational science: The NIH Undiagnosed Diseases Program experience. *A.E. Links[1], D. Draper[1], E. Lee[1], V. Lebedev[2], M. Didenko[2], D. Adams[1], M. Brudno[3], S. Dumitriu[3], M. Girdea[3], W.P. Bone[1], B. Coessens[4], S. Verhoeven[4], C.F. Boerkoel[1], W.A. Gahl[1], M. Sincan[1].* 1) Undiagnosed Diseases Program, National Human Genome Research Institute, Bethesda, MD; 2) RURO Inc., Frederick, MD; 3) Department of Computer Science, University of Toronto, Ontario, Canada; 4) Cartagenia NV, Leuven, Belgium.

The Undiagnosed Diseases Program (UDP) is a translational medicine initiative focused on bridging the gap between the patient bedside and the research bench. The application of translational medicine presents two major challenges: “translational research,” or bench-to-bedside integration, and conduction of complex and multidisciplinary research through collaborative research networks. Moreover, the UDP deals with many independent cases that each have individual research and clinical life cycles. In an effort to overcome these two challenges and manage the complexities of the UDP, clinical and research processes were mapped and used for development of a Translational Research Information System (TRIS) known as the Undiagnosed Diseases Program Integrated Collaboration Server (UDPICS). The UDPICS ecosystem includes a workflow based translational research information management server, an ontology based phenotyping application, a biosample management system, an animal tracking system, a cloud-based genomic analysis platform, and an electronic laboratory notebook. It contains a complicated data structure with over 80 active workflows and over 2000 discrete data elements for efficient tracking of research projects and clinical services. In addition, the UDP has designed a collaborator instance of UDPICS to allow external collaborators to work with non-personally identifiable information (non-PII) specific to the individuals involved in the collaboration, upload reports for easier feedback to the clinical staff, and engage in recorded discussions with members of the UDP team. In conclusion, the UDP has developed a system that is capable of supporting both translational research and clinical services while providing a collaboration platform to allow the research consortia to tackle the complex and diverse scientific problems posed by the many different undiagnosed diseases.

2593T

Assessment of the variant annotation interpretive gap among major variant databases. *M. Lee, B. Vecchio-Pagan, G.R. Cutting.* McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD.

The interpretive gap is a growing divide between variants which have been identified and those which have been annotated. This gap is a significant obstacle when seeking to assign disease liability to variants detected by exome sequencing and when interpreting sequencing results in diagnostic laboratories. The breadth of the interpretive gap may be assessed by evaluating the consistency and depth of data across traditional disease-focused as well as "normal" variation databases. We surveyed five databases: OMIM (curates disease phenotypes), HGMD (curates disease variants from the literature), ClinVar (inventories variants from user submissions), LOVD (inventories variants from user submissions), and The 1000 Genomes Project (inventories variants detected in "normal" individuals). There were 2780 genes curated in OMIM, 4733 in HGMD, 18645 in ClinVar, 20172 in LOVD, and 17406 in 1000 Genomes. Across the inventory databases (ClinVar, LOVD, and 1000 Genomes), 67.7% of genes were present in all three; however, only 10.4% of genes were found in all five. To assess variant depth, 2179 genes annotated in all five databases were extracted. Variant counts reported for each gene differed by as much as an order of magnitude per gene. To determine consensus and quality of variant annotation, we selected the 30 disease-implicated genes with the highest number of variants in OMIM and extracted variant data for each gene from OMIM, HGMD, ClinVar, and 1000 Genomes. MAF data from dbSNP and 1000 Genomes revealed that 60.0 ± 39.2% (range: 17.1-99.1%) of variants reported in the 30 genes had a dbSNP rsID, and 26.8 ± 45.9% (range: 0.114-95.0%) had population MAF, indicating that high quality feature data is missing for a substantial fraction of variants in genes with strong disease associations. Remarkably, only 41 (0.437%) of 9383 unique variants are annotated by all four databases. Furthermore, variant type (e.g. SNV, deletion, insertion, indel) distributions of the 30 genes in HGMD, ClinVar, and 1000 Genomes were strikingly different among the curated variant types. A suspected overrepresentation of SNVs was seen, and represented 72.6% of variants in HGMD, 82.6% in ClinVar, and 98.5% of variants in 1000 Genomes. These findings reveal that the interpretive gap is both wide and multidimensional, as there is an absence of consistency regarding the number, composition, and associated feature data of variants annotated in widely used variant databases.

2594S

Clinical whole exome sequencing (WES) production update at Baylor Whole Genome Laboratory (WGL): Improved procedures for faster TAT and better disease gene coverage. *Y. Ding[1], S. Matakis[1], C.J. Buhay[1], M. Wang[1], N. Veeraraghavan[1], T. Chiang[1], A.C. Hawes[1], W. Liu[1], N. Saada[1], J. Ma[1], L.K. Dolores-Freiberg[2], J. Chandarana[2], C.J. Qu[2], R. Najjar[2], M.N. Bainbridge[1], Y. Han[1], H. Dinh[1], J.V. Korchina[1], Q. Wang[1], E. Boerwinkle[3], J.R. Lupski[2], S.E. Plon[2], A.L. Beaudet[2], C.M. Eng[2], R.A. Gibbs[1], D.M. Muzny[1], Y. Yang[2].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.

Baylor WGL has been providing clinical WES services since October 2011. To date, approximately 3800 WES cases including 3650 germline WES, 150 cancer WES as well as ~3100 mitochondrial genomes have been sequenced on the Illumina HiSeq2500 or HiSeq2000. The current sample volume is approximately 200 samples per month. NGS sequencing of mitochondrial genomes has been included in the clinical WES test starting with samples received on 10/15/2012. The mitochondrial genome was amplified by a single long-range PCR before subjecting to paired-end library constructions. Mitochondrial and WES libraries from the same sample were barcoded and sequenced in the same lane (1:30) on Illumina's HiSeq platform. This process eliminates virtually all sequencing cost for the mitochondrial genome. Paired-End library constructions for WES and mitochondrial genome were automated using Beckman Biomek NXP/Span-8 robots, and show high throughput (up to 400 samples per month), data quality and reproducibility. The enrichment for exome sequences utilized a solution based whole exome capture probe set, VCRome 2.1, which covers approximately 200,000 coding exons (35 Mb). An average of >10 Gb data were generated for each WES and 95% of the targeted exome regions were sequenced at a depth of >20X (average coverage >140X). The mitochondrial genome was sequenced at >20,000X to facilitate the detection of large deletions. Further, we have validated and implemented additional procedures initially developed at the Baylor Human Genome Sequencing Center (HGSC) in order to achieve better sequencing coverage and shorter turnaround time (TAT) (Muzny et al. 2014 ASHG abstract). First, by applying the "spike-in" reagents (version 1) targeted at 1800 GeneTests genes, the number of 100% covered genes increased by 200-300 for samples with 10 Gb output. Starting from April 2014, all clinical samples (450 so far) are captured using the "spike-in" exome reagents. Second, the newly implemented rapid hybridization procedure reduces the probe capture time from 72 to 24 hours. Further validation of additional quick WES (QWES) procedures developed at HGSC will enable the completion of the wet lab exome sequencing process within 5 working days at WGL. Improved clinical WES protocols and automation have enabled Baylor WGL to produce higher quality WES data with better disease gene coverage and shorter TAT and prepared the lab for more challenging services such as WES for prenatal or neonatal ICU samples.

2595M

Clinical actionability of incidental findings: application of a semiquantitative metric to assess actionability of over 1200 genes. *A.K.M. Foreman[1,2], J.K. Booker[1,3], L. Boshe[1,2], K.R. Crooks[1,3], J.P. Evans[1,2], B.C. Jensen[1,4], K. Lee[1,2], D.K. Nelson[1,5], J.M. O'Daniel[1,2], B.C. Powell[1,2], C.M. Powell[1,6], M.I. Roche[1,6], C. Skrzynia[1,2], N.T. Strande[1,2], K.E. Weck[1,3], K.C. Wilhelmsen[1,7,8], J.S. Berg[1,2].* 1) UNC Chapel Hill School of Medicine, Chapel Hill, NC; 2) Department of Genetics; 3) Department of Pathology and Laboratory Medicine; 4) Department of Medicine, Division of Cardiology; 5) Department of Social Medicine; 6) Department of Pediatrics, Division of Genetics and Metabolism; 7) Department of Neurology; 8) Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC.

As exome and genome sequencing are increasingly applied in clinical scenarios, incidental findings (IFs) are discovered as a matter of course. Debate about the most appropriate handling of such findings is ongoing, but it is widely recognized that IFs vary in their clinical utility or actionability. Evaluating actionability is critical to identifying genes for routine return and enabling informed decision-making by patients. The NCGENES (North Carolina Clinical Genomic Evaluation by NextGen Exome Sequencing) study returns "clinically actionable" IFs to patient-participants undergoing exome sequencing. To define these IFs for NCGENES, we developed a semiquantitative metric to assess actionability of gene-phenotype pairs by scoring five elements of actionability on a 0-3 scale: (1) **severity** and (2) **likelihood** of the adverse outcome, (3) **efficacy** and (4) **burdens or risks** of the proposed intervention, and (5) level of **knowledge**. Higher scores suggest greater actionability. We initially scored 161 genes selected from OMIM as plausible candidates for routine return as medically actionable IFs. The median score was 11, and 73 genes scored 12 or higher. To investigate the robustness of the metric, we scored a random sample of 1000 genes in RefSeq. The majority of genes scored 0 by default because an associated phenotype either did not exist in OMIM or OrphaNet, was caused by somatic mutation, or was simply a modest influence on disease risk based on association studies. The remaining 114 genes were formally scored using the metric, with a relatively flat distribution of scores from 5-10 and a median of 8. We then applied the metric to the 56 genes recommended by the ACMG for routine return when discovered as IFs. The median score was 11 (range 7-14). High-scoring genes included *MLH1* (with its associated phenotype, Lynch syndrome) with a score of 13, and *RYR1* (malignant hyperthermia) with a score of 12. The *NTRK1* gene, which was dropped from the ACMG list after being included on a preliminary version, scored 10. The ACMG 56 contains nineteen genes with a score ≤10. Our proposed metric promotes transparency in decisions about return of IFs, provides a basis for comparing genes with disparate phenotypes, and is customizable. We evaluated the impact of weighting individual elements, such as efficacy or knowledge, on the scores and discuss challenges to use of the metric, including pleiotropy, and factors important to return of IFs not captured by the metric.

2596T

Project of Iwate Tohoku Medical Megabank Organization toward preemptive medicine. *A. Shimizu[1], T. Hachiya[1,2], K. Tanno[3,4], A. Fukushima[5,6], Y. Shiwa[2], H. Ohmomo[1], R. Furukawa[2], K. Yamamoto[6], K. Ono[1], M. Satoh[1,2,3,7], J. Hitomi[8,9], K. Sobue[10,11].* 1) Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University; 2) Division of Biobank and Data Management, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University; 3) Division of Community Medical Supports and Health Record Informatics, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University; 4) Department of Hygiene and Preventive Medicine, School of Medicine, Iwate Medical University; 5) Division of Innovation and Education, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University; 6) Department of Clinical Genetics, School of Medicine, Iwate Medical University; 7) Division of Cardioangiology, Department of Internal Medicine and Memorial Heart Center, School of Medicine, Iwate Medical University; 8) Deputy Executive Director, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University; 9) Department of Anatomy, School of Medicine, Iwate Medical University; 10) Executive Director, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University; 11) Department of Neuroscience, Institute for Biomedical Sciences, Iwate Medical University.

The Tohoku Medical Megabank Organization was established building upon a partnership between two Japanese Universities: Tohoku University and Iwate Medical University as a national project to promote recovery from the Great East Japan Earthquake and the establishment of an advanced medical system. The organization has established a biobank that integrates the medical information of 150,000 participants and genome information collected from over 1,000 participants and is conducting a large-scale genome cohort study in order to develop and provide preemptive medicine; and at the same time, it also provides health checkups to the participants free of charge and dispatches doctors to the disaster-affected area. Iwate Medical University Tohoku Medical Megabank Organization is carrying out the following studies on its own: 1) technical examination of integration of several cohorts; 2) technical examination of transportation of the specimens and information stored in the cohorts and the biobanks; 3) technical examination of the use of information on polymorphism, DNA methylation and gene expression for predicting disease risks; 4) consideration of informing the participants of their disease risks estimated with the use of their genetic information; and 5) questionnaire surveys to study the impacts that this project has had on the health of the participants. We would like to integrate these studies in an organic manner, proceed with the technical examinations so that our genome cohort research will contribute to improving the health awareness of the participants, and eventually realize preemptive medicine in the future.

2597S

The ClinGen framework for defining the clinical validity of monogenic disease associations for use in research and clinical analyses. *J.S. Berg[1], O. Birsoy[2], A.J. Butte[3], M. Giovanni[4], K. Goddard[5], A. Hamosh[6], L. Milko[1], M.F. Murray[4], E. Riggs[4], P.N. Robinson[7], W.S. Rubinstein[8], A. Santani[9], A. Scott[6], T. Sneddon[3], N. Strande[1], H.L. Rehm[2], S.E. Plon[10], C.L. Martin[4].* 1) Genetics, UNC Chapel Hill, Chapel Hill, NC; 2) Laboratory for Molecular Medicine, Partners Healthcare Center for Personalized Genetic Medicine, Boston, MA; 3) Genetics, Stanford University, Stanford, CA; 4) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 5) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 6) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 7) Institute for Medical Genetics, Charité Berlin, Germany; 8) National Center for Biotechnology Information, Bethesda, MD; 9) Children's Hospital of Philadelphia, Philadelphia, PA; 10) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

The rapid pace at which new gene-disease associations are reported for monogenic disorders poses a tremendous challenge to the clinical interpretation of genome-scale sequencing data in patients and research participants. The NIH-funded ClinGen consortium is working with the NCBI to generate a publicly available resource for delineation of clinically relevant genes and variants. As part of this effort, a workgroup was assembled to develop standardized procedures for curating genes and their relationship to monogenic disorders, in order to provide well curated, up-to-date information for use in research and clinical analyses. The workgroup first addressed the problem of categorizing the clinical validity of gene-disease associations according to strength of evidence.

The workgroup established seven categorical designations for strength of evidence for a causal role in disease: Definitive, Strong, Moderate, Limited, No Reported Evidence, Disputed, and Evidence Against. These categories are defined according to the type and strength of evidence provided in the published literature or other public sources, the presence or absence of contradictory evidence, and whether the initial report of disease-gene causality has been replicated.

This framework is initially being applied to genes relevant to three clinical domains: hereditary cancer, cardiology, and inborn errors of metabolism. We are actively curating target genes within these areas and evaluating results of the categorization, consistency between reviewers and frequency of category change. Our overall goal is to provide the categorization and supporting evidence for every gene implicated in monogenic disorders, and to develop a standardized system for community curation in order to support updated gene-disease relationships and associated evidence as new information is published. This framework will be useful to researchers investigating human genetic disorders, so that they can apply consistent criteria when claiming evidence of disease association, and to highlight those claims of disease association that require more data. Categorization of the clinical validity of gene-disease associations will also be of substantial clinical benefit given the increasing use of large multi-gene panels and genome-scale analysis in diagnostic genetic testing.

2598M

PCRstable: Chemical stabilization technology for ambient genetic testing. *G. Dodson, V. Liberal, S. De los Rios, R. Muller.* Research & Development, Biomatrica, Inc., San Diego, CA.

The need for cold storage and shipping of genetic screening reagents makes the use of these reagents difficult in regions where access to freezers and dry ice is either non-existent or undependable. PCRstable is a novel technology developed to stabilize PCR-based testing components dry at ambient temperatures. We have developed and tested the stability and functionality of stabilized PCR and RT-PCR reagents for genetic screening, which are described in this study. We combined the test reagents with proprietary biostability compounds and applied a simple air drying procedure for stabilization. Following accelerated aging studies at elevated temperatures, we performed both PCR and RT-PCR reactions to assess the ability of the stabilized reagents to perform the respective reactions. For both PCR and RT-PCR assays, we found excellent stability at ambient temperatures for both end-point and quantitative PCR. Preservation of PCR and RT-PCR reagents with PCRstable biostability compounds lead to high retention of both reagent stability and activity at ambient and elevated temperatures. These procedures could be applied to PCR-based genetic screening methods to eliminate their cold chain requirements. This would allow such genetic testing in locations currently lacking access to reliable cold storage by significantly reducing storage and associated shipping costs .

2599T

Public perceptions of disease actionability and severity and their potential utility for making decisions about Genomic Testing Results.

P.S. Sinicrope[1], K.D. Graves[2], Y. Zhou[2], J.B. McCormick[3], S.T. Vadaparampil[4], N.M. Lindor[1]. 1) Health Sciences Research, Mayo Clinic , Scottsdale, AZ; 2) Lombardi Comprehensive Cancer Center, Georgetown University, Washington D.C; 3) Divisions of General Internal Medicine and Health Care Policy & Research, Biomedical Ethics Program , Rochester MN; 4) Moffitt Cancer Center, Tampa, FL.

Background: While much research highlights disease actionability and severity as important factors informing the return of genomic results, few studies actually address public perceptions of these concepts. We examined perceptions of disease actionability and severity and their relationship to clinical, attitudinal, and sociodemographic characteristics. We also explored individuals' perceptions of the utility of disease actionability and severity in making hypothetical decisions about general genomic testing results. Methods: We conducted a cross-sectional online survey with a representative sample of US adults using the YouGov opt-in panel. Our survey was informed by focus groups, the multi-disciplinary investigative team, and cognitive interviews. The survey contained 136 items and 3 videos covering the basics of genetics and disease actionability and severity. Sections included: 1) Overview and Motivations for Genetic Testing; 2) Perceptions of Severity; 3) Perceptions of Actionability; 4) Perceptions of Severity/Actionability combined; 5) You do the Binning; and 6) Assessing the Binning Process/Demographics. Results: Respondents included 900 adults (51% Female, 65% White, 12% Black, and 15% Hispanic) with an average age of 45 (SD=17). About 64% reported knowing "something, but not very much" about genetic testing and 60% preferred to make their own decisions about genomic results they would want to learn about vs. letting medical experts decide. The majority (>85%) found the concepts of actionability and severity to be useful individually and combined; 46.6% indicated a preference for using actionability over severity. Over half (53.8%) reported being very/extremely confident in their ability to score for actionability and severity. Participants' scoring of actual medical scenarios varied for both actionabilty and severity ($p<0.0001$). Conclusions: We found that actionability and severity are perceived quite variably, raising the question of how meaningful current binning efforts are for patients. Despite the challenges of scoring disorders, respondents indicated confidence in their ability to score, to make decisions for themselves versus expert decisions, and to receive results of all types, with only a minority want to be more selective. These results may inform development of patient-centered approaches to decision making in genomic testing.

2600S

Conductive, Sensorineural, and Mixed Hearing Loss in Patients with Down Syndrome. *A. Musso[1], A. Umrigar[1], K. Foley[1], D. Mercer[1], M. Marble[1,2], F. Tsien[1].* 1) Genetics , LSUHSC New Orleans, New Orleans , LA., USA; 2) Genetics, Children's Hospital of New Orleans, New Orleans, LA, USA.

According to the Centers for Disease Control and Prevention, an estimated 75% of patients with Down syndrome suffer from some form of hearing loss. The hearing loss can be conductive hearing loss (CHL), sensorineural hearing loss (SNHL), or mixed hearing loss, which is a combination of CHL and SNHL. CHL in Down syndrome patients is usually due to otitis media with effusion which is managed medically with antibiotics and pressure equalization (PE) tubes. Causes of SNHL in the Down syndrome population are congenital inner ear abnormalities including internal auditory canal hypoplasia, enlarged vestibular aqueducts, cochlear nerve hypoplasia, and others. Due to the irreversible nature of SNHL, hearing aids and speech therapy are the optimal intervention for patients with Down syndrome. The incidence of SNHL in Down syndrome children ranges from 4% to 55%. We are conducting a retrospective study of at least 200 patients evaluated in the Down syndrome clinic at Children's Hospital of New Orleans and their satellite genetics clinics throughout the state of Louisiana. Based on audiograms of the hearing loss, we will discern whether the hearing loss is conductive, sensorineural, or mixed. We will evaluate the following standard audiological procedures: otoscopy, tympanometry, otoacoustic emissions, and pure tone audiometry. Hearing loss type and severity will be determined by comparing air conduction and bone conduction pure tone thresholds. This study will provide further data on the incidence of hearing loss in children with Down syndrome. Geneticists can play an important role in enhancing the communicative outcomes of Down syndrome patients through referral for early hearing and speech intervention.

2601M

HUMAN COMMUNICATION DISORDERS IN PATIENTS WITH DOWN SYNDROME. *A. Romero-Diaz[1], J.M. Aparicio-Rodriguez[2].* 1) Phoniatrics Otoneurology; 2) Genetics, Hospital para el Nino Poblano, Puebla, Mexico.

Down syndrome is one of the most frequent chromosomal abnormalities live births. Most patients with a trisomy of a portion or all of the chromosome. Its prevalence is approximately 1/800 live births. The human communication disorders are common in patients with Down syndrome. Morphological abnormalities of the head and neck frequently in patients with Down syndrome as midface hypoplasia, characterized by malformation of the eustachian tube, short palate, macroglossia, and narrowing of the oropharynx and nasopharynx, joined as systemic factors associated with muscular hypotonia, results in a high incidence of recurrent otitis media and sleep apnea syndrome in these patients. HEARING.- The prevalence of hearing loss in children with Down syndrome is high, especially the transmission or conductive hearing loss since between 50-70% of cases have chronic serous otitis media between 3 and 5 years; the neuro-sensory deafness represents only 4%. LANGUAGE DEVELOPMENT.- Language development is usually delayed. This delay correlates with cognitive ability; Most oral communication, which may be more or less significant defects language is obtained; and in some cases can not develop oral language and require augmentative and / or alternative communications. CHANGES IN THE VOICE.- Typically have a high palatal vault, large tongue, generalized hypotonia and maxillary hypoplasia. No skeletal abnormalities affecting craniofacial features. Although his brain is structurally normal can be observed brachycephaly and a flattened shape in the center of the face due to a malformation of the frontal bone. DOWN SYNDROME AND STUTTERING.- Have certain degrees of learning difficulties that can affect the ability to understand and produce speech and language. One of the biggest problems for some people with Down syndrome is the unintelligibility of their speech toward others. Unintelligibility and disfluency often go hand in hand. MEMORY.- Children with Down syndrome usually show memory skills short better than visual-term verbal memory. This means that they will learn more easily if the information presented to them visually if they are presented verbally. ATTENTION.- In Down syndrome there are alterations in the brain mechanisms involved when changing focus of attention. So often have difficulty sustaining attention for extended periods of time and ease of distraction from diverse and novel stimuli.

2602T

Corpus Collosum Agenesis and Psychomotor Retardation in a Female Patient with 15.4 Mb Deletion of 14q12→q21.2 and 550 kbp deletion of 18p11.23; Microarray Delineation of Imbalanced Chromosomal Rearrangement. *D. Torun[1], M. Arslan[2], H. Akar[1], Y. Tunca[1].* 1) Dept. of Medical Genetics, Gulhane Military Medical Faculty, Ankara, Turkey; 2) Dept. of Pediatric Neurology, Gulhane Military Medical Faculty, Ankara, Turkey.

Agenesis of the corpus callosum (ACC) is characterized by the partial or complete loss of the nerve fibers that connect the cerebral hemispheres. ACC is one of the most common central nervous system abnormalities and the etiology of ACC is associated with several chromosomal, single gene and environmental causes. A 9 month female patient from non-consanguineous healthy parents has been referred for psychomotor retardation and ACC detected by cranial MRI in postnatal period. Prominent forehead, hypertelorism, upslanted palpebral fissures, nystagmus, depressed nasal bridge, prominent philtrum, thin upper lip, deep palmar creases observed on her clinical examination. Physical exam revealed that height, weight and head circumference was below the 3th percentile. She treated for congenital hypothyroidism. Abdominal ultrasound, echocardiography was normal. Karyotype analysis of patient revealed a de nova apparently balanced translocation between chromosomes 14 and 18 [t (14; 18) (q13; p11)]. To investigate if the translocation was really balanced, analysis on the Affymetrix Cytoscan 750K was done. Microarray analysis revealed a deletion spanning about 15.4 Mb at the chromosomal region 14q12-14q21.2 {arr[hg19] 14q12q21.2(28,431,057-43,915,132)x1} and 550 kbp deletion at 18p11.23 chromosomal region { arr[hg19] 18p11.23(7,281,346-7,816,276)x1}. Deleted regions on chromosome 14 and 18 were consisting of 40 and 1 OMIM genes, respectively. The presence of ACC and microcephaly is in agreement with previously published report [PMID: 18627055]; however, seizure has not been detected in current case. This case further shows haploinsufficiency of multiple genes resulting in ACC and developmental delay. Microarray analysis is a sensitive method to detect imbalances in reciprocal translocations.

2603S

Deletion of 17p11.2 encompasses *FLCN* with increased risk of Birt-Hogg-Dubé in Smith Magenis Syndrome: Recommendation for Cancer Screening. *A.C.M. Smith[1], L.R. Fleming[2], A.M. Piskorski[3], A. Amin[3], C. Phorphutkul[4], S. de la Monte[4], E. Stopa[4], W. Introne[1], T. Vilboux[1], F. Duncan[1], J. Pellegrino[5], B. Braddock[5], L.A. Middelton[6], C. Vocke[6], W.M. Linehan[6].* 1) Office Clinical Dir, NHGRI/NIH, Bethesda, MD; 2) Medical Genetics Branch, NHGRI/NIH, Bethesda, MD; 3) Department of Pathology, Warren Alpert SOM Brown Univ., Rhode Island Hosp., Providence, RI; 4) Department of Pediatrics, Warren Alpert SOM Brown Univ., Rhode Island Hosp., Providence, RI; 5) Medical Genetics, Upstate Medical University, Syracuse, NY; 6) Urologic Oncology Branch, NCI/NIH, Bethesda, MD.

We report two unrelated adults with Smith Magenis syndrome (SMS) who were confirmed to have renal histology consistent with Birt-Hogg-Dubé (BHD) syndrome. These two cases suggest the need to expand previous management recommendations for SMS to include cancer surveillance starting at 21 years of age.

BHD is an autosomal dominant disorder characterized by increased risk of cutaneous fibrofolliculomas, pulmonary cysts, spontaneous pneumothorax, and renal tumors. BHD renal tumors are most commonly chromophobe or hybrid renal cell carcinoma, although clear cell RCC or oncocytoma may also occur. Haploinsufficiency for folliculin (*FLCN*), a gene that maps within the common SMS del 17p11.2 region, poses a risk to develop BHD. While 88% of BHD cases are attributed to *FLCN* point mutations, deletion or duplication of *FLCN* in BHD has been described. As the majority of SMS cases result from a 3.7Mb deletion in 17p11.2 which includes *FLCN*, there has been a theoretical concern about increased cancer risk in SMS patients. However, these two cases constitute the first known association of BHD characteristic tumors in SMS patients.

Case 1 was a 50-year female cytogenetically diagnosed with SMS (del 17p11.2) after undergoing radical nephrectomy for bilateral chromophobe renal cell carcinoma and hybrid tumors. Her history was negative for skin lesions or lung cysts. Case 2 was a 45 year old adult male with SMS, who presented with a high grade B-cell lymphoma; chemotherapy was declined and he died one month later. A full autopsy revealed diffuse lymphomatous involvement of all major organs, oncocytosis of the right kidney, and incidental oncocytomas of the left kidney. Molecular SNP analysis confirmed a 7.1 kb 17p11.2 deletion (14,374,758-21,539,613; HG Build 18) that encompassed *RAI1* and *FLCN* with a second *FLCN* mutation (c.736delA; S246fs) seen in the renal specimens.

Conclusion: Given these observations, we propose clinical surveillance for BHD-associated renal tumors by baseline abdominal and pelvic MRI or CT scan with/without contrast for SMS adults beginning at age 21 years and every 3 years thereafter.

† Trainee Award Finalist

2604M

AMMECR1 gene disruption and expression impairment in a balanced X-autosome translocation patient. *M.M. Oliveira[1], R.S. Guilherme[1,2], V.A. Meloni[1], N. Kosyakova[2], T. Liehr[2], G. Carvalheira[1], M.I. Melaragno[1].* 1) Departamento de Morfologia e Genética, Universidade Federal de Sao Paulo, Sao Paulo, Brazil; 2) Institute of Human Genetics, Jena University Hospital, Friedrich Schiller University, Jena, Germany.

Detailed molecular characterization of women with balanced X-autosome translocations has proven to be valuable in searching candidate genes for diseases. Patients with balanced chromosomal rearrangements involving an X-chromosome generally present a skewed X inactivation, with the normal X-chromosome being preferentially inactivated. Thus, a gene disruption in X-chromosome frequently results in absence of a functional copy of the affected gene. Here we report a seven year old female patient who presented with karyotype 46,X,t(X;9)(q23;q12)dn, disproportionate short stature, microsomia, minor facial dysmorphisms, septal atrial defect with thin aneurysmal septal tissue, scoliosis, diffuse bone demineralization with peripheral osteopenia, recurrent sinusitis, and normal cognition. Replication banding showed a preferential inactivation of the normal X-chromosome in lymphocytes, and genomic array indicated for a balanced translocation. The autosomal breakpoint affects a heterochromatic region. X-chromosome breakpoint mapping was performed by array painting technique using a glasneedle-microdissection derived DNA probe, exclusively of the derivative chromosomes. Array painting determined the disruption of AMMECR1 gene at intron 1, within the genomic coordinates chrX:109,509,444-109,552,503 (GRCh37/hg19). The breakpoint localization was validated by FISH using the BAC probe CTD-3066N24. AMMECR1 mRNA whole blood expression was assessed by qRT-PCR in the patient and eight female controls. Two TaqMan assays were used and both showed absence of AMMECR1 expression in the patient and high gene expression levels in all controls. Therefore, AMMECR1 gene function impairment may be related to her clinical abnormalities. Although AMMECR1 gene function is unknown, its expression is ubiquitous in human organs and tissues, with exception of the nervous system. AMMECR1 gene expression occurs primarily in organs and cell types that were affected during the patient's development. AMMECR1 protein is highly conserved throughout evolution, presents nuclear localization, several phosphorylation sites and similar conformation as proteins that interact with nucleic acid, suggesting a regulatory role. This patient may represent a unique model for the evaluation of the phenotypic consequences of AMMECR1 impairment in humans and the claim of this gene as a candidate for the observed phenotype requires further functional and genetic validation. Financial support: FAPESP.

2605T

GENETIC BASIS OF AGING, TELOMERES AND TELOMERASE. FOUR PEDIATRIC PATIENTS WITH PREMATURE AGING OR COCKAYNE'S SYNDROME. *M. Barrientos- Perez[1], J.M. Aparicio-Rodriguez[2,3].* 1) Endrocrinology; 2) Genetics, Hosp para el Nino Poblano, Puebla; 3) Estomatology Benemérita Universidad Autónoma de Puebla, México.

There has been a direct relation between Telomere, telomerase and aging due to multiple investigations that are carried out on them, so it has been associated with human youth and the cure of cancer. There is a direct relationship in the alteration of the Telomeres and Telomerase with stress and inbreeding. With respect to consanguinity there is a syndrome of premature aging called Cockayne syndrome. Four cases of cockayne syndrome (CS) from two families with first degree consanguinity living in a small town where radioactive uranium (UR) mines were found. The clinical manifestation of CS were analyzed in the Hospital Para el Niño Poblano (HNP) by a multidisciplinary study. The main clinical manifestations were dermatological alterations as photosensitivity to sunligth and predisposition to skin cancer "xeroderma pigmentosum" (XP); endocrinology alterations as dwarfism and senile appearance; ophthalmological findings from cataracts to pigmentary retinal degeneration; neurological alterations as mental retardation and sensorial hearing loss; and upper limbs, lower limbs and vertebral column degeneration was also found. Four cases with Cockayne Syndrome were diagnosed in two Mexican families with consanguinity from an area with radioactive contaminants, showing that consanguinity is the etiological factor of this syndrome. Keywords. Consanguinity, autosomal recessive inheritance, Cockayne's syndrome, Telomeres, telomerase, uranium, xeroderma Pigmentosus.

2606S

Copy Number Variation in Oculoauriculovertebral Spectrum. *M. Colovati[1], S. Bragagnolo[1], R.S. Guilherme[1], A.G. Dantas[1], C.A. Kim[2], A.B. Perez[1], M.I. Melaragno[1].* 1) Genetics Division, Department of Morphology and Genetics, Universidade Federal de São Paulo, São Paulo, Brazil; 2) Genetics Unit, Instituto da Criança, Universidade de São Paulo, São Paulo, Brazil.

Oculoauriculovertebral spectrum (OAVS, OMIM 164210) is a clinically heterogeneous condition that includes unilateral or bilateral ear anomalies, hemifacial microsomia, ocular defects, conductive and/or sensorineural hearing loss, orofacial clefts, vertebral malformations, and more rarely, cardiac, renal and cerebral malformations and intellectual impairment. Most cases are sporadic with rare familial cases exhibiting autosomal dominant inheritance with incomplete penetrance and variable expressivity. Several chromosomal abnormalities have been associated with OAVS but no recurrent abnormalities were identified to define candidate regions. We described 62 patients with OAVS clinical diagnosis that fit in the minimal criteria according to Tasse et al. (2005), including seven familial cases. All showed normal karyotypes, except one: 46,XX,inv(12)(q15q24.1). One patient presents a der(4) from a critical maternal translocation t(X;4)(p22.31;p15.33) detected by FISH. Genomic arrays identified 13 potentially pathogenic copy number variations (CNV): 7 deletions (4p16.3p15.33, 4q13.3q21.1, 8q13.3, 10q26.2q26.3, 16p13.3, 22q11.21, Xp22.33) and 6 duplications (2q32.1, 4p16.1.1, 16p13.11, 17q11.2, Xp22.33). The 4p16.1 duplication (~950 kb) was observed in two non-related patients. Two patients showed overlapping CNVs in Xp22.33 (~3,2 Mb deletion and ~9 Mb duplication). None of these CNVs were described in the Database of Genomic Variants (DGV), except the duplication 4p16.1 that was possibly relevant because involves HMX1 gene. Some variants of unknown clinical significance (VOUS) were also identified when DGV and the internal control database were considered. In this study some deleted and duplicated regions comprise genes with relevance to the OAVS phenotype, such as BAPX1 in 4p15.33, responsible for hemifacial microsomia; HMX1 in 4p16.1, involved in the oculoauricular syndrome; EYA1 in 8q13.3, important for the craniofacial development; YPEL1 and ERK1 in 22q11.2, involved in branchial arch development. Besides, larger deletions, duplications and VOUS found in our patients include many genes described in OMIM and genomic regions which may be related to the phenotype. These candidate genes and/or regions will deserve further investigation. The identification of the OAVS genes and genomic position are important to understand better the molecular pathways and clinic heterogeneity of the syndrome. Financial support: FAPESP, Brazil.

2607M

A rare combination: mosaic Turner syndrome by isochromosome Xq with 17p13.3p13.2 microduplication and congenital cataract with autosomal dominant inheritance and incomplete penetrance in the same individual. *J.A. Rojas Martínez[1], J.C Acosta Guio[2,3].* 1) Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogotá, Colombia; 2) Instituto de Ortopedia Infantil Roosevelt. Bogotá. Cundinamarca. Colombia; 3) Instituto de Investigación en Nutrición, Genética y Metabolismo. Universidad El Bosque. Bogotá. Cundinamarca. Colombia.

The combination of Turner syndrome with congenital cataract with autosomal dominant inheritance and incomplete penetrance has not been reported, since even this form of cataract in isolated presentation is exceptionally rare. There are also no reports presentations with rearrangements on chromosome 17. Next, we present the case of a twenty month old girl born in Bogotá (Colombia), product of first pregnancy maternal, non-consanguineous parents. She was diagnosed a month living with bilateral congenital cataract and also has a history of neurodevelopmental delay and background in the family of mother, grandmother and maternal uncle with bilateral congenital cataract. Physical examination found multiple minor facial anomalies, and other abnormalities. Neurological examination with generalized hypotonia with truncal predominance, and scarce and unclear language. Suspecting chromosomal abnormalities, diagnostic study begins with high resolution BG karyotype reporting: mos 46,X,iso(Xq)[82]/45,X[18], with the presence of a mosaic given by major cell line with a isochromosome of the long arm of the X chromosome and another cell line with complete monosomy of the X chromosome. Comparative Genomic Hybridization (CGH) was performed, confirming the previous cytogenetic analysis. Additionally, a 274 Kb duplication was identified in 17p13.3p13.2 (with breakpoints in 3561131-3835801) that compromises 8 genes. Congenital cataract with clear autosomal dominant inheritance present in this patient exhibits incomplete penetrance, as the mother of the patient must have the mutated allele, because she transmits to her progeny, but is not expressed in her phenotype. Furthermore 8 genes were identified in the region corresponding to the duplicate 17p13.2p13.3 segment in this patient. One of these genes, GSG2 encodes a protein called haspina whose overexpression has been shown to be associated with genomic instability and aneuploidy, so the authors suggest that the microduplication 17p13.2p13.3 ontogenetically preceded and probably contributed to the development of aneuploidy and abnormality of the X chromosome present in this case. As in other reports of chromosomal abnormalities that occur simultaneously, in the above case the clinical manifestations result from the superposition of two phenotypes, therefore this report contributes to a better definition of the clinical features of 17p13.3 microduplication syndrome.

2608T

Polycystic kidney disease and multiple malformations in a patient with tetrasomy 2q13q21.1. *L. Dupuis[1], C. Roadhouse[1], R. Badilla-Porras[1], D.J. Stavropoulos[2], R. Mendoza-Londono[1].* 1) Clinical Genetics, Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada; 2) Department of Paediatric Laboratory Medicine, The Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada.

Autosomal tetrasomies are rare events which typically cause a severe clinical phenotype. Well characterized autosomal tetrasomies include cat-eye syndrome (tetrasomy 22), Pallister Killian (tetrasomy 12p) and tetrasomies of 9p and 18q. We describe a case of chromosome 2q13-q21.1 triplication, resulting in tetrasomy for this segment. This is the first report of a liveborn with a triplication of chromosome 2 of this magnitude (19.876 Mb). The patient presented at day two of life with dysmorphic features, bilateral hip dislocation, overlapping toes and bilateral pneumothoraces. Abdominal ultrasound revealed didelphys uterus, large polycystic ovaries and dysplastic cystic kidneys that led to end stage renal disease and kidney transplant. At her most recent evaluation at the age of four years, medical issues included global delay, lung hypoplasia, bilateral hip dislocation, G-tube dependence, strabismus and mild sensorineural hearing loss. MRI of the brain revealed appropriate myelination and prominent lateral and third ventricles but no evidence of hydrocephalus. Microarray analysis (44,000 oligo) showed that the area of tetrasomy involved 115 RefSeq genes and 13 OMIM Morbid Map genes. A thorough review of the literature revealed two previous cases of tetrasomy of chromosome 2q that partially overlapped the area involved in our patient. Mercer et al (2009) reported an adult female with a 7.28Mb triplication of 2q12.3-q13 with infertility and had left polycystic kidney, patent ductus arteriosus, left streak ovary, bicornuate uterus and deafness. Wang *et al* (1999) reported a stillborn with a larger area of tetrasomy 2q11.2-q21 with brain malformations, multicystic kidneys, absence of the right thumb and cleft lip and palate. The area of overlap between these three cases comprises a 3Mb region on 2q13 (chr2:111,158,401-114,416,131 hg18) that harbors four OMIM genes. These include *MERTK* a tyrosine kinase that when absent can result in retinitis pigmentosa, *IL1B* and *IL1RN* which participate in the immune system, and the transcription factor *PAX8* that when mutated can result in lesions of the excretory system and thyroid dysplasia. This case series narrows the critical region for renal cyst and genital malformation susceptibility on chromosome 2 and expands the phenotype of patients with tetrasomy of this region. It is not presently clear which genes in excess dosage from this area of triplication may be responsible for different aspects of the phenotype.

2609S

Neurodevelopmental Profile and Cognitive Variability in Two Females with the Rare 48, XXXX Chromosomal Disorder. *D.C. Gibbs[1], A.L. Gropman[2,4], T. Sadeghin[3], C. Samango-Sprouse[1,2,3,4].* 1) The Focus Foundation, Davidsonville, MD; 2) George Washington University School of Medicine, Washington, DC; 3) Neurodevelopmental Diagnostic Center, Davidsonville, MD; 4) Children's National Medical Center, Washington, DC.

48, XXXX is a rare chromosomal aneuploidy associated with neurocognitive deficits, speech and language disorders and executive dysfunction. Less than 60 cases have been reported in the literature however limiting our understanding of the 48, XXXX neurodevelopmental phenotype. To our knowledge, this is the first study to report on the neurodevelopmental, cognitive and behavioral profile of two females with 48, XXXX of similar ages. Patient 1 (age=11.0) and patient 2 (age=10.9) were evaluated using the Wechsler Intelligence Scale for Children-4th Ed. (WISC-IV), the Leiter International Performance Scale (LIPS-R), Expressive and Receptive One-Word Vocabulary Tests-4th Ed. (ROWPVT-4/EOWPVT-4) the Beery Buktenica Test of Visual Motor Integration-5th Ed. (Beery-VMI), the Child Behavioral Checklist (CBCL) and the Behavioral Rating Inventory of Executive Function (BRIEF). Verbal IQ's were 44 and 68 on the WISC-IV and 56 and 80 on the LIPS-R for patient 1 and 2, respectively. Overall vocabulary was stronger in patient 2 than patient 1, but receptive vocabulary was stronger than expressive in both patients (Patient 1: 62 / 55; Patient 2: 83 / 79; ROWPVT-4, EOWPVT-4). Both 48, XXXX girls had significantly impaired visual motor capacities in graphomotor and perceptual domains (<5th percentile Beery-VMI subtests) as well as significant executive dysfunction and behavioral/social deficits (<10th percentile on CBCL and BRIEF subtests). Patient 1 had an extensive history of family learning disabilities (FLD) likely contributing to her more severely delayed cognitive development. We hypothesize that the co-existing ADHD in both subjects compounded their existing social and behavioral deficits. The visual-motor deficits are novel findings in 48, XXXX and likely to contribute to patients' delayed cognitive development. These distinct and overlapping features are characteristic of other X and Y chromosomal variations and may be of great use in the diagnosis, treatment and counseling of 48, XXXX patients and families.

2610M

48,XXYY Syndrome: a wide spectrum of phenotypic presentation. A case report. *L.F. Imbachi[1], A.I. Sanchez[1], P.M. Hurtado[1, 2].* 1) Pontificia Universida Javeriana Cali, Cali, Colombia; 2) Centro Medico Imbanaco Cali, Colombia.

Introduction: Sex chromosome aneuploidies are the most common chromosomal abnormalities with an incidence of 1 in 400 births. Tetrasomy and pentasomy conditions (48,XXYY, 48,XXXY, 49,XXXXY) are less frequent with an incidence of 1:18,000-1:100,000 male births. 48,XXYY syndrome is the most common of these three occurring in 1:18,000-1:40,000 male births and is related to a wide spectrum of physical findings, congenital malformations, cognitive impairment, developmental delay and behavioral problems. Case Presentation: This is a 3-year-old (y/o) male patient who was born by cesarean section from non-consanguineous parents after a first full term uneventful gestation. Maternal and paternal ages were 29 y/o. No significant family history. He started to have myoclonic, afebrile seizures at 4 months old. EEG showed generalized discharges. He was started on anticonvulsants with complete remission of seizures. At 2 and a half y/o aggressive behavior and motor stereotypies were noticed. Clinical examination revealed global developmental delay with plagyocephaly, hypertheloríism, severe truncal hypotonia with head lag communication problems. MRI spectroscopy revealed symmetric signal abnormality in thalamus, basal ganglia and brainstem. Urine organic acids were negative; pyruvate, plasma aminoacids and Acylcarnitine profile were normal. Mutations or deletions were not detected in mitochondrial DNA screen or in FOXG1/MECP2 sequence analysis. Karyotype showed 48,XXYY chromosome abnormality. Discussion: This is a male patient with 48,XXYY syndrome who has physical and neurodevelopmental features consistent with it. Common findings include hyperthelorism, dental problems, clinodactyly and hypogonadism. The first two were evident in our patient as well as hypotonia, which occurs in 75% of patients. Congenital malformations like cleft palate, heart and kidney defects were ruled out in this patient. About 10-15% of individuals have seizure disorders, like this child. Elevated rates of autism (28-45%), mood disorders (46,8%) and attention deficit and hyperactivity disorder (70%) have been seen. Therefore this could explain patients' aggressive behavior and repetitive conducts. Conclusions: 48,XXYY syndrome has a wide spectrum of phenotypic presentation. It is important to identify the variability in physical features and cognitive functions to establish an early diagnosis, initiate management and to offer appropriate genetic counseling and therapies for rehabilitation.

2611T

Clinical features of 5p13 duplication syndrome. *T. Kuchikata, S. Itou, H. Yoshihashi.* Division of Medical Genetics, Tokyo Metropolitan Child Medical Center, Tokyo, Japan.

5p13 duplication syndrome (OMIM #613174) is newly recognized chromosomal microduplication syndrome characterized by developmental delay and intellactual disability, facial dysmorphisms. NIPBL in the duplicated region is known as a part of causative gene. Loss-of-function mutations in NIPBL cause cohesin abnormality, which leads to chromosome segregation abnormality during the mitotic and meiotic cell cycles. A part of cohesinopathy causes Cornelia de Lange syndrome (CdLS). Here,we describe a Japanese case with 5p13 duplication syndrome. Subject: 4y-male. He was delivered at term without any critical problems during the perinatal period ,except for a birth weight of 1.9kg. His unrelated parents and siblings are healthy. He had a delay in psychomotor development with aging. At the age of 3 month old, brain MRI revealed there was no definitive abnormality and myelination was not retarded. At the age of 2 year old, the weight was 11.2kg (25th percentile) and the height was 83.3cm (25th percentile). He had mild bilateral auditory disorder, strabismus, right cryptochidism, left accessory ear and dysmorphic features which are prominent forehead, downslating palpebral fissure, hypomalar, broad nasal root, long philtrum and long slender fingers. The gestalt was suggestive of Kabuki syndrome. KMT2D and KDM6A mutations were not detected. Radiological survey for skeletal dysplasia revealed no definitive abnormality. A peripheral karyotype was 46,XY and there was no submicroscopic subtelomeric chromosomal rearrangements. Array-CGH(ISCA:CGH180K, Agilent)showed an arr[hg19]5p13.2(36374209-37331912)×3. The chromosomal abnormality arose de novo. The duplicated region contains NIPBL and SCL1A3, C5orf43, NUP155. Discussion: This is the first case of 5p13 duplication syndrome in Japan. This syndrome is a counterpart of CdLS with many characteristic features,but has no cardinal feature of diagnostic value. Although the phenotype in this case overlaps those of cases with the 5p13 duplication syndrome in previous reports, it seems to be difficult to recognize this syndrome from the point of view of dysmorphic features. The phenotypic effect of NIPBL haploinsufficiency is well known as CdLS and the other diseases related to cohesion abnormality in clinical and research work. However, to our knowledge, there is no sufficient scientific knowledge about the gene dosage effect of NIPBL. Further cases may provide insight into the consideration of clinical information.

2612S

Klinefelter Syndrome (48, XXXY) in a Patient With Mild Mental Retardation and Psychotic Personality Traits. *H. Pachajoa[1, 2], M.F. Hernandez[1].* 1) Centro de Investigaciones en Anomalías Congénitas y Enfermedades Raras (CIACER), Universidad Icesi, Cali, Colombia; 2) Fundación Clínica Valle del Lili, Cali, Colombia.

Introduction: Klinefelter syndrome is the most common aneuploidy in males with a prevalence of 0.1-0.2% in the general population, which rises up to 3% in males with fertility issues, although only 35% of cases are diagnosed. The affected males tend to be tall, have narrow shoulders, wide hips, sparse body hair, gynecomastia and small testis; they present androgen deficiency and azoospermia. Besides the previous physical characteristics, there have been reports on the expression of behavioral and cognitive traits that tend to be very variable, possibly according to the type of aneuploidy, with an established association between the number of extra X chromosomes and cognitive deficit, and a not so clear association of different chromosomal variants of Klinefelter and psychotic behavior, with some authors proposing the origin of the extra chromosome as a determinant of different behavioral traits. Case report: 13 year old male with karyotype 48,XXXY, who besides presenting the classical physical features, is under psychiatric treatment because of presenting trouble at home and at school for being aggressive and impulsive. A review on the literature is made. Conclusion: When approaching a patient with Klinefelter syndrome is necessary to acknowledge the importance of an opportune start of hormonal therapy and a multidisciplinary approach including endocrinology, pediatrics, genetics, neuropsychology and psychiatry when needed.

2613M

Maternal Uniparental Disomy Prader-Willi Syndrome in an XYY boy. *P. Phowthongkum[1], J. Berkowitz[2], A. Schneider[2].* 1) Internal Medicine, Einstein Medical Center Philadelphia, Philadelphia, PA; 2) Genetics Division, Einstein Medical Center Philadelphia, Philadelphia, PA.

Prader-Willi Syndrome (PWS) is a neurodevelopmental disorder characterized by neonatal hypotonia, failure to thrive, intellectual disability, hyperphagia, short stature, and characteristic behaviors. PWS is associated with chromosome 15q11-13 deletions, maternal UPD or an imprinting abnormality. 47,XYY is a sex chromosome disorder which is usually subtle. We report long term follow up of a ten year old boy with PWS and 47,XYY. He was born at term by emergency C-section for fetal distress to a 27-year old G1P0 mother and 30-year old father. At delivery he was hypotonic with poor gag reflex. Examination revealed cryptorchidism and subsequent feeding difficulty. Karyotype of 47,XYY did not explain his phenotype and prompted testing for PWS. Methylation study was consistent with PWS. FISH for del15q12 was negative. Microsatellite studies confirmed maternal UPD. He was started on growth hormone at 6 months. His height is at the 75th percentile, which may be attributed to the 47,XYY. After having some periods of hyperphagia and food searching behaviors as a toddler, he demonstrated very good self-control with his eating until recently. He has developed general anxiety and obsessive behaviors, and is exhibiting aggression due to frustration from being unable to keep up with peers. He has become more self-aware of his condition and is having more difficulty with peer-to-peer interaction. His neurodevelopmental evaluation revealed an IQ of 72. He performed very well on math and expressive reading but poorly on language comprehension and sound awareness. He does not have autistic spectrum features. There are few reports about genotype-phenotype correlation for PWS. Maternal age is considered a risk factor for UPD15 PWS, but his mother was only 27 at delivery. Evaluation of PWS individuals who developed psychosis, found the majority have matUPD15. Further longitudinal study is required to establish the natural history and clinical course for this subgroup of patients prone to neuropsychiatric issues. In terms of XYY involvement, as non-disjunction happens independently in maternal oogenesis and paternal spermatogenesis, the odds of recurrence of this mixed phenotype is extremely low. The evolution of this child's PWS manifestations and how XYY syndrome may have mitigated some of the physical features will be discussed. We will also explore how XYY may be adding to the neuropsychiatric manifestations.

2614T

The Extended Phenotypic Spectrum of 7p14.3-15.3 deletion syndrome. *M. Michelson-Kerman[1,2], D. Lev[1,2], C. Vinkler[1,2], L. Blumkin[1,3].* 1) Inst Med Genetics, Wolfson Medical Ctr, Holon, Israel; 2) Metabolic Neurogenetic clinic, Wolfson Medical Ctr, Holon, Israel; 3) Pediatric Neurology Unit, Wolfson Medical Ctr,Holon, Israel.

Structural aberration of chromosomes are associated with various syndromes. Microdeletion of the short arm of chromosome 7p is rare. Phenotypic presentation is variable and include craniofacial malformations, hypoplastc ears, heart defects, hypotonia, short fingers with tapering phalanges and severe developmental delay. Most of the cases are de novo deletions. Few cases are due to unbalanced rearrangements. We present a twenty months-old girl with severe developmental delay, dysmorphic features, hypoplastic ear cruses, tapering fingers, hypotonia, severe ataxia and congenital alacrima. Karyotype in leucocytes was normal, 46 XX. Affymetrix Cytogenetics Whole-Genome HD Array chip revealed a 10375.473 KBP loss in chromosome 7, between p14.3 and p15.3. This region harbors 96 known genes, including HOXA gene cluster and several cytogenetic regions associated with genetic disorders . Parental studies are normal. Although most of phenotypic features were previously described in patients with 7p microdeletion, our patient presents with distinct features including severe congenital alacrima and severe ataxia. These features have not been described previously.

2615S

Pigmentary mosaicism type Ito in a balanced X-autosome translocation with no gene disruption at the breakpoint. *M. Melaragno[1], R.S. Guilherme[1], C.E. Steiner[2], G.M. Carvalheira[1], N. Kosyakova[3], T. Liehr[3], M.M. Oliveira[1].* 1) Disciplina de Genetica, Universidade Federal de Sao Paulo, Sao Paulo, Brazil; 2) Department of Medical Genetics, Universidade Estadual de Campinas, São Paulo, Brazil; 3) Institute of Human Genetics, Jena University Hospital, Friedrich Schiller University, Jena, Germany.

Pigmentary mosaicism type Ito (PMI) is a heterogeneous symptom complex characterized by hypopigmented whorls and streaks following Blaschko's lines, with neurological deficits, epilepsy and/or asymmetric abnormalities in other organs. Although its etiology is not fully understood, PMI is typically associated to apparently balanced X-autosome translocations involving Xp11, especially the band Xp11.2. Several hypotheses have been proposed to explain this association, such as disruption of X-linked genes, effect position, mosaic functional Xp disomy and functional autosomal monosomy due to a random X-inactivation pattern. We report the first patient with PMI and balanced X-autosome translocation involving Xp11 that was submitted to a high resolution breakpoint definition. An 18-year-old girl presented with karyotype 46,X,t(X;21)(p11.22;p11.1), Blaschko linear hypopigmentation in the trunk, neurodevelopmental delay, and electroencephalographic abnormalities. The normal X-chromosome was preferentially inactivated in lymphocytes, genomic array indicated for a balanced translocation, and the autosomal breakpoint affects a heterochromatic region. X-chromosome breakpoint mapping was performed by array painting technique using a glasneedle-microdissection derived DNA probe, exclusively of the derivative chromosomes. Array painting determined the Xp11.22 breakpoint as intergenic, within the genomic coordinates chrX:51,760,718-51,776,770 (GRCh37/hg19). The breakpoint localization was validated by FISH using the BAC probe RP11-1129M7. There are no breakpoint neighboring genes that could be a candidate for pigmentary dysplasia. Although we observed a skewed X inactivation pattern in lymphocytes, the hypopigmented skin may present a random X-inactivation pattern causing a functional Xp disomy. Therefore, the X-inactivation pattern study only in lymphocytes is not sufficient to assess the pathogenic mechanism leading to PMI in patients with rearrangements involving X-chromosome. Most descriptions of balanced X-autosome translocation associated with PMI date back to the 80's and 90's, when the rearrangement investigation was limited to low resolution techniques and interpretation of a rearrangement's clinical impact was based upon a broad chromosome region. The report of a patient with PMI and balanced X-autosome translocation involving Xp11.22 with an intergenic breakpoint reduces the strength of the hypothesis of an X-linked loci disruption. Financial support: FAPESP.

2616M

Down syndrome before Lejeune, Gauthier, and Turpin: Historical myths and reality. *E.B. Hook[1,2].* 1) School of Public Health, University of California, Berkeley, CA., USA; 2) Department of Pediatrics, UCSF, San Francisco, CA, USA.

A critical review of the literature reveals that a number of historical aspects of Down syndrome are often misrepresented or misunderstood. i) (J. Langdon) Down's report (1866) of those with the syndrome, which classified some types of "idiocy" as atavistic reversions of Europeans to those of a "lower" race, is usually regarded as an unchallenged example of Victorian racism. But at the very same meeting at which he presented his conclusions, they were criticized forcibly. At the height of the Victorian era there were from the first those who rejected Down's racial interpretation. ii) Even before Down, Seguin (1846) had reported the syndrome, but as a type of "cretinism". And primarily because of the superficial similarity of those with Down syndrome and cretinism, some kind of causal association of hypothyroidism with Down syndrome has been the longest lasting and most prevalent hypothesis until the cytogenetic era. iii) While Waardenburg (1932), Bleyer (1934), and Fanconi (1939) suggested a chromosome abnormality, there was nothing forcing about their hypothesis. Moreover, Mittwoch (1952) could not find a chromosomal abnormality when she searched. iv) Other genetic theories were offered which seemed at the time to explain the etiology just as well. Macklin (1929), an early specialist in medical genetics, proposed 5 separate recessive genes or 2 dominant and 4 recessive, and Penrose 2 dominant genes. v) Some investigators who suggested genetic explanations also posited interactions with non-genetic factors, e.g. Baroff (1958), a year before the report of Lejeune et al., because of both the low average familial proportion of those affected and the maternal age association. vi) Bleyer (1961) attributed his (correct) cytogenetic hypothesis to "deductions...from close clinical study", but as McKusick (1961) implied, many incorrect theories were also derived from close clinical study. Wallin (1949) and Engler (1949) reported several dozen for Down syndrome. McKusick noted many are absurd. Some, like racial theories reflect contemporary prejudices, but others when proposed, appeared plausible and defensible to many based on existing knowledge at the time. vii) Many past outstanding clinicians drew incorrect inferences about the cause of the condition. We cannot judge these past investigators because we lack detailed knowledge of their perspectives and their limits in time and place.

2617T

IMPLEMENTATION OF HIGH-RESOLUTION SNP ARRAYS IN THE INVESTIGATION OF PATIENTS WITH NEURODEVELOPMENTAL DISORDERS: 6 YEARS OF CLINICAL EXPERIENCE. *O. PALUMBO, P. PALUMBO, R. STALLONE, M.P. LEONE, T. PALLADINO, L. ZELANTE, M. CARELLA.* MEDICAL GENETICS UNIT, IRCCS CASA SOLLIEVO DELLA SOFFERENZA, SAN GIOVANNI ROTONDO, FG, Italy.

Genomic copy number variations (CNVs) contributes to the etiology of global developmental delay (DD), intellectual disability (ID), autism spectrum disorders (ASDs), and multiple congenital anomalies (MCAs). Chromosomal microarray analysis (CMA), with a 10-20% diagnostic yield, can identify CNVs≤1 Mb and now is considered the first-tier clinical test for detection of CNVs among patients with developmental disabilities. We report our experience with the use of the Affymetrix SNP Arrays in 1800 Italian patients during the past 6 years (2008-2013). Blood samples were analyzed with CMA at a resolution ranging from 10 kb to 1 kb. CNVs have been considered pathogenic if the variant was responsible for a known syndrome, if involved gene/s reported in association with the clinical phenotypes of the patients, if occurred de novo or, when inherited, if the parent was variably affected. We identified CNVs with a high score of pathogenicity in 486 (27%) patients. Among them 167 (34.4%) showed a CNV overlapping with a known syndrome, 319 (65.6%) a likely pathogenic rearrangement. Of particular interest, we found some CNVs useful to further delineate the clinical features associated with deletions in 8q12.1q12.3, in 15q25.2, in 17q21.31, in 2q24.1q24.2, in 22q11.2, and duplications in 16p13.3 and in 11p13. Some CNVs were useful to describe new syndromes such as a 1.7 Mb deletion in 3q13.2q13.31. Also, we have identified a large group of small CNVs (<1.0 Mb) encompassing, either in whole or in part, functionally related genes to the phenotypes such as CASK, CNTN6, SNTG2, HIP1, DLG2, NRXN1, MCPH1 and CHL1. Among these small CNVs, we have reported a FOXP1 gene microdeletion in an adult patient with autism and speech delay, and a de novo interstitial deletion of 0.122 Mb at 2q24.2 region harboring only TBR1 gene in a boy with moderate to severe intellectual disability. Variants of uncertain significance (VOUS) because unreported, containing genes of uncertain clinical significance or non-genic but potentially regulating nearby gene expression, were identify in 144 individuals (8%). Interestingly, using SNP arrays, we were also able to identify some copy neutral events including uniparental disomy (UPD), run of homozygosity (ROH) and low level mosaicism revealing in some cases important clinical correlations.

2618S

Idic(15) syndrome: clinical studies of 32 new individuals. *A. Battaglia, T. Filippi.* Dev Neurosciences, Stella Maris nst/Univ Pisa, Pisa, Italy.

Chromosome region 15q11q13, known for its instability, is highly susceptible to clinically relevant genomic rearrangements, such as supernumerary marker chromosomes formed by the inverted duplication of proximal chromosome 15. Inv dup(15) results in tetrasomy 15p and partial tetrasomy 15q. The large ones, containing the Prader-Willi/Angelman syndrome critical region (PWS/ASCR), are responsible for the inv dup(15) or idic(15) syndrome. Diagnosis is achieved by standard cytogenetics and FISH analysis, using probes both from proximal chromosome 15 and from the PWS/ASCR. Microsatellite analysis on parental DNA or methylation analysis on the proband DNA, are also needed in order to detect the parent-of-origin of the idic(15) chromosome. Array CGH has been shown to provide a powerful approach to detect the duplication and its extent. Differential diagnosis concerns the possible occurrence of double supernumerary isodicentric chromosomes derived from 15, resulting in partial hexasomy of the maternally inherited PWS/ASCR. Large idic(15) are nearly always sporadic. Antenatal diagnosis is possible. Here, we report on 32 new idic(15) patients; all sporadic. 1/32 had a double idic(15). Phenotypic features were quite variable, and 60% presented with a distinct "autistic-like" profile, providing a behavioral signature for ASD arising from the proximal 15q. Epilepsy occurred in 90%, and in only 20% could be controlled by AEDs. The interictal EEG showed slow/sharp waves, and/or biphasic spikes-polyspikes, spike/wave complexes, and an excess of fast activity over both fronto-temporal areas. Intellectual disability was severe-profound in 85%, with expressive language limited to dysillabic sounds/single words. Structural CNS defects were seen in 30%. Early central hypotonia was present in all. Follow-up ranged from 2 to 24 years.

2619M

Large distal duplication of chromosome 22q. *D. Ortiz, L. Karger, M. Babcock, L. Edelmann, A. Babu, L. Mehta.* Dept. of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

About 25 cases of live-born trisomy 22 have been reported in the literature with characteristic dysmorphic features and major organ anomalies typically resulting in demise soon after birth (Heinrich et al, 2012). Other microduplication syndromes of chromosome 22 include Emanuel syndrome [der(22)t(11;22)], cat-eye syndrome (tri- or tetrasomy 22q11.2) and 22q11.2 microduplication syndrome. We describe a live-born female infant with a large 27.3 Mb duplication of chromosome 22q11.23q13.33 resulting in trisomy 22q11.23→qter. During the pregnancy there was persistent asymmetric IUGR with severe microcephaly (<3rd percentile) and oligohydramnios. Fetal echocardiogram was normal. Amniocentesis was offered and declined. The patient was born at 36.3 weeks gestation and had dysmorphic features, poor tone and poor respiratory effort. Additional features included a secundum atrial septal defect, bilateral hip dislocation, and tracheomalacia. No renal or brain anomalies were found. Karyotype revealed a derivative chromosome 22, 46,XX,der(22)t(22;22)(p12;q12.1), with no duplication of the DiGeorge region TUPLE1 probe but with duplication of the 22q13.3 ARSA probe. Array CGH further defined the duplicated region as 22q11.23q13.33 (22201006-49525123) (NCBI36/hg18). Parental studies were normal. The patient died on day of life 77 due to respiratory failure. Her facial appearance was strikingly similar to patients with complete trisomy 22 including microcephaly, cranial anomalies, hypertelorism, flat nasal bridge, dysplastic ears and shortened neck, and she shared many additional features including IUGR, near-term delivery, genital anomalies, congenital heart lesion and respiratory issues. A patient with a similarly large partial trisomy 22 (23.5 Mb between 22q11.1-22q13.1) has been described, but had findings consistent with cat-eye syndrome, including coloboma, preauricular ear tag, hypospadias and renal anomaly (Karcaaltincaba et al, 2010). Our patient expands the literature on complete and partial trisomy 22, abnormalities with limited viability, and helps clarify the contribution of the distal duplication to this distinctive phenotype and associated malformations.

2620T

Prader-Willi syndrome: Toward Automated Identification of Phenotypic Differences. *L. Wolf[1], C. Clericuzio[2], S.B. Cassidy[3], J.E. Allanson[4].* 1) FDNA, Herzliya, Israel; 2) Department of Pediatrics, University of New Mexico Health Sciences Center, Albuquerque, New Mexico; 3) Division of Medical Genetics, Department of Pediatrics, University of California, San Francisco, San Francisco, California, USA; 4) Hospital of Eastern Ontario, University of Ottawa, Canada.

In Prader-Willi syndrome (PWS), the two common etiologies are paternal deletion of chromosome 15q11-13 (~70% of affected persons) and maternal uniparental disomy15 (UPD) (~25%). Cassidy et al. [Am J Med Genet 68; 433-440, 1997] evaluated 37 individuals with deletion PWS and 17 with UPD to determine if there were phenotypic differences and reported that those with UPD have a higher likelihood of an atypical face, often rounder displaying a broader forehead and a flatter/broader nasal bridge. Allanson et al. [Proc Greenwood Gen Ctr 19: 117, 2000] followed up with an objective study of 109 individuals with PWS. Measurements of eyes, ears, nose, mouth, facial widths, depths, lengths and circumferences were taken. Results showed agreement with the previous study: persons with UPD have a longer face, more protuberant nose with a narrower base, smaller mouth, and broader or more prominent mandible. In this study we revisited the 2000 study to validate the analysis using different statistical machinery, examine the age-related differences between the two etiologies, and evaluate image-based differentiation. 73 persons are included in our retrospective study, 37 under the age of 8 (7 UPD, 30 deletion), and 36 above (15 UPD, 21 deletion). The R Matching package was used to match individuals based on gender and age. The results reveal that, in the younger age group, nasal base and mouth width are significantly reduced in UPD, while in the older, inner canthal distance and ear length are larger in UPD. After normalizing measurements by ofc, these results remain significant. Reduced minimum frontal distance is another differentiator of UPD under the age of 8, while reduced mouth width differentiates above age 8. The successful utilization of normalized measurements supports the future use of ratios of distances extracted from images. Next, facial images of 16 persons (4 UPD, 12 deletion) from the same cohort were evaluated with automatic image analysis tools for UPD/deletion differentiation. This type of analysis provided good discrimination between UPD and deletion individuals. Since UPD is the less prevalent etiology and has a subtly different phenotype, there is a risk of under diagnosis. In a future study we plan to test this hypothesis using blinded dysmorphologist assessment and automatic facial analysis tools. We would assess whether this risk could be reduced by training automatic systems to identify the individual etiologies.

2621S

Wiring the Brain in Down Syndrome: Unique Identical Twins Discordant for DS. *J.R. Korenberg[1], L. Dai[1], J.O. Edgin[3], C. Vachet[2], J. Anderson[4], G. Gerig[2].* 1) Center for Integrated Neuroscience and Human Behavior, Department of Pediatrics, Brain Institute, University of Utah, Salt Lake City, UT; 2) School of Computing, University of Utah, Salt Lake City, UT; 3) Department of Psychiatry, University of Arizona, Tucson, AZ; 4) Department of Neuroradiology, University of Utah, Salt Lake City, UT.

Genetic disorders as Down syndrome (DS) provide a solution to the challenge of human neuroscience, to develop a unified view of the brain across spatial, temporal and functional levels integrated with emergent properties such as cognition. However, subtle brain and cognitive phenotypes are obscured by genome-wide variation. We report a multimodal analysis including genetics, cognition, structural and functional MRI of the only living pair of Identical male twins, age 22, discordant for DS. These resulted from an extra chromosome 21 in an otherwise identical genome, with the remainder of variance contributed by environmental and stochastic effects. We established MZ identity by Illumina genomic arrays, and established typical clinical and chromosomal features of DS in one, >96% trisomy 21 in cord blood, blood and fibroblasts, and in the other a normalized phenotype, >92% normal karyotype in these tissues. The twins completed the Arizona Cognitive Test Battery for DS, an assessment including neuropsychological measures of hippocampal and prefrontal function. The 8% mosaic twin had a full-scale IQ of 123, typical performance on a frontal measure (CANTAB IDED), but impaired spatial working and associative memory (CANTAB Spatial Span and Paired Associates Learning (PAL)), markers of hippocampal dysfunction. Early language delay was noted; at baseline testing he presented with a 24 point split in verbal and performance IQ. The 96% mosaic twin showed scores typical for the DS range (full scale IQ=42), with prominent impairments in (PAL). The twins were also scanned on high resolution MRI/DTI and compared with 22 typical DS and 17 controls. Overall, normalized volumes of the 96% mosaic twin were very close to the volumes of the DS group, and the 8% mosaic twin volumes were similar to controls. An exception was DS-like volume in the 8% mosaic twin of the right fusiform gyrus, involved in visual memory and language. Integration of MRI with cognition in full trisomy, revealed correlation of hippocampal volumes with PAL results. Only the 96% mosaic twin demonstrated increased internetwork synchrony across networks seen in typical DS during functional MRI. The twins are a graphic vision of how a simply trisomy for 21 disturbs brain wiring and the results implicate the hippocampal and fusiform gyrus formation in the neural systems responsible for visual-spatial (PAL) and linguistic deficits of DS.

2622M

Congenital Myasthenia Syndrome: Uniparental disomy of chromosome 2 and homozygous mutation of GFPT1. *S. Rangasamy[1,2], R. Richholt[1,2], K.M. Ramsey[1,2], A. Siniard[1,2], J.J. Corneveaux[1,2], I. Schrauwen[1,2,3], J. Krate[1,4], A. Kurdoglu[1,2], M. De Both[1,2], S. Szelinger[1,2,5], B.E. Hjelm[1,2], S. Swaminathan[1,2], M. Huentelman[1,2], D. Craig[1,2], V. Narayanan[1,2].* 1) Neurogenomics Division, Translational Genomics Research Institute, Phoenix AZ, USA; 2) Dorrance Center for Rare Childhood Disorders, Translational Genomics Research Institute, Phoenix AZ, USA; 3) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; 4) University of Arizona College of Medicine, Tucson, AZ; 5) Molecular and Cellular Biology Interdisciplinary Graduate Program, School of Life Sciences, Arizona State University, Tempe AZ, USA.

Congenital myasthenia syndrome (CMS) includes a group of genetic disorders involving genes encoding components of the neuromuscular junction. We describe an unusual case of a child with CMS who was found to have long stretches of homozygosity on chromosome 2. The female child was born full term but required emergency C-section, followed by prolonged intubation. She was later diagnosed with hypotonia, weakness with reduced spontaneous movement, and was areflexic. Genetic testing for spinal muscular atrophy, Prader-Willi syndrome, and congenital myotonic dystrophy were negative. Muscle biopsy study at 1 month of age showed normal fiber type size and grouping, but there was coarse staining of the T-tubule system suggestive of a mitochondrial disease with reduced cytochrome C oxidase levels. EMG showed an electrodecremental response, and she had a partial response to treatment with pydostmine. Gene testing for SURF1, SCO1, SCO2, and CO×10 were negative and the chromosomal microarray test showed normal gene dosage. Chromosomal microarray identified a long (57.6 Mb and 90 Mb) contiguous regions of homozygosity restricted to chromosome 2: arr 2p21q11.2 (45,013,724 - 102,641,201) hmz; 2q22.1q36.3 (140,091,291 - 229,908,245) hmz. The Congenital Myasthenia Group at the Mayo Clinic confirmed a CMS diagnosis. At the time, CMS gene panel was negative. Family trio based whole-exome sequencing (WES) was performed at the Dorrance Center for Rare Childhood Disorders as a part of research study. Exome data identified a homozygous, frameshift, c.865-866insG insertion on chr2, at exon 9 of GFPT1 gene; a low coverage region by WES. Insertion was confirmed by Sanger sequencing which demonstrated that father was homozygous reference and mother was heterozygous. Based on our findings, we propose that this child had two identical (maternal) copies of large segments of chromosome 2, implying maternal uniparental disomy, harboring the GFPT1 insertion that causes autosomal recessive congenital myasthenia. Previously reported patients with recessive GFPT1 mutation show reduced GFPT1 levels and decreased protein glycosylation. Glycosylation at neuromuscular junctions is important for key intracellular signaling protein function and compromised glycosylation in GFPT1 mutation leads to altered neuromuscular phenotype. The precise pathogenic mechanism for our observed GFPT1 insertion will be determined with future studies.

2623T

CRANEOFACILALES AND FEATURES ASPECTS ORAL OF CHILDREN WITH THE POBLANO CHILD HOSPITAL GOLDENHAR SYNDROME IN THE PERIOD 2013-2015. *D. Reyes Ramirez[1,2], R. Aparicio[1,2], G. Muñoz Quintana[1].* 1) Pediatric Estomatology, Benemerita Universidad de Puebla, Puebla, Puebla; 2) Genetics, Hospital para el niño Poblano, Mexico.

The facial skull malformations and chromosomal aberrations are considered alterations in the phenotypic and secondary structure error of the birth. All these changes must be observed during childhood and confirmed by chromosomal karyotype or deficiency in DNA repair. Pediatric patients were evaluated and showed different congenital and chromosomal alterations. These changes at the level of its multiple system and organic structure, chromosomes were analyzed and classified as numeric and structural alterations respectively. Another group of genetic disorders known as mutations and they are inherited in different generations. A wide range of pediatric patients with congenital and genetic diseases by alterations mainly associated with craniofacial development during pregnancy, prenatal, or at birth are described in this study to analyze characteristics clinical, medical, or surgical procedures and medical evolution according to the syndrome of malformation in study. The wide variety of Craniofacial anomalies often makes them unclassifiable. Goldenhar, also known as the first Syndrome and is second Gill arch or oculo-auriculo-vertebral spectrum, is a complex of Craniofacial and vertebral anomalies. These findings can be found in solitary or usually associated with microtia, mandibular hypoplasia, or congenital vertebral malformations. The incidence is limited and varies between 1 in 45,000 to 2 in 100,000 inhabitants. Currently it should be considered as a bilateral malformation, which would differentiate it from the Microsomía Hemifacial. Know the relationship of oral diseases in children with Goldenhar Syndrome, as well as craniofacial features will give us a broad overview in order to provide better care odonto/stomatological pediatric patients. The earlier and excellent medical and surgical treatment is important to improve their life quality.

2624S

Intragenic *CNTN4* Deletion in a Child with Profound Speech Apraxia and Absence of Autistic Features. *J. Fleischer[1], H. Al-Kateb[2], M. Shinawi[1].* 1) Genetics, Washington University, St. Louis, MO; 2) Department of Pathology and Immunology , Washington University School of Medicine, St. Louis, MO.

Background: Contactin 4, a cell adhesion molecule and a member of the immunoglobulin super family, is expressed primarily in the nervous system. Deletions and duplications encompassing the *CNTN4* gene have been reported in children with autistic spectrum disorder and developmental delay. Intrestengly, all previously documented cases were inherited form unaffected fathers raising questions regarding reduced penetrence and pathogenenicity of these rearrangements. Methods: Clinical and molecular cytogenetic characterization studies were performed on a 2-year-old boy who presented with severe speech apraxia, mild gross motor delay and absence of autistic features. The findings in the proband are compared with previously reported cases. Results: The proband is non-verbal, but displays age appropriate receptive speech skills, cooperative play, joint attention, and eye contact. Additional findings include intrauterine growth retardation, premature birth, a large secundum atrial septal defect, and mild facial dysmorphisms. Chromosomal microarray analysis revealed a 128Kb deletion within the *CNTN4* gene corresponding to the coordinates chr3: 2,878,363-3,006,381 (hg19). The deletion takes out exons 7-12 and introduces a frame-shift mutation at amino acid S154 resulting in a termination codon 22 amino acids downstream the deletion site. Two splice variants (4 and 5) are not affected by this deletion. Fluorescence in-situ hybridization studies revealed a paternal origin of the deletion. Conclusion: Although the *CTNT4* intragenic deletion in our proband is predicted to result in loss of *CNTN4* function, the child did not exhibit autistic features, the hallmark of *CNTN4* microdeletion-associated phenotype. Our data provide additional support for the significance of *CNTN4* gene in language development and emphasize the phenotypic heterogeneity. We discuss potential factors contributing to reduced penetrence and attempt to find possible phenotype genotype correlation.

2625M

A 284kb microduplication at 7q21.11 involving SEMA3A possibly linked to agenesis of corpus callosum and visual impairment in a patient with spastic quadriparetic cerebral palsy. *D. Ma[1], N. Singh-Bhatia[2], P. Levy[2], R. Marion[2], R. Naeem[1].* 1) Pathology, Montefiore medical center, Bronx, NY; 2) Genetics, Children Hospital at Montefiore, Bronx, NY.

A 284kb heterozygous duplication was identified by array comparative genome hybridization (aCGH) on chromosome 7q21.11 in a patient with spastic quadriparetic cerebral palsy, cerebral dysgenesis including agenesis of corpus callosum, visual impairment and severe global developmental delay. This duplicated region, involving the regulatory elements of SEMA3A along with its exon 1 and the transcriptional start point, has not been documented in any publicly available case or control databases. Within the family, another two brothers and one sister also presented with cerebral palsy and global developmental delay, suggesting a candidate gene in a possibly autosomal dominant inheritance mode with variable expressivity and penetrance. An in depth examination of the SNP data revealed five loss of heterozygosity (LOH) regions on 2q23.2-q24.1, 3q25.2-q26.31, 6q14.2-q14.3, 6q22.1-q22.31 and 8q11.21-q11.23 with various lengths from 3.1Mb to 16.8Mb. A search for potential disease-causing genes in a recessive inheritance mode did not reveal obvious candidates that could explain the clinical features. A literature review of SEMA3A indicated a high expression of the gene in all adult and fetal brain regions and heart as well as fetal skeletal muscle. Gene expression has been associated with the growth of the human corpus callosum and the effect on retinal pigment epithelial cell activity, which is quite well matched with the multiple congenital anomalies in this patient. In addition, a recent case report has linked the deletion of the similar region with a novel type of syndromic short stature. Though the gene dosage of SEMA3A appears to be critical for the normal development of brain, the functional impact of the duplication on its regulation of gene expression remains to be elucidated.

2626T

Craniosynostosis in Williams Syndrome. *N. Okamoto, K. Ueda.* Dept Medical Genetics, Osaka Med Ctr/Res Inst, Osaka, Japan.

Williams syndrome (WS) is caused by a microdeletion on chromosome7q11.23 encompassing the elastin gene. WS is characterized by distinctive facies, congenital cardiovascular malformations, connective tissue abnormalities and intellectual disability. Craniosynostosis has seldom been described as a complication of WS. We identified seven patients with craniosynostosis. They showed abnormal cranial morphology, prominent microcephaly or headache. Four had sagittal craniosynostosis (dolichocephaly), one had metopic craniosynostosis (trigonocephaly), and two had both. Two of them underwent cranioplastic surgery. Patients with WS often show hyperactivity together with developmental delay. If craniosynostosis remains undiagnosed, disabilities may be worse. Increased intracranial pressure may modify behavioral abnormalities. Therefore, proper diagnosis and neurosurgical intervention are necessary. If patients with WS show prominent microcephaly or cranial deformation including dolichocephaly or trigonocephaly, neuroradiological evaluation including 3D-CT of the cranium is necessary. We suggest that craniosynostosis is an important complication of WS. Craniosynostosis should be considered in every patient with WS and cranial deformation.

2627S

Neurodevelopmental and neurobehavioral characteristics in males and females with CDKL5 duplications. *P. Szafranski[1], S. Golla[2], W. Jin[1], P. Fang[1], P. Hixson[1], R. Matalon[3], D. Kinney[4], H-g. Bock[5], W. Craigen[1], P. Magoulas[1], J.L. Smith[1], W. Bi[1], A. Patel[1], S.W. Cheung[1], C. Bacino[1], P. Stankiewicz[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Departments of Pediatrics and Neurology, University of Texas Southwestern Medical Center, Dallas, TX; 3) Division of General Academic Pediatrics, Department of Pediatrics, The University of Texas Medical Branch at Galveston, Galveston, TX; 4) Memorial Children's Hospital Navarre Pediatrics South Bend, IN;; 5) Department of Pediatrics, University of Mississippi Medical Center, Jackson, MS.

Point mutations and genomic deletions of the *CDKL5* (*STK9*) gene on Xp22 have been reported in patients with severe neurodevelopmental abnormalities, including Rett-like disorders characterized by early-onset seizures, infantile spasms, severe intellectual disability with absent speech, and microcephaly. The phenotypic resemblance to Rett syndrome has been proposed to result from similar functions or interactions in the neuronal molecular pathways between CDKL5 and MECP2. In contrast to *MECP2*, the clinical consequences of increased dosage of *CDKL5* are poorly understood. To date, only larger-sized (8-21 Mb) duplications harboring *CDKL5* have been described. We report six females and four males from seven unrelated families with *CDKL5* duplications ranging in size between 540 kb and 935 kb. Three families of different ethnicities had identical 665,915 bp duplications of complex genomic structure containing only the shorter *CDKL5* isoform expressed at low levels in the fetal brain but not the adult brain, indicating that it had likely arisen many generations ago. Four affected boys, ages 9-14, and three affected girls, ages 8-9, manifested autistic behavior, developmental delay, language impairment, and hyperactivity. Of note, two girls and one boy had macrocephaly. Two carrier mothers of the affected boys reported a history of problems with learning and mathematics while at school. None of the patients had epilepsy. Consistent with *CDKL5* mutations and deletions, the X-inactivation pattern in all six studied females was random. We propose that the increased dosage of *CDKL5* might have affected its interactions with MECP2, leading to perturbation of synaptic plasticity and learning; manifestations of this perturbation include autistic behavior, developmental and speech delay, hyperactivity, and macrocephaly.

2628M

Subtelomeric investigation in individuals with microform of HPE. *L. Ribeiro-Bicudo[1,2,3], V. do Ó[2], B. Gamba[3], A. Richieri-Costa[2].* 1) Genetics, Goias Federal University, Goiania, Goias, Brazil; 2) Hospital for Rehabilitation of Craniofacial Anomalies, Bauru, Sao Paulo, Brazil; 3) Genetics, Sao Paulo State University, Botucatu, Sao Paulo, Brazil.

Holoprosencephaly (HPE) is a malformation sequence where the cerebral hemispheres fail to separate into distinct left and right haves. Three levels of increasing severity are described in HPE: lobar, semi-lobar and alobar. Another milder subtype of HPE called the middle interhemispheric variant (MIHF) or syntelencephaly, has been recognized. A new phenotype was first delineated in a series of Brazilian patients, the so-called holoprosencephaly-like (HPE-L) phenotype. The etiologies and the variable clinical spectra of HPE and HPE-L, also called HPE minor form or microform, seem to be related and extremely heterogeneous. Data about frequencies of classic HPE are well known. It is considered one of the most common human malformation, with an estimated prevalence of at least 1/16,000 live births and in 1/250 conceptuses, and it is usually classified according to type of CNS involvement or facial anomalies observed. On the other hand, similar statistics concerning HPE-L and/or microforms are unknown, but certainly it represents a challenge in several areas, such as clinical findings, differential diagnosis, prognosis, and genetic counseling. HPE-L can be seen as a condition important to the understanding of many of the mechanisms involved in normal and abnormal craniofacial development. Attention should be given to newborn babies with isolated large cleft lip/palate where the definitive diagnosis only will be possible with the phenotypic evolution. It is estimated that the cause of HPE is due to cytogenetic anomalies in 30-50% of indivuduals, to their environmental causes and unknown genetic alterations in 10-15%, and to mutations in HPE genes in 5-10%. The present work was based on the observation of previous data showing subtelomeric aberrations in HPE individuals. We selected individuals presenting the microform of HPE with normal MRI and no developmental delay. These individuals had facial characteristics such as hypotelorism, single central incisor, flat nose, and cleft lip/palate. We performed MLPA screening based on the P036 kit on 64 DNA samples. No subtelomeric aberrations was found in the samples analyzed. Even considering that the sample is small, it seems that the subtelomeric aberration is not a common event in the microform of HPE.

2629T

Discontinuous microdeletion at 1p13.3 involving NBPF4 but not ALX3 in a patient with severe frontonasal dysplasia. *A. Ali[1], R. Patel[1], S. Jaiswal[1], S. Singh[2], R. Raman[1,3].* 1) Centre for Genetic Disorders, Banaras Hindu University, Varanasi, Uttar Pradesh, India; 2) GS Memorial Plastic Surgery Hospital Hospital, Varanasi; 3) Department of Zoology, Banaras Hindu University, Varanasi, Uttar Pradesh, India.

Frontonasal dysplasia (FND) arises due to improper embryonic development of head and face. It is defined by at least two of the following features; hypertelorism, broad nose, clefts of nose, absent nasal tip, cleft lip and/or cleft palate, anterior cranium bifidumoccultum, or a widow's peak hairline. Three types of frontonasal dysplasia are reported; type-1 and type-2 which are correlated with mutations in ALX3 (1p13.3) and ALX4 (11p11.2) genes respectively, whereas type-3 is correlated with mutations in ALX1 gene (12q21.31). Here, we report a sporadic case of FND with some overlapping features of all the three types of FND, characterized by hypertelorbitism, absent cupid bow, cleft of bilateral nostrils, lacking midline bony nasal septum, deviation of nasal septum to right side, frontal bossing with depression in centre, left eye anophthalmia and right eye micropthalmia, left orbit without lens, bilateral epicanthal folds, overjet teeth, flexion deformity of little finger of left of left hand and unilateral cryptorchidism (left side). Cytogenetic microarray analysis using Affymetrix CytoScan® HD array was performed for the patient, both the parents, and 7 unaffected controls. A deletion of 125 kb region at locus 1p13.3 was detected in the patient only and none of the parents and unaffected controls. The deleted region contains two genes SLC25A24 and NBPF4. SLC25A24 encodes a carrier protein that transports ATP-Mg which exchanges it for phosphate. NBPF4 is a neuroblastoma breakpoint family gene known to be associated with a number of developmental and neurogenetic diseases such as microcephaly, macrocephaly, autism, schizophrenia, mental retardation, congenital heart disease, neuroblastoma, and congenital kidney and urinary tract anomalies. ALX3, a known candidate gene for FND, is close to the deleted region but is in normal copies. This study suggests, deletion of NBPF4 as the cause of frotonasaldysplaia in the patient due to the following two reasons. One, NBPF4 is known to cause variable phenotypes including cranial abnormalities. Second, hypermetyhylation of ALX3, a known candidate for FND, cause neuroblastoma, and since NBPF4 is also known to cause neuroblastoma, we hypothesize that ALX3 probably regulates NBPF4 expression and therefore ALX3 mutation cause FND through downregulation of NBPF4. Therefore haploinsufficiency of NBPF4 causes FND. Investigation is underway to further confirm the role of NBPF4 in FND.

2630S

Growth Standards for Children and Adolescents with Smith-Magenis Syndrome. *L.R. Fleming[1], F. Duncan[2], W. Introne[2], J. McGready[3,5], K. Schulze[4,5], A. Alade[5,6], J. Hoover-Fong[5,6], A.C.M. Smith[2].* 1) Medical Genetics Branch, NHGRI/NIH, Bethesda, MD; 2) Office of the Clinical Director, NHGRI/NIH, Bethesda, MD; 3) Dept of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 4) Center for Human Nutrition, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 5) Greenberg Center for Skeletal Dysplasia, Johns Hopkins University, Baltimore, MD; 6) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD.

Background: Smith-Magenis syndrome (SMS) is a complex genetic disorder, with an incidence of 1/15,000, characterized by distinctive craniofacial features, cognitive impairment, sleep disturbances, and neurobehavioral abnormalities. Poor linear growth and weight gain are common in infancy and short stature has been reported in 50-78% of individuals. However SMS-specific growth curves have not been available previously to allow clinicians to assess growth. Methods: Anthropometric data (height, weight and head circumference) were collected by clinician measurement, review of the medical record and parent submission for 143 individuals (82 Females, 61 Males) with SMS in Europe, the USA and Australia. Of these, 138 (96.5%) had deletions and 5 (3.5%) had point mutations in *RAI1*. Penalized cubic smoothing splines were used to estimate SMS curves for height and weight for periods of 0-36 months and 2-12 years. These curves were compared to WHO (0-36 month) and CDC (2-18 year) norms. Results: At birth, the majority of term infants with SMS are within the clinically normal range for weight and length. Infants with SMS exhibit a decline in height velocity, with the median height tracking < 5th centile of average stature by 3 years of age. Females with SMS show a more gradual decline in height velocity than males. Relative short stature persists through early adolescence, with median height for individuals with SMS through age 12 near the 25th centile for sex specific CDC age norms. By 6 months of age there is substantial attenuation of weight velocity in SMS males. A more gradual attenuation in weight velocity over the first year of life is seen in SMS females. Weight curves for 2-12 years of age show accelerated weight gain beginning in late childhood (~7 yrs). In comparison to CDC norms, the median weight in SMS males and females falls above the CDC 50th centile through 12 years of age. Discussion: This study represents the first set of standardized growth curves created for individuals with SMS. Syndrome specific growth curves can be used to set expectations for growth and to manage syndrome related symptoms. These curves can also be used in future research to better understand the natural history of SMS and the genotype-phenotype relationships in these patients.

2631M

Phenotype of double de-novo Williams and DiGeorge microdeletion syndromes. *A. Shukla[1], K. Mandal[2], SJ. Patil[3], SR. Phadke[2], KM. Girisha[1].* 1) Medical Genetics, Kasturba Medical College, Manipal, Karnataka, India; 2) Department of Medical Genetics, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India; 3) Naryana Hrudayalaya Institute of Medical Sciences, Bangalore, Karnataka, India.

DiGeorge syndrome (DGS, chromosome 22q11.2 deletion syndrome, OMIM 188400) and Williams syndrome (WS, chromosome 7q11.23 deletion syndrome, OMIM 194050) are well characterized microdeletion syndromes. Here, we report a child with de novo deletions in both these regions with the resulting unique phenotype that has not been reported till date in the literature. This 6-years-old child presented with speech delay and history of persistent hypocalcemia with seizures in the neonatal period. He is born to non-consanguineous healthy parents and has a healthy younger sister. The development was markedly delayed with independent walking achieved at 3 years of age. He does not speak even a single word though is able to communicate his basic needs non-verbally. He is yet to achieve bowel and bladder control. He has severe growth retardation with microcephaly. The height is 98 cm (-4 SD below the mean) and OFC is 46.5 cm (-5 SD below the mean). The craniofacial features are a mixture of both syndromes. He has the typical features of full lower cheeks with malar flattening and thick vermillion as in Williams syndrome but the nose and ears are similar to those of DGS. The hooding of the upper eyelids too resembles that of DGSS. Surprisingly, though congenital heart defect is a major feature of both these syndromes, he has a normal echocardiography. The behavioral phenotype matches more with that of DGS and he lacks the sociable behavior of the Williams syndrome. The persistent hypocalcemia can be attributed to the 22q11 deletion. Interestingly, he was clinically diagnosed to have Williams syndrome by one clinical geneticist and DiGeorge syndrome by the other. Detection of microdeletion was done using Multiplex Ligation Dependent Probe Amplification (MLPA) technique (MRC Holland, Amsterdam) using Probe mix P245-Microdeletion-1. Deletion was noted at 22q11.21 (all 3 probe sets in this probe mix for that region) and also at 7q11.23 (both of 2 probe sets in this probe mix for that region) suggesting co-existing DiGeorge syndrome and Williams syndrome. Either parent lacks these deletions. The case illustrates chance occurrence of two common microdeletions at these two regions and the resulting combined phenotype with overlapping features of these two syndromes.

2632T

One cannot presume siblings with similar clinical findings result from the same underlying genetic cause!—familial genome instability, leaky proof reading mechanism, or true diagnosis? *A. Tsai, K. kovak, K. keller, J. kushner, C. rogers, S. Moore.* Molecular and Medical genetics, OHSU, Portland, OR.

When siblings present with similar clinical findings, we usually presume they result from the same etiology. It is not uncommon for a clinical geneticist to order all testing on one child in the family, and subsequently offer siblings specific testing for the revealed genetic disorder. For a cytogenetic anomaly identified by microarray, a reference lab often suggests a FISH study for parents or at-risk individuals in the family rather than a full microarray. We present three families in which multiple children within the family are affected with similar clinical features. Family 1 is fraternal twins. They were born SGA with similar dysmorphic features. A 180k oligo array revealed a copy gain of 1.94 Mb: 5q35.2q35.3 (175,491,045-177,427,809)x3 in the female twin, which includes the entire NSD1 gene (deleted in Sotos syndrome). Her result and features were consistent with the previously reported reverse-Sotos phenotype. A maternal half brother has a similar gain of 1.98 Mb: 5q35.2-q35.3 (175,491,045-177,470,488)x3 and similar features. The male twin's array showed a single copy gain of 435.7 kb (maximum size 458 kb) in 5q35.2q35.3 (176,700,522-177,136,262)x3 that resides within/overlaps with a portion of the sister's, is predicted to disrupt several exons of the NSD1 gene and present with Sotos syndrome. Family 2 is fraternal twins, both with autism and small optic nerves. Microarray in one twin revealed a distal deletion 22q11.21 (19,408,946-19,790,658) x1 and the brother showed a 16p13.11 duplication (15126709-16292235) x3, 1.166-2.133 Mb. Family 3 is 2 maternal half siblings with dysmorphic features and intellectual disability. The brother has a 969.6 kb duplication at 1p13.3, and the sister has a 598 kb deletion at 16p11.2 consistent with the 16p deletion syndrome and the 969.6 kb duplication. Siblings only share 25-50% of genes that are subject to recombination, individual epigenetic exposures and stochastic events. One cannot presume siblings with similar clinical findings result from the same underlying genetic cause. Family 1 shows co-existence of 2 different CNVs from different parental origins. Families 2 and 3 reveal different CNVs and, if only FISH were offered, the anomalies in siblings of the proband would have been missed. CNV can be a red herring. Further investigation is required if the phenotype cannot be explained by known microarray anomalies. Appropriate genetic causes should be considered individually, especially in half siblings.

2633S

An interstitial microduplication in 17q11.2 encompassing the NF1 gene, in a girl presenting with a Prader Willi like syndrome: expanding the NF1 microduplication. *C. Vinkler[1], T. Lerman-Sagie[2,3], D. Lev[1,2,3].* 1) Inst Med Gen, Wolfson Med Ctr, Holon, Israel; 2) Metabolic Neurogenetics Service, Wolfson Medical Center, Holon Israel; 3) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

Interstitial microduplication of chromosome 17q11.2 is rare condition with less than 20 reported cases. In most cases described so far, the microduplications include the NF1 gene. Previous reports trying to delineate the phenotype classified the patients according to the size and breakpoints of the microduplication. Type 1 with 1.4 Mb duplication, type 2 with 1.2 Mb duplication and type 3 with 1Mb duplication. The clinical phenotype includes developmental delay, facial dysmorphism and sometimes seizures. Eating disorder with severe obesity has not been described in this group of patients. We present a 14y old girl who was followed in our clinic for several years. She had dysmorphic features, obesity, hypotonia, global developmental delay and atrioseptal heart defect. Prader-Willi syndrome was ruled out by molecular tests. Family history is unremarkable. At 14y of age she is still overweighed (due to eating disorder) intellectually disabled and has verbal dyspraxia. Her dysmorphic features include round face, short and narrow forehead, small orbital fissures, broad and sparse eyebrows, short and small nose, widely spaced inverted nipples, camptodactyly and clinodactyly of the 5th finger bilaterally. Brain MRI is unremarkable except for a mild enlargement of the left atrium. CMA test showed a de novo 8.746Mb duplication on chromosome 17q11.2 encompassing the NF1 gene along with other 14 OMIM morbid genes. None of the genes has been associated with eating disorders. Obesity has been mentioned in two other cases of 17q11.2 duplication. We propose that CMA test should be considered in cases of eating disorders. More cases might allow to define the common critical area of eating disorder and perhaps help in understanding this abnormality.

2634M

Bilateral absence of the ulna in 4q terminal deletion syndrome. *M. Marble[1], T. Meaux[2].* 1) Dept Ped/Clinical Gen/LSUHSC, Children's Hosp New Orleans, New Orleans, LA; 2) LSUHSC.

We report a patient born with bilateral absence of the ulna who was found to have a terminal deletion involving the long arm of chromosome 4. The deletion was identified by chromosome microarray analysis and determined to be 21.98 Mb in size and to involve 4q32.3-4qter. The molecular coordinates encompassed by the deletion were 4:169,154,141-191,133,858. The patient was born at 38 weeks gestation by spontaneous vaginal delivery. The pregnancy was complicated by intrauterine growth retardation, oligohydramnios, and maternal hepatitis C. Birth weight was 3000 grams. Other clinical features included atrial septal defect, patent ductus arteriosus, micrognathia, cleft palate, inner epicanthal folds and long philtrum. At 2 years of age, weight was 10.6 kg (10-25th percentile) and height was 80 cm (5th percentile). She had delay in language skills with an approximately 15 word vocabulary and no sentences. She could sit on her own and could pull up on furniture but was not yet able to walk. The atrial septal defect and patent ductus arteriosus closed spontaneously. Radiographs showed a short humerus and absent ulna bilaterally. On the upper extremities there was only a single digit. The digit was noted to have a hypoplastic metacarpal, a small well-developed proximal and distal phalanx, but apparent absence of the mid phalanx. Past reports have suggested an association between chromosome 4q deletions and variable ulnar ray defects but most of these reports used traditional cytogenetic methods. We found a few cases in the literature with absent ulna associated with 4q terminal deletions but very few for which comparative genomic hybridization was documented. We compared the deletion in our patient to previous reports in the literature. The findings in our patient, along with other reports, suggest that the terminal 4q chromosomal segment contains a region important for ulnar ray development.

2635T

Congenital heart disease and Sturge-Weber syndrome in a young female with 22q11.2 triplication. *L. Mota-Vieira[1,2,3], L.M. Pires[4], S. Vaz[5], P. Maciel[5], I.M. Careira[4], R. Pires[1,2].* 1) Molecular Genetics and Pathology Unit, Hospital of Divino Espirito Santo of Ponta Delgada, Azores Islands, Portugal; 2) Centre for Biodiversity, Functional and Integrative Genomics (BioFIG), Faculty of Sciences, University of Lisbon, Portugal; 3) Instituto Gulbenkian de Ciência, Oeiras, Portugal; 4) Laboratory of Cytogenetics and Genomics, Faculty of Medicine, University of Coimbra, Portugal; 5) Department of Pediatrics, Hospital of Divino Espirito Santo of Ponta Delgada, Azores Islands, Portugal.

In the scope of the population-based epidemiological study on congenital heart disease (CHD) performed in the São Miguel Island (Azores archipelago, Portugal), we are investigating the genetic basis of these complex diseases which are a major cause of serious morbidity and mortality in childhood. Since 22q11.2 deletions and duplications are frequently implicated in CHD, we carried out a systematic analysis of this genomic region in CHD patients from this Atlantic Island. The analysis consisted of clinical evaluation and searching for copy number variations (CNVs) by Multiplex Ligation-dependent Probe Amplification (MLPA), according to MRC-Holland protocol. The MLPA results detected in a female patient higher doses than 22q11.2 duplication, suggesting the presence of a double disomy (2:2, *i.e.*, two chromosomes with 22q11.2 duplication) or a tetrasomy (1:3, *i.e.*, one normal chromosome and a 22q11.2 triplication in the other). These two hypotheses were investigated by FISH analysis which confirmed a tetrasomy. Further, we screened parental samples, by both techniques, in order to ascertain if the 22q11.2 triplication was inherited or acquired *de novo*. The results showed that her non-affected father presented 22q11.2 duplication (1:2). Therefore, the triplication here described could be explained by an unequal crossing-over of 22q11.2 region that triggered an expansion of the duplication to a tandem triplication during spermatogenesis, in a non-allelic homologous recombination mechanism mediated by the presence of low-copy repeats. To the best of our knowledge, the present study reports the second documented case of 22q11.2 triplication, being the first one described by Yobb and collaborators (2005). Our patient displays the dysmorphic facial features, cognitive deficit and heart defect (restrictive interventricular communication and membranous sub-aortic stenosis) usually seen in patients with 22q11.2 deletion and microduplication syndromes, as well as phenotypic characteristics of Sturge-Weber syndrome (in particular the presence of nevus flammeus — port-wine stain —, behavioral disorders, and visual fields defects). Taking this into consideration, we propose that 22q11.2 triplication is a variation of 22q11.2 microduplication syndrome, with aggravated phenotype due to the major dosage of implicated genes.

2636S

Familial 17q12 duplication presented as SGA/IUGR and microcephaly during pregnancy: A counseling dilemma. *A. Singer[1], I. Maya[2], C. Vinkler[3].* 1) Dept Clinical Gen Unit, Barzilai Med Ctr, Tel Aviv, Israel; 2) Institute of Medical Genetics, "Rabin" Petah-Tikva, Israel; 3) Institute of Medical Genetics,"Wolfson" Holon, Israel.

Recurrent genomic rearrangements of chromosome region 17q12, ranging from 0.3 to 2.1 Mb, have been described to be associated with different clinical phenotypes. Patients carrying a 17q12 duplication present with intellectual disability, developmental delay of various degree, epilepsy, schizophrenia, autisim, brain abnormalities, esophageal atresia and various renal and urinary tract abnormalities. In the normal population, duplication of 17q12 is quiet rare (<0.02%) and it penetrance estimate is around 21%. Hence, prenatal counseling poses a dilemma. We present a baby boy with prenatal and postnatal microcephaly and IUGR/SGA. He was born to a healthy nonconsanguineous couple. Late amniocentesis (32 week) because of symmetric SGA/IUGR, microcephaly and echogenic intracardiac focus found 1.4 MB 17q12 duplication. Parents' CMA showed that it was inherited from his father. During counseling the father was found to have normal HC but he has dyslexia and some speech disturbances. Counseling in this case posed a dilemma since 17q12 duplication is known to have a variable phenotype and incomplete penetrance. The parents decided to continue with the pregnancy. The boy was born at term. Examination was normal except for microcephaly (HC< -2Sd). Prenatal microcephaly has not been previously described in 17q12 duplication syndrome. The clinical significance of this prenatal finding adds to the counseling dilemma.

2637M

Atypical 22q11.2 deletion at the distal end of the common 3Mb deletion. *N. Bhatia[1], T. Goldwaser[2], P. Reingold[2], S. Klugman[2], P. Levy[1], B. Morrow[3].* 1) Children's Hospital at Montefiore, Bronx, NY., USA; 2) Montefiore Medical Center, Bronx, NY, USA; 3) Albert Einstein College of Medicine, Bronx, NY, USA.

Background: 22q11.2 deletion syndrome is a contiguous gene deletion characterized by congenital heart disease, palatal defects, immunodeficiency, hypocalcemia, renal anomalies, developmental delay and facial dysmorphisms. Up to 64%; of individuals with 22q11.2 deletion syndrome exhibit psychiatric disorders, including bipolar disorder and schizophrenia, with wide phenotypic variability. Deletions in the 22q11.2 region usually include a common 3Mb deletion (~90%) or a proximally nested 1.5Mb deletion (~7%). Other atypical deletions include distally nested deletions and deletions distinct from the 3Mb region. The 3Mb and 1.5Mb deletions both include the *TBX1* gene, associated with the congenital heart disease seen in this syndrome.

Case: A 20-week primigravid was referred to our prenatal genetics department for counseling after the fetus was noted to have a Dandy-Walker variant on 16 week anatomy scan. The proband was the product of a nonconsanguineous union between a 23 year old mother and 37 year old father. The mother has a history of bipolar disorder, ADD, OCD, and developmental delay, diagnosed in adolescence. Amniocentesis was performed and arrayCGH revealed a 0.369Mb maternally-inherited deletion in chromosome 22, specifically, arr [hg19] 22q11.21 (21081260-21449911) x 1 mat. At birth, FISH with cosmid probes Tuple1 and N25 were normal. Brain MRI noted a minimally prominent cisterna magna. The proband presented to our pediatric genetics department at 13 months of age. Physical exam was normal and the child was developmentally appropriate with no medical issues. Postnatal arrayCGH confirmed the prenatal finding. Simultaneously, the mother presented to our prenatal department for her second pregnancy, desiring an amniocentesis after anatomy scan revealed a polycystic kidney. This fetus was also found to carry the same familial 22q11.2 deletion, and the mother is currently in her third trimester of pregnancy.

Discussion: This case presents an atypical deletion of the 22q11.2 region, with notable phenotypic features of 22q11.2 deletion syndrome - namely, psychiatric issues as well as velopharyngeal insufficiency and craniofacial dysmorphisms. Further study will aid in an improved genotype-phenotype correlation and enhance our understanding of distal deletions of the 22q11.2 region.

2638T

A novel microdeletion affecting *SNRPN* but preserving distal gene expression leads to Prader-Willi Syndrome. *T. Diallo[1], R. Begay[1], D. Slavov[1], S.L. Graw[1], T. Boyle[2], M.R.G. Taylor[1], P.R. Baker II[3].* 1) Adult Medical Genetics Clinic, University of Colorado Denver - Anschutz Medical Campus, Aurora, CO., United States; 2) Departments of Medicine and Pathology, University of Colorado Denver - Anschutz Medical Campus, Aurora, CO., United States; 3) Clinical Genetics and Metabolism, The Children's Hospital, Denver, CO., United States.

Background: Prader-Willi Syndrome (PWS) is an imprinting disorder characterized by initial neonatal hypotonia, feeding problems, and failure-to-thrive, followed by hyperphagia, obesity, and developmental delay. The most common causes of PWS are 5-7 Mb class I or II deletions of paternally expressed genes at the 15q11.2 locus. Atypical microdeletions have implicated various genes as critical for the development of pathology and continue to shed light on our incomplete understanding of PWS. **Methods and Results:** A 23 year-old woman with typical features of Prader-Willi Syndrome was evaluated clinically. As a neonate she was severely hypotonic with early feeding problems and failure-to-thrive. By 8 years of age her parents noted lack of satiety, obesity, and food hoarding. She exhibited delayed puberty, mild developmental delay, unique behavioral characteristics, and her IQ was considered to be low-normal. Her overall PWS clinical score was 13 (based on Holm et al. 2002 criteria). Methylation analysis, classic PWS FISH, and clinical *SNRPN* expression studies were normal, but oligoarray studies revealed a microdeletion of up to 132 kb. SNP homozygosity analysis narrowed the deletion to approximately 70 kb between intron 2 and intron 4 of *SNRPN*, and to not extend into the *SNURF* transcript. RT-PCR analysis was compatible with loss of *SNRPN* and retention of *SNURF* transcripts. Importantly, the patient's deletion did not include the *SNORD116* cluster or the imprinting center, with preserved expression of transcripts distal to *SNRPN*. **Conclusion:** Rare cases of PWS with attenuated phenotypes have been associated with microdeletions in snoRNA clusters within the *SNRPN/SNURF* locus (Sahoo et al. 2008, DeSmyth et al., 2009; Duker et al. 2010). In these cases methylation testing and *SNRPN* expression testing were still diagnostic. Our proband, with a striking, albeit attenuated, clinical picture represents the first patient with a PWS phenotype and a microdeletion involving only a small part of *SNRPN* (exons 2&3) which tested falsely negative on methylation, FISH, and expression analysis. Because of the size and location of the deletion the use of oligoarray was invaluable in the diagnosis of this patient. Further, and more mechanistically important, we have demonstrated a PWS phenotype with preserved expression of the *HBII-85/SNORD116* snoRNA cluster, challenging the hypothesis that this is the only PWS critical region.

2639S

Overgrowth in association with 3q25 microdeletion. *K. Enomoto[1,2,3], N. Kurosawa[2], R. Satomi[1], Y. Sugawara[3], A. Watanabe[2], S. Watanabe[2], K. Kurosawa[4].* 1) Department of Pediatrics, JA Toride Medical Center, Toride, Ibaraki, Japan; 2) Department of Pediatrics, Tsuchiura Kyodo General Hospital, Tsuchiura, Ibaraki, Japan; 3) Department of Pediatrics and Developmental Biology, Tokyo Medical and Dental University Graduate School, Tokyo, Japan; 4) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan.

Prevalence of congenital disorders associated with overgrowth is not so much compared to what with growth failure. Generally speaking, the rare and specific physical features should sometimes reduce the difficulty with diagnosis in some unknown cases. However, the diagnosis in association with overgrowth is often difficult because of few of other distinctive symptoms or nothing but nonspecific or confusing dysmorphic faetures. It's not easy to detect disease-causing alterations of chromosome with traditional investigations such as karyotype, though some characteristic syndromes can be analyzed by using technic of molecular biology or genetics, e.g. fragile X syndrome, Beckwith-Wiedemann syndrome (BWS), and Simpson-Golabi-Behmel syndrome. We have limited cases showing overgrowth previously reported associated with chromosome anomalies, e.g. duplication in 4p, deletion in 5q35 (Sotos syndrome), duplication in 11p15 (BWS), duplication in 12p, 22q13 deletion syndrome, and Pallister-Killian syndrome (mosaic 12p tetrasomy). We herein present a new case of 3q25 microdeletion extending to 700kb involving *SHOX2* and some other genes identified by microarray CGH examination: arr 3q25.32(157,649,680-158,323,134)x1 (hg19). The patient who is 5-year-old boy at present was born at 38 weeks gestation. His birth weight, length, and head circumference were 3,828g (+2.89 SD), 53.0cm (+2.66SD), and 36.0cm (+2.25SD), respectively. He was transferred to our hospital in early newborn owing to his systemic angioma including facial invasion without hemihypertrophy or limb asymmetry. He had not been complicated by any other problems in newborn to early childhood. He exhibits obvious overgrowth especially in his postnatal weight and head circumference (macrocephaly), mild developmental delay, and facial dysmorphism. Facies is characterized by frontal bossing, downslanting palpebral fissures, bulbous nose, and mildly dysplastic external ears. Angioma is now getting thinner and fairly disappearing. Our investigation reveals that he has the smallest deletion range in the previously reported cases. As our observations delineate the phenotypic spectrum associated with a clearly defined deletion of chromosome 3q25, we suggest that the haploinsufficiency of the genes involving the deleted region contributes to the above mentioned some characteristic features, and this part of chromosome can be the almost critical lesion of 3q25 microdeletion syndrome.

2640M

Diamond-Blackfan anemia and intellectual disability: a new contiguous gene syndrome at 15q25.2. *M. Gorce[1], A. Guichet[1], A. Donzeau[2], A. Ziegler[1], M. Barth[1], I. Pellier[2], D. Bonneau[1], E. Colin[1].* 1) Department of Biochemistry and Genetics, CHU Angers, ANGERS, France; 2) Department of Pediatrics, CHU Angers, Angers, France.

15q25.2 microdeletion is an emergent CNV locus for intellectual disability, dysmorphic features and congenital anomalies. Two distinct microdeletions have been described at this locus: 1) a distal deletion (11 cases) responsible for neurodevelopmental and neuropsychiatric disorders and 2) a proximal deletion (8 cases) which is a susceptibility locus for cognitive deficit, diaphragmatic hernia and possibly Diamond-Blackfan anemia (DBA). This proximal deletion is said to predispose to DBA because it contains the gene RPS17 encoding for a ribosomal protein, responsible for 2% of DBA. Until now, however, DBA has been diagnosed with certainty in only one case of 15q25 proximal deletion. The additional case reported here had a history of intrauterine growth retardation. At age 18 months, he had a moderate developmental delay, dysmorphic features and musculoskeletal anomalies. He had a normochromic macrocytic aregenerative anemia with elevated erythrocyte adenosine deaminase activity and elevated HbF (3,2%) highly suggestive of DBA. A 15q25.2 microdeletion of 2.2Mb including RPS17 was identified using SNP array and the deletion of RPS17 was confirmed by FISH using a specific probe. The deletion was absent in patient's father and was impossible to test in mother. To date, only few mutations in RSP17 have been reported in patients with DBA. Anemia was mentioned in 4 cases among the 8 previous reported cases of 15q25.2 proximal deletion but the definite diagnosis of DBA was made in only 1 case. The present report confirms that patients with 15q25.2 deletion involving RSP17 are at risk for DBA and possibly DBA-associated malignancies.

2641T

Expanding the Phenotypic Profile of Kleefstra Syndrome: a Female with Near-Normal Intelligence and Developmental Dyspraxia. *D. Sisson[1], D.C. Gibbs[1], A.L. Gropman[2,3], C. Sprouse[2], T. Sadeghin[3], C. Samango-Sprouse[1,2,3,4].* 1) The Focus Foundation, Davidsonville, MD; 2) Children's National Medical Center, Washington, DC; 3) George Washington University School of Medicine, Washington, DC; 4) Neurodevelopmental Diagnostic Center for Young Children, Davidsonville, MD.

Kleefstra syndrome is a rare neurogenetic disorder most commonly caused by deletion in the 9q34.3 chromosomal region and is associated with intellectual disabilities, severe speech delay and behavioral problems. To our knowledge, this is the first patient with a 9q34.3 deletion that is not intellectual disabled but has a developmental disorder with overlapping features expanding the phenotypic profile and prognosis for patients with Kleefstra syndrome. The patient is a 6 year-old female whose neurodevelopmental and cognitive profile was assessed using the Beery-Buktenica Visual Motor Integration tests (Beery-VMI), the Expressive and Receptive One Word Picture Vocabulary Tests-Revised (EOWPVT-R/ROWPVT-R), Preschool Language Scale-4th Ed. (PLS-4), the Wechsler Intelligence scales (WPPSI-III) and the Social Responsiveness Scale (SRS-2). Verbal IQ was 81 and Global IQ was 94 on WPPSI-III. Beery subtests indicated severe visual motor deficits (0.08 percentile, VMI; 0.05 percentile, Visual Perception; 0.02 percentile, Motor Coordination). Receptive vocabulary (standard score=99, ROWPVT-R) was within normal range and significantly stronger than expressive vocabulary (standard score=82, EOWPVT-R). Auditory Comprehension and Expressive Communication were within normal limits for age but in the low-average range. SRS-2 indicated delays in social cognition (T-score=89) social communication (T-score=81) and autistic features (T-score>90) but the patient did not meet criteria for an Autism Spectrum Disorder. These results confirm the patient's previous diagnosis of developmental dyspraxia associated with motor planning deficits affecting her speech, oculomotor and graphomotor abilities for which the patient received early intervention services. These services may have contributed to the patient's above average speech and intellectual capabilities than previously reported cases with a 9q34.3 deletion, expanding the phenotypic profile associated with Kleefstra syndrome. Thus, dyspraxia should be considered in the diagnosis and prognosis of patients with a 9q34 deletion as this case-report suggests that targeted intervention therapies may significantly improve patients' cognitive outcome.

2642S

Craniofacial Dysmorphism And Mild Intellectual Disabilities In A Child With A Paternally Inherited 14q32.1 Deletion. *Y. Wang, J. Martinez.* Dept Pathology, Univ South Alabama, Mobile, AL. 600 Clinc dr. Mobile, AL 36688.

Deletions in chromosome 14q32 have been reported in association with intellectual disabilities and congenital anomalies. This is an imprinted genomic region with paternally and maternally expressed genes. We present a child with dysmorphic features and mild intellectual disabilities associated with a paternally inherited 14q32.1 chromosomal deletion. This is a 9 yr old white male seen in the Genetics clinic with history of developmental delay and unusual features. The child was born at term to the first pregnancy of a 20 yr old mother from a pregnancy complicated by oligohydramnios. Birth weight was 8 pound 6 ounces. Birth length was 20 and ¾ inches. He had no neonatal problems but he was later noted to have global developmental delay, feeding and sensory processing difficulties and speech abnormalities. He was attending school at the 3rd grade level within an I.E.P and he is having learning difficulties. He was also evaluated for constipation and GI studies revealed a "small colon". A rectal biopsy was negative for Hirschsprung disease. The family history is significant for the child's father who has similar craniofacial features but he does not have intellectual disabilities. Physical examination revealed an OFC of 51 cm (25th%); Height of 129.5 cm (25th%) and Weight of 28.6 Kg (50th%). He has craniofacial dysmorphism including a narrow bifrontal diameter, high nasal bridge, downslanted palpebral fissures, puffy cheeks, thick lips and a high and narrow palate. There is also increased joint laxity and skin elasticity. Otherwise the physical examination is negative. A CGH microarray revealed a paternally inherited 14q32.12q32.13 encompasses 1.3 Mb and thirteen genes including RPS6KA5, GPR68, CCDC88C, FBLN5, TRIP11, NDUFB1, CPSF2, SLC24A4, RIN3, LGMN, GOLGA5 and CHGA. The common phenotype of 14q32 deletion syndrome includes varying degrees of developmental delay or intellectual disabilities, speech and communication abnormalities and dysmorphic features. Our patient has similar clinical findings previously identified in individuals with deletions of chromosome 14q32.2. Furthermore, the chromosome 14q32 region harbors a cluster of imprinted genes. Deletions affecting this cluster could disrupt regulatory elements increasing the recurrence risk. Deletions of chromosome 14q32.1 region however have not been reported and in that context, our family offers are rare opportunity of understanding mechanisms that alters genetic expression in this region.

2643M

Clinical features and molecular characterization in a subject with an interstitial deletion of 2q24.2. *H. Yoshihashi[1], S. Ito[2], T. Kuchikata[1].* 1) Div Med Gen, Tolyo Metropolitan Childrn's Hos, Fuchu-shi, Japan; 2) Div Nursing, Tolyo Metropolitan Childrn's Hos, Fuchu-shi, Japan.

Microarray analysis has identified an interstitial deletion syndrome in 1q21, 9q22.3, 15q13.3, 15q24 and 16p11.2. A deletion of 2q24.2 is rarely reported cytogenic aberration in patients with developmental delay(DD)/intellactual disability(ID). However, there have been only a few reports describing characteristic findings on the basis of molecularly defined breakpoints. Here, we present clinical manifestations and review current reports of a deletion in 2q24.2. Subject: 7-year-old male. He was born at term after an uneventful pregnancy and postnatal period, except for low-birth-weight infant (2392g), mild hypotonia and VSD. He had moderate delay in psychomotor development with aging. Conventional chromosomal study showed normal male karyotype. At the age of 3 year old, he represented the facial dysmorphism, including a broad forehead, arched eyebrows, long eye lashes, up-turned nares, long philtrum, full lips, thick ear lobes and low hair line. Further genetic testing using array CGH (ISCA 180K manufactured by Agilent) revealed an interstitial deletion (approximately 6.5Mb) of 2q24.2. Parental analysis showed no clinically significant CNVs. The precise breakpoints were determined [arr2q24.1q24.2 (157,612,737-162,620,308)x1 dn]. Discussion: At least 10 patients with a deletion of 2q24.2 have been reported and partly delineated clinical features and breakpoints, but it may be difficult to be suggestive of this cytogenic aberration from dysmorphic features without microarray analysis. The deleted region in our subject encompassed 23 genes and the smallest overlapping region (SOR: approximately 1.8Mb) indicated by Yokoyama et al.(2014), which involved eight genes (LY75, PLA2R1, ITGB6, RBMS1, TANK, PSMD14, TBR1, SLC4A10). The haploinsufficiency of genes in SOR will be responsible for common phenotypic features, including low birth weight, DD/ID and hypotonia in most patients with a deletion of 2q24.2. TBR1 and SLC4A10, critical for early cortical development and regulation of the intracellular pH of neurons respectively, may have an influence on pathogenesis of psychomotor development. RMBS1 and PSMD14, associated with decreased cell proliferation and severe IUGR in animal models respectively, may be noteworthy as causative genes of growth failure. The clinical features and molecular characterization in our subject support that an interstitial deletion of 2q24.2 may cause an emerging deletion syndrome.

2644T

22Q11 DELETION SIZE IN CHILEAN PATIENTS AND ASSOCIATION WITH CLINICAL FEATURES. *G.M. Repetto[1,2], C. Vial[1], L. Leon[1], K. Espinoza[1], F. Benavides[1], M.L. Guzman[1], T. Guo[3].* 1) Center for Genetics and Genomics Clin Alemana- Univ Desarrollo, Santiago, Chile; 2) Hospital Padre Hurtado, Santiago, Chile; 3) Department of Genetics, Albert Einstein College of Medicine at Yeshiva University, New York, USA.

Three common deletion sizes have been reported in the majority of patients with chromosome 22q11 microdeletion syndrome (22q11DS): 3Mb in more than 90% of patients, 2Mb and 1.5 Mb in a smaller proportion. Their correlation with clinical phenotype is unclear, in part due to the limited number of patients with small deletions. We evaluated precise deletion size and location in a large group of Chilean patients with 22q11DS, correlated the findings with the presence of congenital heart disease (CHD) and palate anomalies (PA) and compared the results with similar data from the literature. We analyzed DNA samples from 218 individuals with Affymetrix v6.0 SNP array and with MLPA P250 DiGeorge probemix. There was concordance between the array and MLPA results. We found that 201 patients (92.2%) had a 3Mb deletion, 5 (2.2%) had a 2Mb deletion and 12 (5.5%) had a 1.5Mb deletion. In ten individuals, proximal regions PRODH to DGCR10 were found to be non-deleted by SNP array. These findings were not identified by MLPA, as expected by the probemix composition. When precise breakpoints were analyzed, we observed variation within the low copy repeats that flank the common breakpoints. The frequency of CHD was significantly higher in the individuals with the 3 Mb deletion in comparison with the 1.5 Mb deletion (58.9% vs 25%; p 0.045; OR 4,308, 95% CI 1.025-20.787). This difference was not observed for the presence of PA. We then aggregated our results with data from the literature that described deletion size and pertinent clinical data (Kurahashi et al 2007; Michaelovsky et al 2012). This resulted in information from 360 patients: 330 with 3Mb deletion, 9 with 2Mb and 21 with 1.5 Mb. We confirmed the association between the presence of CHD and the 3Mb deletion, compared with 1.5 Mb (p=0.005; OR 4.0, 95% CI 1.46-11.30). This large cohort of Chilean patients confirms that the large 3Mb deletion is the most common one in this syndrome, and that there is variation within the common breakpoints. Individuals with the 3Mb deletions had greater frequency of CHD compared with the ones with 1.5 Mb deletions, suggesting the existence of modifiers in the non-overlapping region in 22q11. The implications of the presence of both copies of the proximal region in a small group of patients need to be carefully evaluated, particularly given the apparent role of PRODH and proline metabolism in the psychiatric phenotype. Funded by Fondecyt-Chile grant #1130392.

2645S

Unknown CNVs found in 52 Bulgarian patients with intelectual disability and congenital malformations. *S.P. Hadjidekova[1], D. Avdjieva-Tzavella[1, 2], B. Rukova[1], R. Staneva[1], D. Nesheva[1], R. Tincheva[1], D. Toncheva[1].* 1) Department of Medical Genetics, Medical Faculty, Medical University- Sofia, Bulgaria; 2) State University Pediatrics Hospital "Prof. Ivan Mitev", Medical Faculty, Medical University- Sofia, Bulgaria.

Introduction: The evergrowing resolution of the DNA microarrays and the optimization of the technique allow detection of larger aberrations and a variety of small copy number variations (CNVs) whose clinical significance is unknown in some cases. Methods: We studied 52 patients with congenital anomalies and intellectual disability. Whole-genome oligo-array CGH was performed using the BlueGnome CytoChip oligo 2×105K microarray, v1.1. Results: We found a total of 247 CNVs, of which 15 pathogenic (7 deletions, 8 duplications), 124 benign (62 deletions, 62 duplications) and 108 with unknown clinical significance (68 deletions, 40 duplications). The unknown CNVs were detected among 34 patients with the following distribution in size : 52.7% less than 100Kb; 37% from 100 Kb to 500Kb; 9.2% from 500Kb to 1 Mb; 0.9% larger than 1Mb. We applied a specific algorithm in the interpretation of the unknown variants. Conclusion: The rates of unknown CNVs in our study were notably high: This indicates that some of these variations may be probably benign for the Bulgarian population and cannot be found in the studied other populations. It is speculated that the Bulgarians are characterized by high genetic heterogeneity. This demonstrates the obvious need for large population studies and mapping of variations in the Bulgarian population for enabling the precise interpretation of unknown CNVs in the clinical praxis. The protocol and informed consent documents were reviewed and approved by the local ethics committee and prepared according the Declaration of Helsinki and local country laws. Acknowledgements: Grant 02/76-21.12.2009, National Science Fund, Bulgaria.

2646M

A rare case of speech sound disorder with a heterozygous BCL11A deletion. *A. Huang[1], B. Peter[1], . Center for Mendelian Genomics[2], Z. Brkanac[3], A. Stocco[4], M. Matsushita[5], J. Wolff[5], W. Raskind[3, 5].* 1) Speech & Hearing Science, Arizona State University, Tempe, AZ; 2) Genome Sciences, University of Washington, Seattle, WA; 3) Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA; 4) Psychology, University of Washington, Seattle, WA; 5) Medicine, University of Washington, Seattle, WA.

Speech sound disorder (SSD) interferes with a child's ability to acquire clearly intelligible speech. One proposed subtype is childhood apraxia of speech (CAS), thought to result from faulty programming of movement sequences for speech. CAS is often comorbid with reading disorder (RD), a disorder of written language. We recently described a child with CAS with a de novo heterozygous microdeletion of BCL11A and 100 kb of surrounding noncoding sequences. Speech and gross motor development showed evidence of poor motor programming and low muscle tone. In eight of ten previously reported cases with microdeletions in 2p15-p16.1, multiple genes including BCL11A were deleted; phenotypic involvement was more severe. BCL11A regulates fetal hemoglobin and is located in an RD candidate region; its role in speech, reading/spelling, and motor development is not yet well understood. A previous study reported 16 CNVs in 12 children with CAS but causality could not be established (Laffin et al., 2012); none of the reported CNVs involved BCL11A. We investigated the role of CNVs of BCL11A and other regions, including candidate regions for CAS and RD, in 9 individuals with CAS and 3 individuals with RD. One control individual had tongue thrust (TT), an SSD subtype unlikely to share biological etiologies with CAS or RD. Also included were three unaffected individuals, biologically related to two CAS participants, and two families with familial CAS, with 9 and 4 affected individuals, respectively. DNA was extracted from buffy coat. Genotyping was performed with 964 k SNP arrays. CNVs were analyzed with cnvHap and PennCNV algorithms. No participant had CNVs involving BCL11A. Overlaps with two CNVs in the Laffin et al. (2012) study were found to be common CNVs. Five participants (1 RD, 1 TT, 1 CAS, 2 controls) carried four copies of an identical region of the FOXP2 gene. Two participants (1 RD, 1 CAS) had a CNV in an identical region of CNTNAP2. There was no evidence that these CNVs segregated, however, and they are unlikely to be pathogenic. Two or more participants with CAS and RD shared 64 other CNVs. We conclude that BCL11A deletions are uncommon in CAS. In future studies, we will evaluate the coding and regulatory roles of this gene in RD and CAS and follow up on CNVs from the present study. - Genotyping and CNV analyses were provided by the University of Washington Center for Mendelian Genomics (1U54HG006493 to Drs. D. Nickerson, J. Shendure and M. Bamshad).

2647T

Congenital asplenia in a patient with chromosome 1p36 deletion. *L. Pisani[1], P. Zachariah[2], A. Gomez[2], J. Reiner[1], N. Cohen[1], L. Mehta[1].* 1) Dept. of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, NY; 2) Dept. of Pediatrics, Columbia University College of Physicians and Surgeons, NY.

Chromosome 1p36 deletion syndrome is a well-described microdeletion, with an incidence of about 1/5000 newborns. Common manifestations are typical facial features, intellectual disability, hypotonia, seizures, structural brain abnormalities, congenital heart defects, ocular anomalies and hypothyroidism. Immunodeficiency has not been previously associated with this syndrome. We report a 2-year-old girl who presented with a complex partial seizure in the setting of an acute febrile illness. She had a history of global developmental delay and was noted to have frontal bossing, open anterior fontanel, low set ears, flat nasal bridge, short palpebral fissures and overlapping 2nd-3rd toes bilaterally. Blood and CSF cultures at admission were positive for Streptococcus pneumoniae. The patient and her siblings had not received any immunizations due to religious reasons. Family history was negative for intellectual disability, consanguinity or birth defects. Microarray revealed the presence of a pathogenic 5.33 Mb deletion of chromosome 1p36.31-p36.33(736471-6070487, build hg18), including 112 genes and transcripts and overlapping the critical region for 1p36 deletion syndrome. The deletion was confirmed by FISH. In addition, a 91.6 kb duplication on chromosome Xp22.31(8557747-8658299) was noted, encompassing exon 2 of KAL1, considered likely benign. Testing of parents is pending. She received a 2-week course of Ceftriaxone but was then lost to follow up. Four months later, the patient was readmitted with new fever and lethargy. Blood culture was positive for Hemophilus influenzae with CSF analysis suggestive of meningitis. Her peripheral smear was notable for Howell-Jolly bodies and abdominal ultrasound showed a 2.5 cm structure at the site of the spleen. Echocardiogram was normal. She was discharged on amoxicillin prophylaxis after completing a course of Ceftriaxone with recommendations for catch-up vaccinations. This is the first report of congenital functional asplenia in 1p36 deletion. The deletion does not include any known genes implicated in heterotaxy, and is of a commonly reported size. Unmasking of a recessive allele could be considered as a potential mechanism, but no relevant genes could be identified in the deletion interval. This report highlights the presence of a rare birth defect in a relatively well characterized microdeletion syndrome.

2648S

De novo deletion of 5q23.2-q31.1 in a boy with global developmental delay, contractures and dysmorphic features: a contiguous gene deletion syndrome involving morphogenesis and DNA repair. *A. Guerin[1], R. Gatti[4], C. Brown[4], M.S. Meyn[2,3], M. Carter[2,3].* 1) Division of Medical Genetics, Department of Pediatrics, Kingston General Hospital, Kingston, Ontario, Canada; 2) Division of Clinical and Metabolic Genetics, Department of Pediatrics, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Program in Genetics and Genome Biology, SickKids Research Institute, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) UCLA School of Medicine Department of Pathology and Laboratory Medicine, Los Angeles California.

Contiguous gene deletion syndromes represent a diagnostic dilemma in interpreting complex phenotypes. Genes that are deleted can have important roles in morphogenesis, which can result in malformations, but genes with ongoing expression, which are essential for cell survival, can also be affected. We present a patient with a de novo 8.4 Mb deletion of 5q23.2-q31.1, a rare microdeletion syndrome. The patient presented with congenital contractures, dysmorphic features, symmetric growth retardation and global developmental delay. The deletion encompassed 12 OMIM genes, including FBN2, which is expressed in early morphogenesis. Deletions of FBN2 have been reported in patients with Beals syndrome (MIM 121050); this is likely the cause of our patient's congenital contractures; however, he did not have the other characteristic features (arachnodactyly, Marfanoid habitus). Our patient also had haploinsufficiency for the RAD50 gene, whose product is part of the MRN complex, required for repairing DNA damage. Complete RAD50 deficiency leads to a Nijmegen breakage syndrome-like disorder (#613078), characterized by microcephaly and growth retardation, "birdlike" facies and chromosome instability. Given the possibility of decreased production of key repair proteins, chromosome breakage and colony survival (radiosensitivity) assays were performed on a lymphoblast line. These studies were within normal limits. Western blot showed normal levels of ATM, NBS, and RAD50. Hetero- zygous missense variants in RAD50 may be low penetrance risk factors for breast cancer, so it remains uncertain if there is an impact of this deletion on future cancer risk. Contiguous gene deletion syndromes represent a challenge in providing counseling and anticipatory guidance regarding microarray findings, especially when genes responsible can have variable clinical phenotype, and are responsible for DNA repair.

2649M

A case with mild phenotype of holoprosencephaly is caused by de novo hemizygosity for chromosome 2q14.1-q14.3 involving GLI2 gene. *E. Kirtas[1], E. Kirat[1], A. Koparir[1], E. Guzel[1], E. Fenercioglu[1], M. Seven[1], H. Ulucan[1], M. Ozen[1,2], G. Guven[1].* 1) Istanbul University, Medical School of Cerrahpasa, Department of Medical Genetics, Istanbul, Turkey; 2) Department of Pathology & Immunology Baylor College of Medicine, Houston, TX, 77030, USA.

GLI2 is mediated as an activator by central transcriptional factor on SHH signaling although its precise role on SHH signaling remains unclear. Defects in this gene are associated with holoprosencephaly (HPE) with extremely variable phenotypic penetrance or HPE-like phenotype, abnormal pituitary gland formation and/or function, and polydactyly. Here we present a case with HPE-like phenotype having a 2q14.1-q14.3 deletion. The case was a female, born to healthy and 1.5 degree consanguineous parents, 28-year-old mother and 42-year-old father. She was delivered at term with a birth weight of 3040 g (25-50p) and birth length of 47 cm (10-25 p). She was referred to our department at the age of 2.5 years for mild neurodevelopment delay. Bilateral postaxial polydactyly, grade 3 vesicoureteral reflux, and recurrent infections were treated short after the birth. In physical examination, her weight was 13 kg (%50p) and her height was 90 cm (%50p) while head circumference was 46 cm (<3p). Urinary and sacral USG were within normal ranges while atrophy of parenchyma in left kidney with 99mTc-DMSA SPECT was detected. Cranial CT, MRI and elbow radiography showed no abnormality. Amino acid and metabolic screen were normal except from moderate high levels of lactate and ammonia. Chromosome analysis from the blood lymphocytes revealed 46,XX,ins del(2)(q14.2;q21)dn,inv(9)(p11q13)mat karyotype. Her mother's and father's karyotype were 46,XX,inv(9)(p11q13) and 46,XY, respectively, indicating maternal origin. A detailed analysis of the constitutional chromosomal changes in proband was defined by array comparative genomic hybridization (aCGH) revealed a 5.6 mb heterozygous deletion spanning from 2q14.1 to 2q14.3 has been found. This region involves whole GLI2, which is thought to be responsible for this mild form of HPE. Our results on GLI2 deletion in this case confirms phenotypic heterogeneity and mild form of HPE.

2650T

Severe fetal phenotype of a dominant mesomelic dysplasia, associated with a 790 kb microduplication of HOXD gene cluster at 2q31.1. *S. Odent[1, 7], C. Jeanne-Pasquier[2], P. Balouet[3], N. Leporrier[4], S. Jaillard[5], V. Jauffret[5,6], E. Cherot[1], M. Fradin[1], C. Quelin[1], L. Pasquier[1], V. David[6,7], C. Dubourg[6,7].* 1) Genetique Clinique, CHU de Rennes Hosp SUD, Rennes CDX 2, France; 2) Service d'Anatomie Pathologique, CHU de Caen, 14033 CAEN , France; 3) Service Maternite gynecologie, CH de Saint-Lo, 50009 saint-Lo, France; 4) Service de Genetique, CHU de Caen, 14033 CAEN , France; 5) Service de cytogenetique, CHU de Rennes, Hopital Pontchaillou, Rennes, France; 6) Service de genetique moleculaire et genomique, CHU de Rennes, Hopital Pontchaillou, Rennes, France; 7) CNRS UMR 6290, IGDR, Université Rennes1, France.

Mesomelic dysplasias are a very heterogeneous group of skeletal disorders. Among them, the 2q31.1 microduplication syndrome and Mesomelic Kantaputra dysplasia (MIM 156232) have been associated with microduplication of a cluster of HOXD genes. Phenotype of early fetal mesomelic dysplasia was essentially described in Langer mesomelic dysplasia (MIM 249700) caused by homozygous deletion or mutation of the SHOX gene or PAR1 region downstream of SHOX. We report here the very early fetal phenotype of a mesomelic dysplasia inherited from the father, who was followed since childhood for a growth retardation (-2.5 SD) with macrocephaly (+4 SD) , ulnar deviation of hands and club feet, a significant shortening of ulna and fibula, bowed forearms and legs, difficulties with flexion / extension of the fingers. At 11 weeks of gestation, ultrasound showed a cystic hygroma of 4.5 mm, a punctate appearance of the forearm's bones , shortening of the lower limbs, clubfeet, and an unilateral pleural effusion. A TOP was performed at 13 weeks of gestation. Mesomelic dysplasia was confirmed by fetal autopsy, X-rays, and a particular histological appearance (organization of chondrocytes in shortened columns, triangular densification at diaphyseal angulations). Fetal DNA analysis by array-CGH revealed a microduplication of a 790 Kb involving genes HOXD13, HOXD12, HOXD11, HOXD10, HOXD8, HOXD4, HOXD3, MTX2, which was confirmed by QMPSF. This microduplication was also identified in the father. Conclusion : the different phenotypes of mesomelic dysplasia with early prenatal expression are close, sometimes more severe than the phenotype of an affected parent; research of a genomic microrearrangement is indicated since it can reveal notably a deletion of the SHOX region in Xp22.33, or 2q31.1 microduplication encompassing the HOXD cluster.

2651S

Adaptive and Maladaptive Behavior, Profiles and Developmental Trajectories in Children with Subtelomeric Microdeletions. *G.S. Fisch, S. Yang.* Epidemiology & Health Promotion, New York University, New York, NY., USA.

Despite its importance to clinicians and pediatricians, there are relatively few studies currently which examine developmental trajectories or profiles of adaptive and maladaptive behavior in children with genetic disorders that produce intellectual disability (ID) and/or autism (ASD). Recently, we undertook to study children with subtelomeric deletions that produce ID and/or ASD. We recruited 47 children, 26 females, 21 males, diagnosed with Wolf-Hirschhorn Syndrome (WHS), or one of three other subtelomeric deletions (invdupdel8p23, 2q37 or 11q23-25 (Jacobsen syndrome; JBS). Mean age for the sample at T1 was 11.1 years (±4.4). Initially, we assessed their cognitive abilities, adaptive behavior (DQ) and maladaptive behavior using standardized instruments at Time 1 (T1). Two years later, at Time 2 (T2), we were able to retest 31 children. We compared adaptive and maladaptive scores at T1, but also compared adaptive and maladaptive difference scores from T1 to T2 for the entire group, and by genetic disorder, DQ score at T1 or maladaptive score at T1. Results show an inverse correlation between DQ score and maladaptive score, but it was not statistically significant. However, results from ANOVA indicate statistically significant test-retest differences in maladaptive scores among the four syndromes ($F=3.336$; $P=0.034$). Post-hoc analysis showed 2q37 and JBS maladaptive difference scores were statistically significantly different ($P=0.046$). Children with WHS were most severely impacted in cognitive and adaptive behavior, but with relatively lower maladaptive behavior scores which tend to improve as children age. Curiously, children with JBS have the highest mean IQ but the worst maladaptive behavior and adaptive behavior scores. We employed a general linear model to analyze maladaptive behavior and found maladaptive scores were significantly related to both genetic disorder and changes in IQ scores from T1 to T2 ($F=3.88$; $p< 0.01$). Aspects of maladaptive behavior, particularly as they relate to attention deficits and hyperactivity are currently being analyzed and will be discussed.

2652M

Mosaic 15q11-q13 maternal duplication without Autism. *N. Urraca[1], E. Pivnick[2], K. McVicar[2], R. Thibert[3], S. Parisotto[4], H. Pedro[4], L.T. Reiter[1,2].* 1) Neurology, UTHSC, Memphis, TN; 2) Pediatrics, UTHSC, Memphis, TN; 3) Neurology, Mass General Hospital, Boston, MA; 4) Genetics, UMC, Hackensack, NJ.

Most individuals with an interstitial 15q duplication share common deletion breakpoints with Prader-Willi/Angelman Syndrome (PWS/AS) and are the result of the reciprocal non-allelic homologous recombination (NAHR) events that forms the PWS/AS deletions. Because the 15q11-q13 region has a cluster of genes preferentially expressed from one parental allele, there is also a parent of origin effect for the duplication as well as the deletion. Maternally inherited 15q duplications are generally fully penetrant for Autism Spectrum Disorder (ASD) while paternal cases have incomplete penetrance for ASD and other features including sleep disorders. In most cases the duplication is de novo; however, there are few informative familial cases that will help us better understand the genes contributing to the ASD phenotype. The majority of duplications found in this region are maternal in origin, most likely due to the presence of ASD. We report two families with maternal inherited 15q duplications: family 1 with a Class II duplication and family 2 with a Class I duplication. Interestingly, in both families the mothers have a maternally derived duplication with no apparent ASD phenotype as would be expected in maternal duplication cases. However, the offspring with the duplication met criteria for ASD by ADOS/ADI-R testing. In family 1, the mother had learning disabilities and ADHD as a child, while the mother in the second family had school/learning issues as well. The Broad Autism Phenotype Questionnaire (BAPQ) and the Social Responsiveness Scale-Adult Research Version (SRS) did not reveal an ASD phenotype in the mother in family 1 and will be completed in the second mother. Fluorescence In Situ Hybridization in peripheral blood lymphocytes and skin fibroblasts showed mosaic 15q duplication for the first mother. Cases of paternal inherited int dup(15) have been reported with no obvious autism phenotype, but the child who inherits this same duplication maternally has ASD. The two families we will present came to our attention because of their affected offspring. We were surprised to find a maternal duplication in the mothers as well, but the mosaic status of this duplication in blood and skin may explain the absence of ASD. These two families confirmed our previous finding of ASD being fully penetrant in maternal 15q duplication cases and indicates that in any maternal case with no ASD phenotype a mosaic status needs to be ruled out.

2653T

Prevalence of "Y" chromosome microdeletions in infertile males of Gujarat Population, India. *T.A. Patel[1], H. Pandya[1], H. Bavishi[2], M.V. Rao[1].* 1) Gujarat Genetic Diagnostic Center (GenDiCe), Department of Zoology, University School of Sciences, Gujarat University, Ahmedabad-09, Gujarat, India; 2) Bavishi Fertility Institute, Near Adani CNG Station, Opp Municipal Garden, Ahmedabad-07, Gujarat, India.

"Y" chromosome microdeletions are the leading genetic cause of male infertility. Hence, the purpose of present study is to screen various "Y" chromosomal STS markers in infertile male patients. The blood samples were collected from infertile male patients (n=41). DNA was isolated by standard protocol and various Y chromosomal STS markers were analyzed. In our study 41 cases, attending the clinic, considered to have infertility were analyzed for semen, hormonal and microdeletion profiles along with other socioeconomic factors including previous history. These data showed 3 out of 41 (8%) showed deletions at Y chromosome level (AZF regions) leading to infertility in relation to other factors. In the remaining patients (38) in our study the causes for infertility depend on other factors like hormonal imbalance etc. The data revealed that microdeletion is also one of the causes for inducing sterility in the male in addition to others. Thus, 8% in this western population, it may be contributory for infertility, although such infertility caused ranges 10-15% in other regions of India. In conclusion, Y chromosomal microdeletions play an important role in male infertility, where our region constitutes 8% within our samples analyzed.

2654S

A microdeletion encompassing only three genes within the Potocki-Shaffer syndrome interval at 11p11.2 associated with intellectual disability and craniofacial anomalies. *J.D.J LABONNE[1], J. Vogt[2], L.C Layman[1], HG. Kim[1].* 1) Department of Obstetrics & Gynecology, Institute of Molecular Medicine and Genetics, Georgia Regents University, Augusta, GA 30912, USA; 2) West Midlands Regional Clinical Genetics Service, Birmingham Women's Hospital, Birmingham, UK, B15 2TG.

Potocki-Shaffer syndrome (PSS) characterized by intellectual disability (ID), craniofacial anomalies (CFA), biparietal foramina, and multiple exostoses is a rare contiguous gene disorder caused by a minimal 2.1 Mb deletion at 11p11.2. EXT2 and ALX4 have been shown to be causative for multiple exostoses and parietal foramina, respectively. Haploinsufficiency of PHF21A was the underlying mechanism of ID and CFA in PSS since affected individuals with balanced translocations and deletions showed a reduction of PHF21A protein level. Furthermore, knock-down of PHF21A in zebrafish resulted in CFA and neuronal apoptosis. The death of neurons is a likely cause of ID in patients. We report a 5-year old boy with ID and CFA with a deletion including PHF21A, providing corroborating evidence for a role of PHF21A in these specific phenotypes. BlueGnome 8x60K ISCA design oligonucleotide microarray revealed a heterozygous deletion of at least 173 kb (chr11:45,936,954-46,110,572 / hg19) at 11p11.2 in this patient. This deletion region encompassing only three genes (PEX16, GYLTL1B and PHF21A) is likely to be the smallest with the phenotypes of ID and CFA, when confirmed. The proximal breakpoint of this deletion lies within intron 2 upstream of the last 5'-UTR segment of PHF21A based on the coordinates of the minimal deletion. This suggests that no truncated PHF21A protein can be produced by the heterozygous microdeletion, although the exact boundary of the deletion will be defined by quantitative PCR (qPCR). The affected individual displays CFA including a broad forehead with a slightly prominent ridge in the centre, mild underdeveloped cheekbones, a flat nasal bridge, a small mouth with a thin upper lip, and glued large ears rotated backwards. He also possesses tapering fingers with clinodactyly and his feet display webbing between the second and third toes. He has a small penis with some surrounding dark hairs and palpable testes. He also has difficulties interacting with other children, and prefers adult company. Because of his ID and clinical features, he has been enrolled in a special needs school. qPCR confirmed the microarray result. RT-qPCR and western blot with the anti-PHF21A antibody displayed that PHF21A transcript and protein levels are reduced approximately by one half in this deletion patient. Reduced expression levels of PHF21A indicates that this gene is responsible for the phenotypes of ID and CFA seen in this deletion patient.

2655M

Recurrent microdeletion 12p13 in maternal half-siblings suggestive of gonadal mosaicism. *A.F. Elias, C. Hudson, K. Berry, C. Schwanke, T. Schwalbe, S. Philips, D. Xu.* Medical Genetics, Shodair Children's Hospital, Helena, MT., Select a Country.

Microdeletion 12p13 involving the terminal deletion of variable segments of the short arm of chromosome 12 represents one of the least common subtelomeric rearrangements resulting in a very rare, yet increasingly recognized cause for cognitive impairment. The condition is associated with a variable phenotype including language delay, a specific speech impairment described as childhood apraxia of speech (CAS), developmental delay and intellectual disability (ID) as well as neurobehavioral problems. Other features observed in 12p13 deletions include microcephaly, mild craniofacial dysmorphism, generalized muscular hypotonia, and joint laxity. To our knowledge, germline mosaicism has not been reported for this microdeletion syndrome. Here, we report the presence of a microdeletion 12p13.33p13.32 in two half-siblings who have the same mother and different fathers. Both probands have neurodevelopmental and neurobehavioral manifestations. The 8 year-old girl presents with global developmental delay, CAS, microcephaly and mild dysmorphism, as well as pectus carinatum. Her 18-year-old maternal half-brother was diagnosed with autism at age 5 and has a pectus excavatum deformity. Chromosomal microarray in both probands revealed a 3.9 Mb terminal deletion of the short arm of chromosome 12 (12p13.33p13.32), confirmed by Fluorescent in situ hybridization (FISH) analysis in both cases. Their mother, who is asymptomatic, would be predicted to be an obligate carrier of this deletion. However, FISH analysis of maternal blood lymphocytes and skin fibroblasts did not reveal a deletion or any other rearrangement. Extended analysis of additional interphase nuclei from maternal lymphocytes also was not supportive of possible mosaic deletion in the mother's blood lymphocytes. These results are suggestive of maternal gonadal mosaicism for the familial 12p13 deletion. To our knowledge this observation is the first documentation of germline mosaicism for terminal 12p deletion syndrome. Additional studies in maternal buccal cells as well as studies in blood lymphocytes from other sibs are under way. The implications for estimating recurrence risk and genetic counseling are discussed.

2656T

Chromosome microarray analysis in patients with cleft lip/ cleft palate. *P. Eydoux[1], F. Kozak[2], L. Ogilvie[2], J. Pauwels[2], B. Tsang[1], L. Brown[1].* 1) Dept Pathology, Children's & Women's Hosp, Vancouver, BC, Canada; 2) Dept Pediatrics Children's & Women's Hosp, Vancouver, BC, Canada.

Orofacial clefts are one of the most common malformations among newborns, affecting approximately one in 500 to one in 1000 births. Due to their detrimental effects on physical and psychological well being, a comprehensive approach is required for the management of these children. Both genetic and environmental factors have been associated with the occurrence of cleft lip with or without cleft palate (CLP), occurring in 1-2 in 1000 births. Its highest incidence is found in Asian and First Nations Peoples followed by those of European and African descent, showing involvement of genetic factors. Although little is known about the molecular mechanisms resulting in CLP, copy number variants (CNVs) have long been recognized as associated with CLP in visible chromosome abnormalities or infra-microscopic CNVs. To characterize the etiology of CLP, we conducted a study on 35 consented patients with CLP, considered to be non-syndromic. Patients under two year of age (16 females and 18 males) were tested with chromosome microarray analysis (CMA) using a high-density microarray (Cytoscan HD, Affymetrix). The data was analysed using the Affymetrix software ChAS; interpretation was conducted visually and using software allowing bioinformatics search of the RefSeq, UCSC and OMIM databases. A search was conducted in RefSeq and OMIM with the key-words "cleft lip" or "cleft palate"; in addition, the UCSC database table gwasCatalog was searched with the same key words. We identified pathogenic CNVs in two patients, within regions related to known syndromes: one patient had a 3.67Mb Smith Magenis Syndrome microdeletion at 17p11.2, and another patient had a 2.3Mb microdeletion at 17p13.3, not encompassing the YWHAE gene. In our patients, we did not identify any copy number containing genes or GWAS regions related to CLP in the OMIM, RefSeq and UCSC databases. In addition, regions of homozygosity were unremarkable. Our data does not support involvement of small CNVs in the causation of CLP. In young children with CLP, CMA should be performed to rule out a microrearrangement which may indicate a more complex developmental disorder.

2657S

Xq27.3-q28 duplication syndrome: a new consideration in the differential diagnosis of Prader-Willi syndrome. *M.R. Garcia, M.J.H. Willis.* Pediatrics, Naval Medical Center San Diego, San Diego, CA.

The purpose of this abstract is to present a family with Xq27.3-q28 duplication syndrome, compare their findings to those of previously reported cases, and provide further delineation of this syndrome that has strong phenotypic overlap with Prader-Willi syndrome. Patient 1, is a 7 year-old male who was born at term after an uncomplicated pregnancy. Development first became a concern to the family when he was not able to sit independently at 1 year of age. He continued to have delays in motor and language milestones. At 3 years of age etiologic evaluation were initiated. Chromosomal micro-array identified duplication of Xq27.3-q28. Diagnosis at presentation to our clinic comprised; mild cognitive disability, strabismus, excessive hunger, obstructive sleep apnea, and hypogonadism. His physical exam is significant for short stature, truncal obesity, synophrys, bulbous nasal tip, small scrotum, cryptorchidism, and hypotonia. Patient 2, maternal cousin of Patient 1, is an 11 year old male identified independently to have the same duplication. His medical problems and exam findings are congruent with Patient 1's with the addition of fatty liver and Chiari malformation. In addition to microarray he has had a muscle biopsy and urine glycosaminoglycans which were normal. The mothers of both patients, as well as a third sister, all carry the dupXq27.3-q28. This is the third family reported with Xq27.3-q28 interstitial duplication. The patients described here show significant phenotypic overlap with cases of dupXq27.3-28 described in the literature with keys features of hypotonia, gross motor delay, mild intellectual disability, hypogonadism, short stature and obesity. Deficiency of a number of genes in the duplicated region is associated with human disease; FMR1, IDS, MAMLD1, MTM1 and VMA21. FMR1 duplications have also been associated with intellectual disability. Other genes in this area are likely also dosage sensitive. Xq27.3-q28 interstitial duplication causes an X-linked recessive syndrome with a remarkably consistent phenotype, has very strong overlap with Prader-Willi syndrome (hypotonia, cognitive disabilities, short stature, obesity/polyphagia) and should be considered in the differential for that syndrome. Any patient with phenotypic findings suggestive of Prader-Willi who has negative methylation testing, or who has family history suggestive of x-linked inheritance, should have testing for copy number variants in this region of chromosome X.

2658M

Behavioral, Biochemical and Anthropometric Characteristics of patients with PWS. *H. El-Bassyouni[1], E. M. Salah[2], S. Kholoussi[3], M. Shehab[4], W. Kandeel[5].* 1) Clinical Genetics Department, National Research Centre, Cairo, Egypt; 2) Child Health Department, National Research Centre, Cairo, Egypt; 3) Immunogenetics Department, National Research Centre, Cairo, Egypt; 4) Cytogenetic Department, National Research Centre, Cairo, Egypt; 5) Biological Anthropology Department, National Research Centre, Cairo, Egypt.

Background: Prader-Willi syndrome (PWS) is a genetic disorder characterized by a recognizable pattern of physical findings with significant cognitive, neurologic, endocrine, and behavioral abnormalities. Objective: To study behavioral, cognitive, hormonal and anthropometric characteristics of children with Prader-Willi syndrome (PWS) compared with an age-and gender-matched control group. Subjects and methods: A case- control study of thirteen children and adolescents with PWS (mean age 7.69±4.44 years), and 14 age-and sex- matched non-PWS controls were enrolled. Measurement of anthropometric parameters and body fat percentage (BF %), screening of children's psychosocial dysfunction, assessment of intellectual function and estimation of plasma adiponectin and leptin levels were carried out in all studied subjects. Specific questionnaire for assessment of the behavioral phenotype of PWS was fulfilled by parents. Results: PWS subjects had significantly higher BF%, adiponectin and leptin levels compared to controls (38.84±9.66 vs. 20.26±3.92, P< 0.01; 38.61±13.43ng/ml vs. 24.32±6.04ng/ml, P<0.01;10.39±7.74ng/ml vs. 4±0.71 ng/ml, P< 0.01 respectively). The majority of PWS subjects (84.6%) (11/13) had mild to moderate mental retardation. The most commonly registered behavior problems in PWS subjects were hyperphagia and cognitive rigidity. Inattention was reported in 76.9% (10/13) of PWS vs. 25% (3/14) of the controls. Conclusion: PWS subjects have substantial behavioral problems which are not associated with the degree of hyperphagia or intellectual disability. The distinct body fat distribution in PWS subjects is associated with significantly higher levels of adiponectin and leptin than their lean controls. The primary focus on management of PWS should be placed on weight control and behavior modification.

2659T

A new chromosomal rearrangement resulting in Axenfeld-Rieger syndrome. *L. El Khattabi[1], I. BenAyed[1], C. Ioos[2], L. Perrin[3], D. Le Tessier[1], JM. Dupont[1], A. Lebbar[1].* 1) Laboratoire de Cytogénétique, APHP, Hôpitaux universitaires Paris Centre site Cochin, Paris, France; 2) Unité fonctionnelle de Médecine Physique et Réadaptation de l'enfant, Hôpitaux Universitaires Paris Ouest site Raymond Poincaré, Garches, France; 3) Unité fonctionnelle de Génétique Clinique, CHU Robert Debré, Paris, France.

Axenfeld-Rieger syndrome (ARS) is a developmental disorder that associates eye abnormalities and multiple congenital malformations. An abnormal migration of neural crest cells results in dysgenesis of the anterior segment of the eye. Other manifestations are mild dysmorphic features, dental, cardiac and umbilical anomalies. ARS is inherited in an autosomal dominant manner with high penetrance of *FOXC1* and *PITX2* mutations. However, over 50% of cases remain of unknown cause. Herein, we report a case of ARS associated with epilepsy in a two year-old boy. Clinical examination found some dysmorphic features and growth delay. Ophthalmologic examination revealed a bilateral corectopia, iris hypoplasia and a bilateral embryotoxon, consistent with ARS eye disorders. Whole genome microarray analysis using Nimblegen 135K array showed a 4.55 Mb *de novo* deletion on 4q25q26 bands (nt 112,009,675 to nt 116,558,399 bp - hg18). The deleted region encompasses 9 OMIM genes and its proximal boundary is located 230 Kb upstream of *PITX2*. *In silico* investigation identified the presence of regulatory sequences of *PITX2* in this region. PCR analysis confirmed the deletion of these sequences. Further studies are carried out to assess the disruption of *PITX2* expression. This report of ARS resulting from deletion of *PITX2* gene cis-regulatory elements sequences suggests that analysis of *PITX2* upstream region could be helpful for ARS patients without *PITX2* or *FOXC2* mutations.

2660S

Fetal Skeletal Dysplasias in a Tertiary Care Centre: Radiology, Pathology, and Molecular Analysis of 112 cases. *D. Chitayat[1, 2], E. Barkova[3], U. Mohan[4], S. Keating[5], A. Toi[6], J. Frank[1], R. Frank[1], G. Tomlinson[7], P. Glanc[8].* 1) Mount Sinai Hospital, Department of Obstetrics and Gynecology, The Prenatal Diagnosis and Medical Genetics Program, University of Toronto, Toronto, Ontario, Canada; 2) The Hospital for Sick Children, Division of Clinical and Metabolic Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Department of Medical Imaging, South Shore Regional Hospital, Bridgewater, NS, Canada; 4) Department of Obstetrics and Gynecology, University of Calgary, Calgary, AB, Canada; 5) Mount Sinai Hospital, Department of Laboratory Medicine and Pathobiology -Perinatal Pathology; 6) Mount Sinai Hospital, Department of Diagnostic Imaging; 7) University of Toronto, Institute of Health Policy, Management & Evaluation; 8) Sunnybrook Health Sciences Centre, Department of Medical Imaging.

Fetal skeletal dysplasias are a heterogeneous group of rare genetic disorders, affecting approximately 2.4-4.5/10,000 births. We performed a retrospective review of the perinatal autopsies conducted between the years 2002-2011. The study population consisted of fetuses diagnosed with skeletal dysplasia with subsequent termination, stillbirth and live-born who died shortly after birth. Of the 2,002 autopsies performed, 112 (5.6%) were diagnosed with skeletal dysplasia. These 112 cases encompassed 17 of the 40 groups of Nosology 2010. The two most common nosology groups were osteogenesis imperfecta [27/112 (24%)] and the FGFR3 chondrodysplasias [27/112 (24%)]. The most common specific diagnoses were thanatophoric dysplasia (TD) type 1 [20(17.9%)], and osteogenesis imperfecta type 2 [20 (17.9%)]. The combined radiology, pathology, and genetics investigations and grouping the cases using Nosology 2010 resulted in a specific diagnosis in 96/112 of the cases.

2661M

Rare case of combination osteogenesis imperfecta and genetic skin disease. *N.M. Marycheva[1], J.Y. Kotalevskaya[2].* 1) Sechenov First Moscow State Medical University, Moscow, Russian Federation; 2) MV Vladimirsky Moscow Regional Research Clinical Institute, Moscow.

Osteogenesis imperfecta (OI) is a clinically and genetically heterogeneous disease characterized by fragility of bones and other short stature. OI XI type is a rare form of this disease, which is inherited in an autosomal recessive manner; mutations are described in the *FKBP10* gene. Epidermolysis bullosa (EB) is a group of inherited diseases of the skin manifested in formation of bubbles on the skin because of mechanical trauma. Mutations of simple type of EB are frequently identified in genes of keratin 5 and 14 (*KRT5*, *KRT14*). Here is a description of the case of the combination of OI type XI and simple type of EB. The proband is a 4 years old girl from the 4th pregnancy (in medical history previously was pregnancy with a girl with inherited EB). Clinical manifestations of EB appeared on the 2nd day and OI was diagnosed at the age of 1 month. She was admitted to the orphanage with a diagnosis of OI (fracture of the left humerus and of the right radius, multiple rib fractures in the medical history), the left hip dysplasia, EB of the simple type. Repeated cases of spontaneous fractures during the observation period happened. Examination: weight -12 kg, height - 99 cm, can't sit by her own, can't walk. Light blue sclera. Swarthy skin, multiple foci of de-and hyperpigmentation, blisters on the skin of the hands, elbow bends were observed. Feet hyperkeratosis. Pigeon breast, thighs saber deformity, scoliosis, flat feet, hypermobility of joints of the hands, limitation of movement in the hip and knee joints, muscular hypotonia. There were no other singularities in other systems. Analysis of the literature showed that there was one case of combination of OI type XI and simple type of EB, where were identified homozygous mutations in *FKBP10* and *KRT14* genes. We conducted the ArrayCGH to confirm the assumption of the similarity of our case with the previously described one and to narrow the candidate genes list for further genetic diagnosis. Results: stretches of homozygosity > 3,000,000 bp were identified in the proband. This value exceeds the population level10 times that may indicate kinship parents. One of the longest stretches of homozygosity located on the long arm of chromosome 17, where the genes *FKBP10* and *KRT14* are located. This result indirectly confirms our hypothesis about the possible mutations in the homozygous state in the genes responsible for the development of OI and BE. Sequencing of these genes in work.

2662T

INCONTINENTIA PIGMENTI: A Case Report Associated With Cleft Lip Palate in a Patient at Smile Operation Foundation, Bogotá - Colombia. *m. montiel, l. briceno, j. Martinez, A. Patino, J. Rincon.* medicine, universidad de la sabana, chia, Colombia.

Introduction Incontinentia pigmenti is a neuroectodermal disease, with an autosomal X-linked transmission, also known as Bloch Sulzberger syndrome (1). Epidemiology The Bloch syndrome is a genetic disease X -linked dominant inheritance, with a reported incidence of 1 case per 50,000 live births (3). Genetics They are two types of the disease: the sporadic and genetic IP. The gene involved in the development of IP1 is located in Xp11 chromosome (4). On the other hand, mutations that occur in the NEMO gene (essential modulator of nuclear factor Kb) located on chromosome Xq28 , is responsible for IP2 , with a dominant X -linked inheritance (4). Histology There are four stages of the disease, which are usually sequential. The inflammatory or vesicular phase. (5). Hypertrophic phase, (4). The third stage is known as hyperpigmentation phase (4). Finally, the atrophic phase has hypopigmetation atrophy or areas that tend to appear from adolescence to adulthood (4). Clinic About 80 % of the population with the disease have extracutaneous features of organs derived from neuroectoderm and neural crest in which the most common are dental, ophthalmological and neurological manifestations (3). There are cases of patients with cleft lip and minor manifestations of the disease. The main cutaneous manifestations reported are alopecia, pigmentary and nail changes (4). Case Report This case is about a 5 years old boy which presents a clinical manifestations of the disease like cleft palate, hypopigmented lesions on the abdomen and thorax associated with strabism, dental abnormalities, and alopecia. The case was diagnosticated at Operation Smile Foundation in 2011. Fig 1. The child presents hypomelanosis of Ito in the back Fig 2. The abdomen has hypomelanosis of Ito Fig 3. Associated cleft palate, dental abnormalities and alopecia. Diagnosis For the diagnosis of incontinentia pigmenti, there are some criteria simplified by Landy et al . (6) , but the management is based on an accurate physical examination which must be addressed to the search for skin disorders which are present in all patients, dental (absence of deciduous or permanent teeth, delayed in the eruption, deformations and conical crowns and hook) teeth, and eye CNS disorders (7). When the physician has the clinical suspicion, the histopathology analysis of the lesions is the method that identifies the type of injury and the stage at which the disease is found (7).

2663S

***MuSK* - a new target for lethal fetal akinesia deformation sequence (FADS).** *M. Wilbe[1], S. Ekvall[1], K. Eurenius[2], K. Ericson[1], A. Ameur[1], G. Annerén[1], M-L. Bondeson[1].* 1) Immunology, Genetics and Pathology, IGP, Uppsala, Uppsala, Sweden; 2) Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden.

Fetal akinesia is clinically and genetically heterogeneous disorders, with the common feature defined as reduced or loss of fetal movement. Several disease genes of fetal akinesia have been described. This includes genes involved in motor neuron development and survival, encoding components of the neuromuscular junction (NMJ), adult skeletal muscle proteins and fetal myostructural proteins. However, the genetic etiology of majority of cases with fetal akinesia is still unknown.

We report on a family with recurrent fetal loss, where the parents had five affected fetuses with fetal akinesia deformation sequence (FADS [MIM 208150]) and one healthy child. The fetuses displayed no fetal movements from the gestational age of 17 weeks, extended knee joints, flexed hips and elbows and clenched hands. There was polyhydramnion and no visible fetal stomach.

Whole exome sequencing (WES) of one affected fetus and the parents were performed. A recessive inheritance model was assumed and data was filtered against dbSNP, ANNOVAR and ~900 exomes in our in-house database. Only one candidate homozygous variant was identified in the fetus, c.40dupA (p.Thr14Asnfs*9), located in the first exon of *MuSK* (muscle, skeletal, receptor tyrosine kinase). The variant was verified using Sanger sequencing. Segregation analysis in the family revealed homozygosity for all affected fetuses, while the variant was not present in the healthy child. The c.40dupA variant leads to a frameshift in MuSK predicting a premature stop codon and likely removal of the resulting mRNA via nonsense-mediated decay pathway, causing the FADS phenotype. Haplotype analysis showed that the mutation most likely represents a founder-mutation.

MuSK is an agrin-dependent receptor tyrosine kinase required for formation of the neuromuscular junction and missense mutations in this gene have previously been described in congenital myasthenic syndrome (CMS). Interestingly, MuSK is located in the same acetylcholine receptor pathway as several other genes reported to cause CMS and/or FADS (*CHRNA1*, *CHRND*, *CHRNG*, *RAPSN*, *DOK7*, *CNTN1* and *SYNE1*). To our knowledge, this is the first time *MuSK* is reported to the spectrum of FADS and we propose that *MuSK* should be included in genetic analysis and prenatal screening for FADS.

2664M

DIAGNOSIS, PLANNING AND EDUCATIONAL EVALUATION IN GENETICS AS INTERACTIVE MATERIAL, WITH STUDENTS FROM A MEDICAL ODONTOLOGICAL UNIVERSITY AND MULTIDISCIPLINARY EVALUATION AT A PEDIATRIC HOSPITAL. turner, klinefelter, criduchat, down, Duchenne, Mucopolysacharidossis, Muscular dystrophy. *R. Aparicio[1,2], A. Torres Velandia[2,3].* 1) Medical Estomatology, Benemerita Universidad de Puebla; 2) Centro de Estudios e Investigación para el Desarrollo Docente AC, Guadalajara Jalisco; 3) Pedagogic Adviser, ICE-UAEM, México.

A Genetic Disorder is caused by abnormalities in genes or chromosomes. While some diseases, are caused by environmental factors. A great variety of Genetic and Disorders were studied in the Department of Genetics with Medical Students, such as; Cornelia de Lange, Dextrocardia, Kabuki, Duchenne Muscular Dystrophy, Achondroplasia, Albinismo, Osteogenesis Imperfecta, Neurofibromatosis, Down, Criduchat, Turner, Klinefelter, Marfan, Fetal Alcohol, Moebius, Ectodermic dysplasia, Goldenhar and Epiolia. As an reference, a study was performed in this hospital in a 20 years period of time, 6480 Karyotypes , Iontoforesis and metabolic studies (MS) among others have been done, in order to obtain a diagnosis, for a better treatment where the Medical Students learned in a more directly manner. A total of 2370 metabolic studies were performed and 102 clinical cases were associated to different inborn errors of metabolism (IEM). It was then analyzed both Teacher and students the great responsibility we have by learning that some diseases can be due to metabolic error for a protein or enzyme absence. It was important that the Students understood that ADN mutation exists, in relation to chromosome aberration, since a great variety of mutation in the patients were observed similar as Down, Turner and Klinefelter among others were evaluated to understad its pathogenesis in the Hospital. Medical Education policymakers need to seriously rethink many of the decisions that we are taking to promote educational change by changing some Paradigms, not only in this country but along the world. We believe that the project contributes to a change in university teaching health-related to directly assess patients with genetic or congenital abnormalities involved in this study. The porpoise of these knowledge strategies are to offer a better education for medical students to obtain an ability for earlier multidisciplinary medical evaluation for genetic diagnosis, and provide then better medical and therapeutic treatment. This study is performed as part of the academic program of Postdoctoral 2014 ;Diagnosis, assessment and educational planning; title given by the Centre for studies in research for teacher development (CENID). we considered important that both Teacher and Student have the ability to understand how important is the patient and his family that need us, in order to provide a better quality of life, for the different kind of patients according to their genetic disease.

2665T

Duplication of approximately 320 kb in the chromosomal region 7p15.1 in a girl with Peho-like phenotype, and in her normal father. *M. Giovannucci Uzielli[1, 5], G. Scarselli[2], M. Fichera[3], L. Castiglia[4], S. Baffini[5].* 1) Scienze della Salute, University of Florence, Florence (Italy); 2) Dept. of Paediatric Neurology, University of Florence (Italy); 3) Dept. of Medical Genetics, University of Catania (Italy); 4) Lab. Genetics Diagnosis, OASI Maria SS., Troina (Italy); 5) Genetic Science, Piazza Fra G. Savonarola, 11 Florence (Italy).

The increasing use of molecular tools in the diagnostic work-up of patients with syndromic and non-syndromic intellectual disabilities, is detecting the base of a growing number of birth defects. We report on a patient with a complex abnormal phenotype, characterized by congenital, peripheral, and persistent oedema, multiple anomalies, and profound mental retardation. The clinical data collected at birth and after a 24 years follow-up gave us the opportunity to well observe the specific aspects of her phenotype and natural history, suggesting the diagnosis of Peho syndrome (or better, Peho-like syndrome), a rare progressive infantile encephalopathy. So far, no candidate genes, no gene's location, no specific biochemical test to confirm the clinical diagnosis are reported in the literature for this rare disorder. Array-comparative genomic hybridization, using Human Genome CGH 60k Oligo Microarray kit (AMADID 21924, Agilent Technology) with Median of the spatial resolution 41.5 kb, showed an interstitial duplication of approximately 320 kb, in the chromosomal region 7p15.1 in both our patient and her normal father. This region encompasses the two genes CPVL and CHN2, also known as chimerin 2, a member of the chimerin family. Vitellogenic-like, CPVL is a carboxypeptidase, whose function has not yet been determined. Normal the molecular study in the mother. Conventional karyotype of the patient and both parents was also normal. The pathological meaning of this duplication is not therefore clear. Recently a similar duplication was included in the data base of Troina: unfortunately the description of the clinical phenotype observed in this patient is limited to "global difficulties".

2666S

An interstitial microdeletion of 4q21 in a girl with pituitary insufficiency associated with empty sella, epilepsy, severe growth impairment, and profound intellectual disability. *E. Nishi[1,2], K. Wakui[2], M. Arakawa[1], S. Hirabayashi[3], Y. Fukushima[2], T. Kosho[1,2].* 1) Dept Genetics, Nagano Children's Hosp, Azumino City, Nagano, Japan; 2) Dept Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Dept Pediatric Neurology, Nagano Children's Hospital, Azumino, Japan.

Microdeletion 4q21 syndrome (OMIM#613509) is characterized by severe intellectual disability, lack of speech, hypotonia, significant growth restriction, and distinctive facial features. The deletion sizes varied from 2.0 to 15.1 Mb, and the common features have been associated with a 1.37 Mb minimal critical region including five known genes: *PRKG2, RASGEF1B, HNRNPD, HNRPDL,* and *ENOPH1* [Mitchell et al., 1981; Nowaczyk et al., 1997; Harada et al., 2002; Friedman et al., 2006; Bonnet et al., 2010; Dukes-Rimsky et al., 2011; Lipska BS et al., 2011; Tsang E et al., 2012; Bhoj et al. 2013] . We report a girl with a de novo interstitial microdeletion of 4q21, who showed pituitary insufficiency associated with empty sella, epilepsy, severe growth impairment, and profound intellectual disability. The patient was the first child of a healthy 26 -year-old mother and a healthy 30-year-old non-consanguineous father. She was born by caesarean-section at 38 weeks and 6 days of gestation. Her birth weight was 2206g (-2SD), length was 43.4cm (-2.4 SD), and OFC was 32.8cm (-0.07 SD). She showed hypotonia and severe developmental delay. She raised her head at the age of 9 months, and walked unsupported in a wide gait at the age of 6 years. She also spoke no words, and profound intellectual disability. Brain MRI showed thinning of pituitary gland with empty sella. At the age of 10 years, she had severe growth impairment with the weight as 20.55kg (-1.8SD), the height as 112.3cm (-3.8SD), and the OFC as 49.1 cm (-2SD). She was found to have growth hormone deficiency and hypoplastic pituitary gland with empty sella, and growth hormone replacement therapy was started. G-banded karyotype was 46,XX, inv(12)(p12.3q13.3)mat. Microarray analysis (135K Oligo, Roche) revealed a de novo 2.16 Mb microdeletion at 4q21 [arr[hg18] 4q21.22q21.23(29,297,619-30,447,117)x1 dn]. The deleted segment encompassed 15 OMIM genes (*HNRNPD, HNRPDL, SCD5, SEC31A, THAP9, LIN54, PLAC8, COQ2, HPSE, HELQ, MRPS18C, FAM175A, and AGPAT9*), which shared only two genes (HNRNPD and HNRPDL) with the previously reported minimal critical region of microdeletion 4q21 syndrome. *HNRNPD* and *HNRPDL* might be the most contributing genes for cardinal features of microdeletion 4q21 syndrome comprising severe growth impairment, severe developmental delay with hypotonia, distinctive facial features, and profound intellectual disability with lack of speech.

2667M

The phenotypic variability of split hand and split foot malformation. *T. Yokoi[1,3], C. Hatano[1], T. Saito[2], J. Nagai[2], Y. Enomoto[1], K. Kurosawa[1], H. Ida[3].* 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Department of Clinical Laboratory, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Department of Pediatrics, The Jikei University School of Medicine, Tokyo, Japan.

Split-hand/foot malformations (SHFM) are a heterogeneous group of malformations, in which the hand and/or foot findings may occur in isolation or with other anomalies. There are five subtypes due to loci of chromosome in SHFM to date. SHFM occurs either sporadically or in families. There are autosomal inheritance and X-linked inheritance. In addition to each subtype's character, each patients has phenotypic variability. Here we report two different types of SHFM cases. Case1 is a-3-year-old-boy who has bilateral split hand showing a deep central cleft and bilateral split foot without any other anomalies and symptoms. His mother is also the patient of SHFM and has the similar features like her son. Case 2 is a-5-year-old-boy who has bilateral split hand showing a deep central cleft and bilateral split foot with developmental delay, sensorineural hearing loss and microtia. He has no families with any anomaly. Micro array analysis showed 430kb duplication of 10q24.21q24.32 including *SHFM3* in case1 and 400kb deletion of 7q21.3 including *DLX6* and *DLX5* in case2. These cases indicate the necessity to treat carefully SHFM patients and their families due to their phenotypic and genetic variability. Array Comparative Genomic Hybridization (array CGH) analysis can be considered as a method of choice for fast and accurate detection of unbalanced structural and numerical chromosomal abnormalities such as SHFM.

2668T

Clinical Implementation of Chromosome Microarray Analysis in Singapore. *H. Law[1], M. Brett[2], M. Yong[3], H. Yon[3], R. Roch[1], M. Tan[3], H. Ee[1], E. Tan[2], B. Cham[1], J. Lim[1], E. Tan[1], I. Ng[1], S. Jamuar[1], A. Lai[1].* 1) Paediatric Medicine, KK Women's & Children's Hospital, Singapore; 2) KK Research Centre, KK Women's & Children's Hospital, Singapore; 3) Department of Pathology and Laboratory Medicine, KK Women's and Children's Hospital, Singapore.

Background: Chromosome microarray analysis (CMA) is recommended as the first-tier genetic test for children with intellectual disabilities, development delay, autism spectrum disorder and/or multiple congenital anomalies. Although CMA is routinely available in USA, our lab only recently started offering it as a clinical test. Method: From May 2013 to May 2014, 100 patients were screened using the Agilent 4x180K CGH+SNP array. The indications for testing included developmental delay, intellectual disability, autism spectrum disorder and/or multiple congenital anomalies. Results: Copy number variants (CNVs) ranging in size from 10kb to 36.4 Mb were found in 38 patients (38%). Pathogenic and likely pathogenic CNVs were found in 18 (18%) patients. These included 14 deletions, 3 duplications and a patient with both a deletion and a duplication. Recurrent microdeletion and microduplication syndromes including the 1p36 microdeletion (2), Williams syndrome (2), Cat Eye syndrome, Cri du Chat syndrome, Miller Diecker syndrome, 3q29 microdeletion, 15q24 microdeletion, and 1q43q44 syndrome were detected in our patients. CNVs of uncertain clinical significance were detected in 20 (20%) individuals: 11 were duplications and 9 were deletions. However, as Singapore is a self-payer health care system, parental testing could not be performed in all 20 cases and hence, significance of these variants could not be established conclusively. Conclusion: CMA is a powerful tool in identifying pathogenic chromosomal copy number alterations. However, in countries with self-payer health care systems like Singapore, cost plays an important role in successful implementation of a diagnostic clinical test. In addition, more data relating to the local population needs to be collected to enhance interpretation of the CMA results.

2669S

Natural history and clinical management of patients with ASXL1 mutations and Bohring-Opitz Syndrome, including the first report of Wilms Tumor in two patients. *B. Russell[1], N. Kramer[2], L. Biesecker[3], J. Johnston[3], W. Rhead[4], A. Pickart[4], A. Dobson[5], L. Clarkson[5], J. Graham[2].* 1) Medical Genetics, Cincinnati Children's Hospital and Medical Center, Cincinnati, OH; 2) Medical Genetics Institute, Cedars Sinai Medical Center, Dept of Pediatrics, Harbor-UCLA Medical Center, David Geffen School of Medicine at UCLA, Los Angeles, CA; 3) National Human Genome Research Institute, National Institute of Health, Bethesda, MD; 4) Section of Medical Genetics, Children's Hospital of Wisconsin, Milwaukee, WI; 5) Greenwood Genetic Center, Columbia, SC.

Introduction: Bohring-Opitz syndrome is a rare genetic condition (35 reported cases) characterized by distinct facial features (glabellar nevus flammeus, wide spaced eyes, depressed and wide nasal bridge, anteverted nares, palatal anomalies, micrognathia, low-set posteriorly angulated ears), typical posture (elbow flexion with ulnar deviation and flexion of the wrists and metacarpophalangeal joints), variable microcephaly, severe intellectual disability, hypertrichosis, and feeding problems. Nine published patients with Bohring-Opitz syndrome have been identified as having a mutation in *ASXL1*. We report natural history and clinical management of five previously unpublished patients with mutations in *ASXL1*. Natural History: The 5 patients we discuss range from 2-12 years of age with complicated medical histories that included feeding issues, cyclic vomiting, respiratory infections, insomnia and Wilms tumor. Severity of illness improved with age after the first 1-2 years of life. Dysmorphic features such as nevus flammeus also faded with age. Severe myopia was present in all patients. They also had distinctive personalities (interactive, happy, and curious), and excessive hair growth which were features not described in the literature. Clinical Management: Permanent feeding tubes due to gastric dysmotility, silent aspiration, chronic emesis and poor weight gain were required in 4 of the 5 patients. Two patients had cyclic vomiting that was managed with cyproheptadine, lorazapam, ondansetron and acetaminophen. Recurrent respiratory infections with components of reactive airway disease occurred in 3 patients and 1 patient required a tracheostomy. Insomnia was a significant challenge for 4 of the patients, 2 of which improved with treatment of their severe anemia and 3 patients had obstructive sleep apnea that improved with CPAP or mandibular distraction. With the occurrence of bilateral Wilms tumor in one of the patients, screening renal ultrasounds were recommended to the other patients. This led to the identification and treatment of bilateral Wilms tumor in a second patient. Given the known association between *ASXL1* and myeloid malignancies along with the two patients presented here, consideration of Wilms tumor screening in patients with *ASXL1* mutations will be discussed.

2670M

Clinical and epidemiological study of orofacial clefts. *S. RASKIN, J. SOUZA.* Group for Advanced Molecular Investigation (NIMA), Graduate Program in Health Sciences, School of Medicine, Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, Paraná, Brazil.

Cleft lip with or without cleft palate (CL±P) or cleft palate (CP) are groups of malformations named orofacial clefts (OC), which are the second leading cause of birth defects. This study aimed to analyze clinical and epidemiological features of Brazilian patients with OC, studying cases treated in the reference center of the state of Paraná (PR). Methods 2,356 charts were reviewed and 1,838 were evaluated by the same clinical geneticist. Data were collected in the reference center, and compared with those of the Health Department of the state of Paraná. Clinical characteristics, presence of other anomalies, and birth prevalence were evaluated. 389 (21.2%) patients had CP, 437 (23.8%) had cleft lip (CL), and 1,012 (55%) had cleft lip and palate (CLP). Syndromic OC were identified in 15.3% of patients, 10.4% of patients with CL±P, and 33.9% of patients with CP. Common additional anomalies were: central nervous system, limbs, cardiovascular, and musculoskeletal defects. The number of syndromic cases was lower when clinical evaluation was performed by other medical specialists when compared to that of the clinical geneticist. Birth prevalence was 1/1,010 live births. Lack of notification with the national birth registry was observed in 49.9% of CL±P. The present data suggests a decrease of 18.52% in the prevalence of non-syndromic OC after folic acid fortification in Brazil. Conclusion Better understanding of clinical and epidemiological aspects of OC is crucial to improve the understanding of pathogenesis, promote preventive strategies, and guide clinical care, including the presence of clinical geneticists in the multidisciplinary team for OC treatment.

2671T

Lateral Meningocele (Lehman) Syndrome: A Rare Connective Tissue Disorder Craniofacial Dysmorphism. *M. Carter[1], S. Blaser[2].* 1) Pediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 2) Diagnostic Imaging, The Hospital for Sick Children, Toronto, ON, Canada.

Multiple lateral meningoceles are a rare anomaly. Lehman et al. (1977) first described a mother and daughter with similar craniofacial dysmorphisms, skeletal sclerosis, and multiple meningoceles. Subsequently, eight more cases have been described (Philip et al., 1995; Gripp et al., 1997; Chen et al., 2005 and Correia-Sa et al., 2013). The causative gene for Lehman syndrome is not yet known. The patient described in this abstract is a two-year-old boy with Lehman syndrome, born to a non-consanguineous Caucasian couple. Prenatally, a complex cardiac malformation was diagnosed (tubular hypoplasia of the aortic arch with coarctation, large ventricular septal defect, hypoplastic aortic valve, and bilateral superior vena cavae). MRI of brain showed middle cranial fossa encephaloceles and Chiari I malformation, with cerebellar tonsils extending to the posterior arch of C2. After discharge at 2 months of age, parents noted bilateral flank swelling. An ultrasound of the abdomen showed lateral lumbar meningoceles displacing the kidneys and causing the "swelling" due to deficient posterior wall musculature. Spine MRI showed enlarged spinal canal, and bilateral extensive lateral neural foramina meningoceles. The baby was discharged with NG tube in situ for feeding, and eventually required gastrostomy due to dysphagia. He was diagnosed with moderate bilateral conductive hearing impairment; hearing aids were prescribed. He has left amblyopia. Cardiac status has been stable. Dysmorphic features include left posterior plagiocephaly and tall cranial vault, sparse hair, hypoplastic supraorbital ridges, epicanthus, hypertelorism, ptosis, downslanted palpebral fissures, midface hypoplasia, small nares, long philtrum and thin upper vermillion, highly arched palate, bifid uvula, short upper lingual frenulum, asymmetric low set ears with short canals, significant microretrognathia, and bilateral single palmar creases. Neurological exam is significant for paucity of facial movement and diffuse hypotonia and joint hypermobility. Development at 13 months of age was significantly delayed. We present a child with Lehman syndrome, and review the cases reported to date. International collaboration will be useful for discovery of the gene(s) responsible for this complex condition.

2672S

A CHARGE syndrome-like phenotype in a patient with *EP300* splicing mutation. *S. Mizuno[1], Y. Tsurusaki[4], Y. Muramatsu[1], K. Maruyama[2], N. Niimi[3], K. Iio[3], N. Matsumoto[4].* 1) Dept Clinical Genetics, Central Hosp, Aichi Human Service Ctr, Kasugai, Aichi, Japan; 2) Dept Pediatric Neurology, Central Hosp, Aichi Human Service Ctr, Kasugai, Aichi, Japan; 3) Dept Pediatric Surgery, Central Hosp, Aichi Human Service Ctr, Kasugai, Aichi, Japan; 4) Department of Human Genetics, Yokohama City Graduate School of Medicine.

We report here the case of a female child with multiple congenital abnormalities and intellectual disabilities(MCA/ID) and growth retardation whose dysmorphic features and anomalies overlap those observed in CHARGE syndrome. In this patient we identified a de novo EP300 splice mutation by whole exome sequencing. The patient was the firstborn of nonconsanguineous parents(37-year-old father and 32 year-old mother). She was born at 37 weeks gestation by caesarian section because of polyhydramnios. Her birth weight and length were 2,350 g (- 1.4 SD) and 41 cm (-3.5 SD), and her occipitofrontal circumference was 32.5 cm (-0.5 SD). She was noted to have a congenital heart disease, esophageal hiatal hernia, small cleft palate, choanal stenosis, deformed earlobes, and congenital hydronephrosis at birth. She underwent several surgical procedures for correction of a large ventricle septal defect and third degree vesicoureteral reflex. At three years of age, she showed severe growth retardation, complete deafness, swallowing problems which required gastrostomy, strabismus, and scoliosis. Her facial features included a square outline, sparse hair, long eyelashes and severely malformed earlobe. She had moderate to severe intellectual impairment with developmental quotient score of 35 at 3 years of age. Although she showed hyperactive behavior, she enjoyed communicating with others. Because the combination of her malformations fulfilled the criteria of CHARGE syndrome, *CHD7* sequencing was performed. However no mutation was detected. Furthermore, microarray analysis did not reveal any pathogenic copy number variation either. By whole exome sequencing, we identified a de novo *EP300* splice mutation c.3671+5G>C in intron 20, which led to exon skipping. This mutation is considered to be responsible for the phenotype observed in this patient. *EP300* is known as the second gene to cause Rubinstein-Taybi Syndrome(RTS), but previous reports showed that only three percent of patients with RTS were found to have an EP300 mutation. The full phenotypic effects of *EP300* mutation remain unclear. Our patient displayed a number of complex anomalies, as well as growth and developmental retardation, which meet the CHARGE syndrome criteria. The only RTS associated phenotypes were curry eyelashes and marginally flat thumb tip. This suggests that , as well as RTS, mutation in *EP300* can lead to MCA/ID with diverse and broader phenotypic effects.

2673M

Detection of *GPC3* gene deletion by chromosomal microarray analysis in a patient with uncharacteristic Simpson- Golabi- Behmel syndrome. *A. Pietrzyk[1], K. Piotrowski[1], H. Chojnacka[2], M. Dera[1], M.B. Czeszynska[3], S. Zajaczek[1].* 1) Cytogenetic Unit, Department of Pathology, Pomeranian Medical University, Szczecin, zachodniopomorskie, Poland; 2) Department of Pediatric Surgery, Pomeranian Medical University, Szczecin, zachodniopomorskie, Poland; 3) Department of Neonatology, Pomeranian Medical University, Szczecin, zachodniopomorskie, Poland.

Introduction: Simpson- Golabi- Behmel syndrome (SGBS, OMIM 312870) is a rare X-linked overgrowth syndrome caused by loss-of function mutations or deletions in glypican- 3 gene *(GPC 3)*. Mapped at Xq26 it encodes a cell-surface heparan sulphate proteoglycan which has been indicated as a major factor in morphogenesis and modulator of cellular response to growth factors. The main clinical findings include: overgrowth, organomegaly, macroglossia, distinctive dysmorphic features, cardiac defects, renal and skeletal abnormalities as well as increased risk of embryonal cancers.

Case report: A four-year-old boy presented for genetic evaluation of dysmorphic features, development delay, and complex heart defect. Prematurity (with its complications), epilepsy, hypothyroidism were remarkable in his medical history. Neonatal cranial ultrasound revealed abnormal image of anterior horns of lateral ventricles. He required regular care of pediatric cardiologist due to pulmonary stenosis and ventricular septal defect. Upon clinical evaluation: trigonocephaly, “coarse” face, hypertelorism, macroglossia, supernumerary nipples, hypotonia, hepatomegaly, hypospadias and cryptorchidism were noted, which belong to SGBS spectrum. Neither prenor post-natal macrosomia as well as neoplasias were observed. Additionally, our patient suffered from mild optic atrophia since neonatal period and developed severe hyperopia. Parents: 25-year-old mother and 29-year-old father were healthy and family history was negative for congenital disorders.

Patient’s karyotype was determined as: 46, XY, inv (Y)(q11qter). Microarray analysis detected interstitial deletion of chromosome Xq.26.2 (395.74-444.63 kb, coordinates according to HG19: 132570975-132966712) containing critical region for SGBS and *GPC3* gene. Further tests showed that mother of our patient was a carrier of aforementioned microaberration.

Conclusion: Clinical manifestation of our patient closely resembles SGBS phenotype but some typical features were not present. Also, other symptoms that we observed couldn’t be found in the available data of similar cases. Confirmed genetic diagnosis has a great significance when a preventive therapeutic measures are possible as well as for a relevant genetic counselling in X-linked disorders considering its implications for reproduction.

2674T

A further case of ESCOBAR syndrome : Definition of novel mutation in CHRNG gene. *i. tekin[1], t. atik[2], h. onay[1], a. aykut[1], o. cogulu[2], f. ozkinay[2].* 1) Department of Medical Genetics, Faculty of Medicine, University of Ege, Izmir, Turkey; 2) Deparment of Pediatric Genetics, Faculty of Medicine, University of Ege, Izmir, Turkey.

Escobar syndrome (ES) is a rare autosomal recessive disorder which is also defined ‘nonlethal form of multipl ptergium syndromes’. ES is characterized by ptozis, low set ears, multiple ptergia, congenital contractures , arachnodactyly, short stature, scoliosis and cryptorchidism. Nonlethal Escobar variant is caused by homozygous or compound heterozygous mutations in CHRNG gene. Here we present a case who was considered to have Escobar syndrome with clinical features subsequent molecular analysis of her CHRNG gene showed a novel missense mutation. An 8-year old female patient born from a consanguineous marriage was referred to pediatric genetics subdivision because of dysmorphological findings recognised at birth. Her height was 113cm(3-10p), weight 22kg(10-25p), head circumference 50cm(<2p). She had downslanting palpebral fissures, hypertelorism, epicantal folds, high arched palate and multipl pterygia of limbs, flexion contractures and bilateral hypoplastic labia majora. DNA analysis revealed a homozygous novel p.P506S mutation in CHRNG gene. It was found a deteriorating mutation in silico analysis. This is the first Escobar case whose molecular analysis has been performed in Turkey.

2675S

De novo heterozygous deletion involving NFIX in a Japanese subject with severe intellectual disability, postnatal growth delay and relative macrocephaly. *D.T. Uehara[1], S. Hayashi[1], S. Mizuno[2], J. Inazawa[1].* 1) Department of Molecular Cytogenetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 2) Department of Pediatrics, Central Hospital, Aichi Human Service Center, Kasugai, Japan.

Intellectual disability (ID) is a genetically heterogeneous condition affecting 1-3% of the population, often associated with other clinical findings broadly referred to as multiple congenital anomalies (MCA). Among the genetic causes of MCA/ID, submicroscopic copy number variants (CNVs) are estimated to contribute to 15-20% of the cases. However, the etiology remains largely unknown in most cases. In order to identify pathogenic CNVs in Japanese subjects presenting with MCA/ID of unknown etiology, we have screened 646 cases using multiple microarrays for nine years. Here we describe one case, a female patient with a 394 kb de novo deletion at 19p13.2 involving NFIX gene (OMIM *164005), detected by SNP arrays. Clinical features include severe ID, hypotonia, small intestinal atresia, low stature (-3.5 SD) and weight (-2.5 SD) along with a relative macrocephaly (-0.7 SD) at the age of seven years old, although her birth size was normal. NFIX belongs to the nuclear factor one (NFI) family, which encodes four transcription factors essential for normal development. Previous reports have identified NFIX heterozygous deletions and point mutations in a few patients whose common features were ID, overgrowth and macrocephaly. A subset of those reports were described in patients with Sotos-like or Marshall-Smith syndromes, two overgrowth disorders characterized by advanced bone age, and in the latter case, skeletal anomalies. Our patient is the first description of haploinsufficiency of NFIX associated with postnatal growth delay. A possible explanation for that could come from two Nfix mouse models with marked differences in phenotypes: in one model, Nfix deficiency produced brain malformation and severe skeletal defects, while the other one primarily showed defects in brain development. Therefore, this present case might help clarify the actual role of NFIX in development. Nevertheless, the contribution of the other genes (LYL1, TRMT1, NACC1, ST×10, IER2 and CACNA1) encompassed by the rearrangement cannot be discarded, as well as other variants elsewhere.

2676M

Facial dysmorphism, skeletal abnormalities and central nervous system abnormalities in two sibs born to a consanguineous couple: A new autosomal recessive condition. *L. Chad[1, 2], M. Thompson[3], I. Miron[1], P. Shannon[3], S. Keating[3], D. Chitayat[1, 2].* 1) Department of Obstetrics and Gynecology, The Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, University of Toronto, Toronto, ON, Canada; 2) Department of Pediatrics, Division of Clinical & Metabolic Genetics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 3) Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada.

We report two female fetuses born to a consanguineous Sri-Lankan couple with facial dysmorphism, central nervous system and skeletal abnormalities. To our best knowledge this is a hitherto new autosomal recessive condition. The fetuses, both female, presented with thickened nuchal folds, echogenic bowel and kidneys, rocker-bottom feet, ventriculomegaly and intrauterine growth restriction. Detailed autopsies following termination of pregnancy at 23.4 and 22.3 weeks gestation respectively revealed short sloped forehead and hypertelorism with webbing of the neck, hydrocephalus with aqueduct stenosis as well as marked narrowing of the spinal canal and platyspondyly with delayed ossification and flattened acetabular roofs, broad hands with brachydactyly and narrow wrists. Microarray analysis was normal on both. These findings likely represent a new genetic syndrome with most probably autosomal recessive mode of inheritance. Whole genome sequencing is being done to try and identify the causative gene.

2677T

Paraspinal neurofibromas in LEOPARD syndrome. *E. Conboy[1], R. Dhamija[2], D. Babovic-Vuksanovic[2].* 1) Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN; 2) Department of Medical Genetics, Mayo Clinic, Rochester, MN.

LEOPARD Syndrome (LS) is an autosomal dominant disorder characterized by lentigines, EKG abnormalities, ocular hypertelorism, pulmonic stenosis, abnormal genitalia, growth retardation and deafness. There is significant clinical overlap between LS and other disorders that result from dysregulated Ras/mitogen-activated protein kinase (MAPK) pathway, including Neurofibromatosis type 1 (NF1) Noonan syndrome, Costello syndrome, Legius syndrome and cardio-facio-cutaneous syndrome. Except for NF1, other RASopathies are not known to be associated with development of neurogenic tumors. We describe two unrelated adult patients with clinical diagnosis of LS and massive paraspinal neurofibromas (dumbbell neurofibromas). Both patients were initially evaluated for NF1, but on the clinical exam they had lentigines, ocular hypertelorism, hearing loss, and positive family history lentigines. One of patients was found to have a heterozygous mutation (T468M) in the PTPN11. Dumbbell neurofibromas are an unusual complication of LEOPARD syndrome and may be an under-recognized manifestation of this disorder. We suggest surveillance for internal neurofibromas in patients with LS, since the risk for development of malignant peripheral sheet tumors in these patients may be increased.

2678S

Transcriptional hallmarks of Neurofibromatosis type I in whole blood cells. *G. Picco[1,2], F. Natacci[3], E. Trisolini[4], D. Cantarella[1], C. Cesaretti[3], G. Melloni[3], E. Riberi[5], F. Dutto[5], M. Cirillo Silengo[5], S. Vannelli[5], E. Medico[1,2], G.B. Ferrero[5].* 1) Department of Oncology, University of Torino, Torino, Italy; 2) Candiolo Cancer Institute, FPO IRCCS, Candiolo, Italy; 3) Medical Genetics Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy; 4) Department of Medical Science, University of Torino, Torino , Italy; 5) Department of Pediatrics, University of Torino, Torino, Italy.

Neurofibromatosis type 1 (NF1), an inherited neurocutaneous disease that has a major impact on the nervous system, eye, skin, and bone, is one the most common congenital non-chromosomal disorders affecting development and growth (1 in 3000 live births). This complex disorder is characterized by an extremely variable clinical presentation and Individuals with NF1 have a predisposition to benign and malignant tumor formation. Aggressive neoplasms as malignant peripheral nerve sheath tumors (MPNST) and juvenile myelomonocytic leukemia (JMML) can be observed in a significant fraction of patients. NF1 is caused by a wide spectrum of mutations affecting the NF1 gene, resulting in loss of function of the gene product neurofibromin, a negative regulator of Ras. These mutations reduce inhibitory activity of the protein, promoting Ras/MAPK pathway signaling, which regulates cell proliferation and differentiation by ultimately controlling gene expression. To investigate the transcriptional consequences of the aberrant activation of the RAS signaling driven by NF1 molecular lesions, we performed Global mRNA Expression Profiling (GEP) in human blood cells, a target tissue of the syndrome. In details, we analyzed 43 samples from molecularly defined NF1 patients and 17 additional samples collected from age- and sex-matched controls. Total RNA extracted from whole blood (PAXgene RNA collection tubes) was processed for expression profiling on Illumina Beadarrays. Subsequently, GEP analysis allowed the identification of a transcriptional signature composed by approximately 100 genes differentially expressed between NF1 cases and control samples. Interestingly, the gene expression patterns highlight a subset of genes subdividing NF1 samples in distinct subgroups. This evidences suggest that the transcriptome of NF1 patients presents clear elements of heterogeneity, possibly reflecting the clinical variability typical of this syndrome. These data establish peripheral blood expression profiling as a powerful tool to appreciate perturbations driven by germline mutations affecting the NF1 gene.

2679M

Fibrodysplasia ossificans progressiva (FOP): A case report. *I.M. Salazar-Dávalos[1], M.A. Aceves-Aceves[1], G. Pérez-García[2], M.I. Ornelas-Arana[2], D. García-Cruz[1,3], S.A. Alonso-Barragán[3], N.O. Dávalos[3], M. Salazar-Páramo[4], I.P. Dávalos[1,3].* 1) Instituto de Genética Humana, DGH, CUCS, Universidad de Guadalajara, Guadalajara, México; 2) Hospital Civil de Guadalajara "Fray Antonio Alcalde" Guadalajara, México; 3) Doctorado Genética Humana, CUCS, Universidad de Guadalajara, CIBO-IMSS, Guadalajara, México; 4) Depto. Fisiología, CUCS, Universidad de Guadalajara, Div. Investigación, UMAE, HE, CMNO , IMSS, Guadalajara, México.

Introduction: Fibrodysplasia ossificans progressiva (FOP) (OMIM #135100) is a rare autosomal dominant disease with complete penetrance caused by heterozygous mutation in the ACVR1 gene (102576) on chromosome 2q23. 95% of FOP cases are sporadic. FOP involves progressive ossification of skeletal muscle, fascia, tendons, and ligaments. FOP has a prevalence of approximately 1 in 2 million worldwide. (Petrie et al. 2009). The treatment is mainly preventive including avoidance of trauma, passive physiotherapy and prevention of chest infections. (Dhamangaonkar et al. 2013). Objective: To present a FOP male patient. Case report: The propositus aged 21 year-old presented, at age of 15 years, intense pain on right dorsal and lumbar region, later developed a painful lump on right costal region and limitation movement of right arm. He was diagnosed with FOP. Physical examination at present: weight 64kg, height 1.72 m, sparse eyebrows, cervical movement restriction, in posterior thorax presented three masses (2 x 2cm each), two on right scapula and one on the left scapula, scoliosis, restricted mobility on right arm and shoulder with restriction of pronation and supination, bilateral short great toes deformities with hallux valgus. He had repeated episodes of acute pain and inflammation of soft tissue, principally in scapular and pelvic girdle. The disease has been progressive with functional limitation. Laboratory studies gave normal or negative results. Thorax and abdominal CT reported new bone formation at right pectoral muscle, trapezium and gluteus medius. Conclusions: We present a sporadic FOP in a male patient. FOP is an incapacitating progressive disease the diagnosis is based mainly on clinical and imaging findings. The treatment is multidisciplinary and basically symptomatic, early diagnosis is fundamental to avoid aggravating factors and improve the quality of life.

2680T

Quantitative phenotype evaluation and management in osteogenesis imperfecta: Egyptian Experience. *M.S. Aglan[1], G.A. Otaify[1], R. ElHoussini[1], M.S. Abdel-Hamid[2], V.L. Ruiz-Perez[3,4], S. Temtamy[1].* 1) Clinical Genetics Department, Center of Scientific Excellence for Human Genetics, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; 2) Medical Molecular Genetics Department, Center of Scientific Excellence for Human Genetics, Human Genetics and Genome Research Division, Cairo, Egypt; 3) Instituto de Investigaciones Biome'dicas, Consejo Superior de Investigaciones Cienti 'ficas-Universidad Auto'noma de Madrid, Madrid, Spain; 4) Centro de Investigacio'n Biome'dica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III (ISCIII), Madrid, Spain.

Osteogenesis imperfecta (OI) is a heterogeneous genetic disorder characterized by bone fragility and fractures. Patients with OI have clinical features that may range from mild symptoms to severe bone deformities and neonatal lethality. We studied more than 150 referred patients with OI. Thirty percent of the referred patients were consistent with autosomal recessive (AR) inheritance indicating a higher percentage of recessive types in our community compared to Western countries as a result of a higher consanguinity rate in our population (20-40%). For clinical purposes, we proposed a quantitative scoring system for assessment of severity in OI. The clinical scoring system (CSS) includes five major criteria of clinical and prognostic value; number of fractures per year, motor milestones, long bone deformities, length/height standard deviation score (SDS), and z-score of bone density. Each criterion in the CSS was assigned a score from 1 to 4 according to standard definitions. Each patient is marked on a scale from 1 to 20 according to these five criteria. The CSS was assessed by its application on patients with different Sillence types of OI and proved to be an easy quantitative tool for assessment of disease severity. Molecular studies carried out for autosomal dominant (AD) and AR-OI patients resulted in new gene identifications and novel mutations in known AD and AR genes causing OI. The CSS applied to these patients, denoted that AR patients were significantly more severely affected than AD patients. Thirty three patients with OI were treated with cyclic biphosphonate injections in the form of zoledronate every 6 months with follow up for a period of 24 months. Significant improvement was noted in different parameters including the number of fractures, bone density, pain and motor milestones. Applying the CSS to these patients before and after therapy proved to be a good indicator for response to treatment. Our experience shows that applying the proposed CSS can quantitatively reflect the degree of severity in OI patients and can be used in complement with the Sillence classification and molecular studies. It also proved to be of significance in the assessment of response to treatment in OI patients.

2681S

Rare Cases of Congenital Arthrogryposis Multiplex without pterygium due to novel CHRNG Mutations. *J. Seo[1], Y. Yongjin[1], L. Youngha[1], K. Jung Min[2], S. Yong Beom[3].* 1) Biomedical Sciences, Seoul National University , Seoul, South Korea; 2) Department of Pediatrics, Seoul National University Hospital, Seoul, Korea; 3) Department of Rehabilitation Medicine, Pusan National University Hospital, Pusan, Korea.

Two unrelated patients visited outpatient clinic for evaluation of scoliosis and gait disturbance. The examination revealed multiple joint contractures in both upper and lower limbs without pterygium. Karyotyping of the proband was normal. Electrodiagnostic study did not reveal peripheral neuromuscular causes of the symptoms. Whole exome sequencing analysis of the patient genomes led to the discovery of identical missense variant (p.Pro143Arg) and frame-shift deletion (p.Pro251fs) on CHRNG, encoding acetylcholine receptor gamma subunit. They comprise rare cases of congenital arthrogryposis multiplex without pterygium related to the novel recessive CHRNG variants in two Korean subjects without apparent kinship. Further understanding of these mutations could achieve a better treatment and follow up plan.

2682M

A case of low frequent somatic and/or germline mosaicism in the ARSE gene detected by deep sequencing using NGS. *T. Kaname[1], K. Kurosawa[2], M. Higa[1], K. Yanagi[1], K. Naritomi[1].* 1) Dept Med Gen, Univ Ryukyus, Nishihara, Okinawa, Japan; 2) Kanagawa Child Medical Center, Kanagawa, Japan.

Chondrodysplasia punctata 1 (CDPX1) is an X-linked recessive disorder characterised by punctate calcifications in radiographs of the feet and other sites, nasomaxillary hypoplasia and brachytelephalangy. It is known that mutations of the arylsulfatase E (ARSE) gene affect the disorder. A male patient clinically diagnosed with CDPX1 was encountered. Examination of the ARSE gene by Sanger sequencing for all exons revealed a missense mutation, c.266G>A (p.S89N) in the patient. The mutation was not found in 250 healthy controls. Prediction of functional effects of the substitution by PolyPhen-2 and SIFT displayed that the mutation affects the protein function. In order to confirm whether his mother is a carrier, a molecular test was done as well. However, the Sanger sequencing could not detect the mutation in the mother, suggesting that the mutation was de novo. Then, we performed deep sequencing of the ARSE gene in the mother using a next generation sequencer, revealed that the mother had a somatic mosaicism of the mutation at 3.2% and 6% of allele frequency in the blood cells and buccal cells, respectively. We concluded that the deep sequencing using NGS could detect low frequent mosaicism, which might be a part of germline mosaicism. In such ‘de novo’ mutation, the deep sequencing should be performed to see whether the parents are mosaic in order to make reliable genetic counselling.

2683T

A further case of Hajdu-Cheney syndrome having a novel mutation in NOTCH2 gene. *A. KAVASOGLU[1], H. ONAY[1], M. ARGIN[2], F. OZKINAY[1].* 1) Department of Medical Genetics, Ege University Faculty of Medicine, Izmir, Turkey; 2) Department of Radiology, Ege University Faculty of Medicine, Izmir, Turkey.

Hajdu-Cheney syndrome is a very rare, autosomal dominant syndrome characterized by skeletal dysplasia, characteristic craniofacial and dental features, acro-osteolysis and proportionate short stature. We describe, here, a 35-year-old male patient with typical clinical and radiological features of Hajdu-Cheney syndrome. His height was 164 cm. He had typical facial features including hypertelorism, bushy eyebrows, micrognathia, dental anomalies, low-set ears, short neck and short fingers. X-ray studies showed wormian bones in the skull, acro-osteolysis of distal phalanges, short bowed long bones. On echocardiography, minimal mitral and aortic regurgitation were observed. Odiological examination revealed a conductive hearing loss. Regarding clinical findings, he was considered to have Hajdu-Cheney syndrome. Molecular analysis showed a heterozygous truncating c.6616 G-T (p.E2206X) mutation in the last exon of NOTCH2 gene. This Hajdu-Cheney case with a novel mutation is the first case whose molecular diagnosis was performed in Turkey and may help to establish phenotype-genotype correlation in the syndrome.

2684S

Expanding the diagnostic spectrum of terminal transverse limb defects: atypical mutations in *ACVR1* result in a phenotype with elements of Adams Oliver syndrome and Fibrodysplasia Ossificans Progresiva. *R. Mendoza-Londono[1], A. Al Maawali[1,2], L. Dupuis[1], F. Hyland[3], C. Scafe[3], T.A. Paton[4], C.R Marshall[4,5].* 1) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children and University of Toronto, Toronto, Canada; 2) Department of Genetics, Sultan Qaboos University Hospital, Sultan Qaboos University, Muscat, Oman; 3) Thermo Fisher Scientific, 200 Oyster Point Boulevard • South San Francisco • CA; 4) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 5) Division of Molecular Genetics, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada.

Reduction limb defects have an estimated incidence of 1 in 1692 live births, with terminal transverse limb defects (TTLD) being the most common. Diagnosis to consider in patients with TTLD include: amniotic bands, teratogenic exposures, Adams Oliver Syndrome (AOS), Poland Anomaly, Moebious and Oromandibular-limb hypogenesis syndromes. Vascular disruption has been thought to play a major role in all of these disorders. We present an 11 year old boy of Italian descent with TTLD, hypotrichosis patchy alopecia suggestive of cutis aplasia, intellectual disability and brain malformations. We performed Proton AmpliSeq whole-exome sequencing with a trio design and prioritized variants in genes known to be associated with AOS (*ARHGAP31*, *RBPJ*, *DOCK6* and *EGOT*). No pathogenic variants were identified in these genes. Further analysis of the sequence data revealed a *de-novo* variant in the *ACVR1* gene (c.983G>A; p.G328E) that resides in a well-conserved glycine residue in the protein kinase domain of ACVR1. Mutations in *ACVR1* have been previously associated with Fibrodysplasia Ossificans Progressiva (FOP), a syndrome characterized by heterotopic ossification of soft tissues. Most patients with FOP carry the same mutation in the glycine-serine rich domain (c.617G>A; p.R206H), have normal intelligence and great toe malformations. Only 3% of cases of FOP have atypical forms, and only 4 patients have ever been described with the G328E mutation. These patients present with TTLD, absent nails sparse scalp hair and cognitive impairment, supporting the causative role of this mutation. In addition the 5 other patients with atypical FOP and TTLD reported to date share mutations within the protein kinase domain (p.G328W and p.G328D). We analyzed the molecular pathways disrupted in AOS and postulate that the Cdc42/Rac1 regulator (*ARHGAP31* and *DOCK6* genes) and the Notch pathways (*RBPJ*) converge in a common signaling route that is critical for vasculogenesis. Activating mutations in ACVR1 in endothelial cells have been shown to cause endothelial-to-mesenchymal cell fate transition, which could result in abnormal vascular development. Together, this data suggests that the original theory of vascular injury and impaired arterial angiogenesis in AOS and TTLD is still valid. With this report, we expand the diagnostic spectrum of TTLD to include atypical mutations of ACVR1 and suggest a unifying pathophysiologic mechanism linking vascular disruption events to single gene mutations.

2685M

A novel homozygous mutation in FGFR3 causes tall stature, severe lateral tibial deviation, scoliosis, hearing impairment, camptodactyly and arachnodactyly. *S.A. Temtamy[1], P. Makrythanasis[2], M.S. Aglan[1], G.A. Otaify[1], H. Hamamy[2,3], S.E. Antonarakis[2,3].* 1) Clinical Genetics Department, Centre of Excellence for Human Genetics, Human Genetics & Genome Research Division, National Research Centre, Cairo, Egypt; 2) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 3) Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland.

Most reported mutations in the FGFR3 gene are dominant activating mutations that cause a variety of short-limbed bone dysplasias including achondroplasia and syndromic craniosynostosis. We report the phenotype and underlying molecular abnormality in two brothers, born to first cousin parents. The clinical picture is characterized by tall stature and severe skeletal abnormalities leading to inability to walk, with camptodactyly, arachnodactyly and scoliosis. Whole exome sequencing revealed a homozygous novel missense mutation in the FGFR3 gene in exon 12. The variant is found in the kinase domain of the protein and is predicted to be pathogenic. It is located near a known hotspot for hypochondroplasia. This is the first report of a homozygous loss-of-function mutation in FGFR3 in human that results in a skeletal overgrowth syndrome.

2686T

Longitudinal observation of clinical and radiological findings in a patient with Spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type caused by a heterozygous mutation in KIF22. *B. Tuysuz[1], S. Yilmaz[2], T. Erener-Ercan[1], C. Celen[2], K. Bilguvar[2], M. Gunel[2].* 1) Cerrahpasa Medical School, Department of Pediatric Genetics, Istanbul University, Istanbul, Turkey; 2) Department of Neurosurgery, Program on Neurogenetics, Yale School of Medicine, New Haven, Connecticut, USA.

Spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type (SEMDJL2), is a rare disorder, and characterized by postnatal short stature, midface hypoplasia and generalized ligamentous laxity. Radiological hallmark includes severe involvement of the epiphysis and the slender appearance of the metacarpals and phalanges. Recently, heterozygous missense mutations in KIF22 gene was reported as the underlying cause in SEMDJL2. The aim of the study was to evaluate of radiological and clinical findings of SEMDL2 during follow-up of a female patient. Evolution of the clinical and radiological findings of the patient from the age of 2 years and 9 months to 11 years was recorded. KIF22 gene was analysed by using whole-exome sequencing. We identified a single nucleotide de novo p.Pro148Leu mutation in KIF22 gene. The patient had postnatal short stature, midface hypoplasia, hip dislocation and generalised laxity of the joints. Walking difficulty due to knee subluxation and bilateral severe genu valgum became prominent after 3.5 years of age. Short stature became evident gradually with increasing age and height was at -3.6 SD at 11 years. Small epiphyses with delayed maturation, metaphyseal vertical striations at the distal metaphysis of femur were observed on initial radiographs. The slender metacarpals and proximal phalanges, progressive degeneration of epiphyses and narrowed epiphyseal plates on both wrist and knees became more prominent after 7 years. In conclusion typical radiological findings became apparent after early childhood. SEMDJL2 should be considered in a child who has short stature with joint laxity and midface hypoplasia.

2687S

Novel mutation in CLTC associated with multiple malformations and developmental delay. *J. DeMari[1], R. Miller[1,3], S. Tang[2], J. Nimeh[3], R. Lebel[1,3].* 1) Section of Medical Genetics, SUNY Upstate Medical University, Syracuse, NY; 2) Ambry Genetics, Aliso Viejo, CA; 3) Division of Pediatrics, SUNY Upstate Medical University, Syracuse, NY.

This female was large for gestational age at 35 weeks, delivered to a 27 year old primigravid Caucasian whose pregnancy was complicated by pre-eclampsia. Neonatal period was notable for hypoglycemia, apnea, bradycardia, hyperbilirubinemia, grade I intraventricular hemorrhage, subdural hematoma, laryngomalacia, hypotonia and feeding difficulties. She presented in the emergency department at 5 months of age with lethargy, emesis and MRI revealing progressive ventricular enlargement with cerebral atrophy; VP shunt was placed promptly. Etiology of hydrocephaly was unknown. She has central apnea and hypothyroidism. Vitamin-K dependent clotting factor deficiency (VKCFD1) was diagnosed, and is now under good control with high dose vitamin K supplementation. At 18 months of age, she was diagnosed with a neuroblastoma. She has numerous minor dysmorphic features. At two years of age, the patient has global developmental delays and nystagmus. Karyotype and oligo-microarray were normal. Whole exome genomic analysis was undertaken. Mutations in 9 genes were identified (14 mutations), and these genes underwent medical review at Ambry Genetics. A de novo, frameshift mutation resulting in a truncated protein was identified in the CLTC gene. CLTC encodes clathrin heavy chain 1 (CHC1), one of the components of clathrin, which serves as the vesicle coat protein involved in intracellular trafficking and endocytosis. Mutations in CLTC have not been previously identified in human disease. However, CLTC is expressed at high levels in the brain, and it has been shown that inactivation of CHC1 in rat and fruit fly models prevents the recycling or release of vesicles in the pre-synaptic terminal. Mutations in CLTCL1, the second member of the Clathrin Heavy Chain family, have been associated with neurological disease: seizures, intellectual disability, autism and schizophrenia. Patients with a deletion of 17q23.1, including CLTC, have been reported with microcephaly, developmental delay, and growth retardation. A single patient with non-syndromic hydrocephalus has been reported with a duplication overlapping the CLTC gene. We propose that this mutation in the CLTC gene is the cause of neurological disease in this patient.

2688M

Two cases of lissencephaly with marked hydrocephalus caused by *TUBA1A* mutation. *N. Ishihara[1,2], S. Yokoi[2,3], H. Yamamoto[2], J. Natsume[2], M. Tsutsumi[3], T. Ohye[3], M. Kato[4], S. Saitoh[5], H. Kurahashi[3].* 1) Dept Pediatrics, Fujita Health Univ Sch Med, Toyoake, Japan; 2) Dept Pediatrics, Nagoya Univ Grad Sch Med, Nagoya, Japan; 3) Div Molecular Genetics, ICMS, Fujita Health Univ, Toyoake, Japan; 4) Dept Pediatrics, Yamagata Univ Faculty Med, Yamagata, Japan; 5) Dept Pediatrics and Neonatology, Nagoya City Univ Grad Sch Med Sci, Nagoya, Japan.

Objective: *TUBA1A* encoding α-1a tubulin is one of the genes responsible for lissencephaly (LIS). Despite of wide spectrum of phenotypes associated with *TUBA1A* mutations, pathophysiology underlining the severity is still under discussion. Here we present two cases of LIS with marked hydrocephalus, the most severe phenotype of the spectrum, to clarify the function of mutated protein. Subjects and methods: Case 1 is a 2-year-old boy. Ventricle dilation was pointed out in fetal period. After birth, he was diagnosed as LIS. He had severe developmental delay with quadriplegia, but he had social smile and he could breathe and swallow by himself. At the age of 8 months, he began suffering from West syndrome, and visited our pediatric neurology office. Case 2 is a 3-year-old girl with marked hydrocephalus pointed out in fetal period. After birth, hydrocephalus rapidly progressed to become hydranencephaly, with cerebellar and brainstem hypoplasia. She did not have either spontaneous breathing or swallowing reflex, necessitating mechanical ventilation and gastrostomy for survival. Both patients were subject to exome sequencing for diagnosis. Results: Case 1 had c.74G>T (p.C25F) missense mutation in exon 2 of *TUBA1A*, and case 2 had c.190C>T (p.R64W) in exon 2 of *TUBA1A*. Both mutations were confirmed by Sanger sequencing. Mapping of the mutations on to 3D structure of microtuble complex showed both C 25 and R63 were not located at the contact surface of the tubulin interaction. Discussion: According to the spectrum of phenotypes associated with *TUBA1A* described by Kumar, case 1 is classified in group 4, most severe type of LIS, and case 2 is much more severe than those of group 4. Microtubules are polymers comprising tandem repeats of α/β tubulin heterodimers which assemble to form a sheet of longitudinal protofilaments, and lateral interactions between neighboring protofilaments cause the sheet to close, thereby forming the microtubule body. Lateral interactions allow microtubules to curve and bend without breaking. Microtubules are arranged in neuronal dense network, so lateral interactions may play an important role for maintaining structural integrity of microtubules especially in neurons. C25 and R64 are predicted to be associated with lateral interactions and mutations in the residues caused most severe phenotype, indicating the disruption in lateral interactions of tubulins may cause extremely severe malformation of the brain.

2689T

Comprehensive clinical characterization of VCP associated multisystem proteinopathy. *V.E. Kimonis[1], A. Surampalli[1], M. Khare[1], M. Wencel[1], C. Nguyen[2], S. Wigal[2], S. Graf[3], A. Wang[4], S. Donkervoort[1,6], M. Milad[1], T. Mozaffar[4], V. Caiozzo[5].* 1) Pediatrics, UC Irvine School of Medicine, Irvine, CA; 2) Child Development Center, Pediatrics, University of California, Irvine, CA; 3) Department of Pediatric Exercise Medicine, University of California, Irvine, CA; 4) ALS and Neuromuscular Center, University of California, Irvine, CA; 5) Department of Orthopedics and Physiology & Biophysics, University of California, Irvine, CA; 6) National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD.

VCP disease, also called multisystem proteinopathy is characterized by hereditary inclusion body myopathy with Paget's disease of bone and frontotemporal dementia (FTD) and other less common varied phenotypes including amyotrophic lateral sclerosis, Parkinson's, and cardiomyopathy,. TDP43 and ubiquitin positive inclusions are seen in affected tissues. We studied 41 individuals (24 affected/ 5 presymptomatic carriers/ 12 unaffected; M18/ F23, mean ages 50.8, 44.9, 49.7 y. respectively) in 8 families harboring three missense mutations, R155H, R155C and R155P; 14 with myopathy, 8 with myopathy and Paget's disease and 2 also with FTD. Average age of onset for myopathy and Paget's disease was 42.47 y. and 37.28 y. respectively. Functional capacity was measured using IBM Functional Rating Scale (IBMFRS), Fatigue Severity Scale (FSS), 6 minute walk test (6MWT); motor function by handheld dynamometry and MRC (Medical Research scale) scores. The mean IBMFRS scores for affected, carriers and unaffected relatives was 28.42±10.37, 38.25±2.36 and 39.83±0.40 (P=.01); MRC scores for all muscle groups was 203.47±64.11; 273.5±9.14; 279.50±1.0 respectively (P=.01) and the Fatigue score was 44.26±13.86; 34.50±24.77; 25.71±13.31 (P=.01 respectively). There was significantly reduced muscle strength in flexors of upper limbs (shoulder & elbow) and lower limbs (hip, knee and ankle); and extensors of knee equally on right and left limbs of affected group, however the strength in the carriers was similar to the unaffected relatives. There was a trend for lower average scores for 6MWT was in affected individuals when compared to carriers and unaffecteds. Spirometry revealed mean scores for FEF25-75% were found out to be significantly lower in affected when compared to controls (2.35±1.0; 4.4 ±0.14, 3.36±1.0, P=.03 for affected group) indicating the early involvement of small airways. Our data also showed that IBMFRS, fatigue scale and MRC scores associated with motor function alongwith FEF25-75% showed correlation and could be used to monitor the progression of respiratory and motor involvement in VCP disease. Early recognition of respiratory muscles involvement, using the functional rating scales could allow an early intervention in VCP patients. This study represents the most comprehensive evaluation of individuals with VCP disease to date which will help establish baseline studies for future clinical trials in this underserved population.

2690S

Identification of a novel variant in TMEM67 gene responsible for JBTS6 by whole exome sequencing. *A. Najmabadi[1], M. Hosseini[2], Z. Fattahi[2].* 1) University of California, San Diego, San Diego, CA., Select a Country; 2) Genetic Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

Joubert syndrome (JBTS) is a clinically and genetically heterogeneous disorder with autosomal recessive pattern of inheritance. The disorder is characterized by cerebellar hypoplasia, intellectual disability (ID), ataxia and oculomotor apraxia. Other clinical features include retinal degeneration, renal anomalies, hepatic fibrosis, and skeletal involvement. The hallmark of JBTS is a radiological pattern in magnetic resonance imaging (MRI), named "molar tooth sign". In 2007 Baala et al. identified a new form of Joubert syndrome designated JBTS6 with a causative mutation in the TMEM67 gene. Until now only 5 mutations in this gene have been identified as responsible for JBTS6. Here, we are reporting an autosomal recessive Iranian family with novel mutations in the TMEM67 gene. This family has two-affected child, one whom has profound ID, seizure, strabismus, and renal failure while the other child only shows signs of ID. Exome sequencing was performed for one of the probands and we were able to identify a number of variants. Following in silico analysis and checking our variant with 300 population specific controls, two candidate pathological changes remained; of which only one of these changes co-segregated in the family. Upon MRI examination of the affected boy, molar tooth sign was observed. We can conclude the missense changes in TMEM67 gene are responsible for Joubert syndrome in our family. Underlying causes of ID remain unknown in many cases because of clinical and genetic heterogeneity; therefor exome sequencing is an effective and helpful technique in detection of de novo mutation in this type of disorders. This approach results in more precise genotype-phenotype correlation and clinical diagnosis.

2691M

Truncating mutation of NFIA causes a brain malformation and urinary tract defect. *Y. Negishi[1], A. Hattori[1], K. Mizuno[2], I. Hori[1], N. Ando[1], F. Miya[3], T. Tsunoda[3], N. Okamoto[4], M. Kato[5], M. Yamasaki[6], Y. Kanemura[7,8], K. Kosaki[9], S. Saitoh[1]*. 1) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Science, Nagoya, Japan; 2) Department of Nephro-Urology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 3) Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 4) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 5) Department of Pediatrics, Yamagata University School of Medicine, Yamagata, Japan; 6) Department of Neurosurgery, Takatsuki General Hospital, Osaka, Japan; 7) Division of Regenerative Medicine, Institute for Clinical Research, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 8) Department of Neurosurgery, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 9) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan.

Chromosome 1p32-p31 deletion syndrome (OMIM #613735) involving the NFIA gene is characterized by corpus callosum hypoplasia or defects, hydrocephalus or ventricular enlargement, and urinary tract defects. Here, we report on the first case complicated by brain malformation and urinary tract defects observed with a truncating mutation in the NFIA gene. This male patient was born to healthy Japanese parents. Callosal agenesis had been suspected from the 28th gestational week. The boy was born by caesarean section on the 41st gestational week due to enlargement of the head circumference and post-term pregnancy. His birth weight was 3180g (+0.4SD), with a head circumference of 38.2cm (+3.3SD). Head MRI on the 3rd day of life revealed multilocular glioependymal cysts, ventricular enlargement, and callosal agenesis. Regarding his developmental milestones, he started walking without support at 1 year and 3 months. He was observed speaking meaningful words at 2 years, with a slight delay in language. Although sharp waves in the frontal head area were detected in EEG at 11 months, epileptic seizures have not been observed to date. Voiding cysturethrogram performed at 5 years of age showed bilateral grade IV vesicoureteral reflux. We performed whole exome sequencing on the proband and his parents, and identified a frameshift mutation (c.1093delC; p.P365fs) in the NFIA gene, which is absent in his parents, indicating that the mutation occurred de novo. Lu et al. reported 5 patients with balanced translocations or interstitial deletions of chromosome 1q31-q32 involving the NFIA gene. Recently, Rao et al. reported a case exhibiting a similar phenotype with an intragenic deletion of the NFIA gene confirmed by CGH microarray. This is the first report, to our knowledge, of a single nucleotide deletion (frameshift mutation) in the NFIA gene associated with brain malformation and urinary tract defects, confirming that the NFIA gene plays a fundamental role in development of brain as well as in urinary tract.

2692T

***De novo* 109 kb microdeletion of *MED13L*: report of a new patient with developmental delay, facial abnormalities and hypotonia.** *E.A. Repnikova[1,3,4], C.E. Lawson[2,4], S.D. Fiedler[1], J.M. Joyce[1], P.V. Thakor[1], L.D. Cooley[1,3,4], H.H. Ardinger[2,4]*. 1) Pathology Dept, Children's Mercy Hospital, Kansas City, MO; 2) Pediatrics Dept, Children's Mercy Hospital, Kansas City, MO; 3) Pathology Dept, University of Missouri-Kansas City Medical School, Kansas City, MO; 4) Pediatrics Dept, University of Missouri-Kansas City Medical School, Kansas City, MO.

The MED13L gene is one of the subunits of the large mediator complex that functions as a transcriptional coactivator for most RNA polymerase II transcribed genes and encodes a primary transcript of ~319 kb in size. MED13L is highly expressed in the brain, heart, skeletal muscle, kidney, placenta, and peripheral blood leukocytes. Missense mutations in *MED13L* have been linked to transposition of the great arteries and non-syndromic intellectual disability. Deletions in *MED13L* have been recently proposed to result in a distinct syndromic phenotype consisting of motor and speech delay, moderate intellectual disability, hypotonic open-mouth appearance and facial dysmorphism. Only three cases with various size deletions (17 kb (exon 2), 41 kb (exons 6-20) and 115 kb (exons 3 and 4) in *MED13L* have been recently reported. In this study, we performed molecular and clinical characterization of a patient with an approximate 109 kb *de novo* deletion of exons 5 through 31 of *MED13L* detected by oligonucleotide array comparative genomic hybridization. The patient is the third child of healthy non-consanguineous parents born at 39 weeks gestation after a pregnancy complicated by polyhydramnios. At the age of 2 years, she had mild hypotonia, wide-based gait, drooling, motor and speech delay, and dysmorphic features that included triangular face with tall forehead, sparse scalp hair (especially on the sides), extra subcutaneous tissue over glabella, sparse lateral brows, columella extending below alae nasi, prominent chin, and mildly uplifted ear lobes. Brain MRI showed nonspecific gliosis. At ages 4 and 6 years, she continued to have delays and poor balance. She had an excessive appetite with height and weight in the 90th percentile. Physical exam did not identify cardiac concerns. Similar to the previously reported cases, our patient has moderate intellectual disability, distinctive facial dysmorphism and hypotonia. However, she lacks cardiac anomaly, a phenotypic feature reported in cases with *MED13L* missense mutations and two cases with deletions of exon 2, 3 and 4. This observation supports the hypothesis recently proposed by van Haelst et al. that the cardiac phenotype may have reduced penetrance. The expression of cardiac phenotype may also be dependent on the type of mutation and its location within the *MED13L* gene. Our study thus adds to the increasing body of cases describing the existence of the novel *MED13L* haploinsufficiency syndrome.

2693S

Novel gene mutation in Schimmelpenning syndrome (nevus sebaceous syndrome). *Y. Kuroda[1, 2, 5], I. Ohashi[1], T. Naruto[1], Y. Enomoto[1], N. Okamoto[3], T. Niihori[4], Y. Aoki[4], K. Kurosawa[1]*. 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan; 2) Department of Pediatrics, University of Tokyo, Tokyo, Japan; 3) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Izumi, Japan; 4) Department of Medical Genetics, Tohoku University School of Medicine, Sendai, Japan; 5) Department of Epigenetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan.

We have identified a novel causative gene for Schimmelpenning syndrome. Schimmelpenning syndrome is a rare neurocutaneous disorder characterized by craniofacial nevus sebaceous with central nerve system, ocular, or skeletal abnormalities. Nevus sebaceous represents yellowish, waxy skin lesion and can undergo neoplastic degeneration in 24% in cases. Somatic mutations of HRAS and KRAS have been reported in patients of nevus sebaceous and Schimmelpenning syndrome. The proposita had nevus sebaceous at right frontal and vertex robe, angioma of the right postauricular skin, and bilateral preauricular tags. Ocular examination revealed retinal coloboma and macular hypoplasia of right eye. Head MRI revealed polymicrogyria and hemimegarencephaly of the right robe and bilateral cerebellopontine lipoma. She had also developmental delay and seizure. No mutation was detected by direct sequencing of exon 1 of HRAS in the nevus. Patient's sample from nevus was sequenced by MiSeq (Illumina Inc., San Diego, CA) by 121 bp pair-end reads, after the enrichment with TruSight Tumor. TruSight Tumor (Illumina Inc., San Diego, CA) was designed for targeted next-generation sequencing for 26 oncogenes and tumor suppressor genes, covering all coding regions, in total, 21 kb. Deep panel sequencing (average coverage of 15896) detected novel mutation of RAS-MAPK pathway gene with the nevus, but not with the lymphocyte, indicating postzygotic mutation. The mutation has been reported as the recurrent mutation in other neurocutaneous disorder. RAS promotes cell growth through activation of the mitogen-activated protein kinase (MAPK) signal transduction pathway. Activating mutations in the RAS-MAPK signaling pathway play a pivotal role in cancer-predisposition shown in epidermal nevus syndrome.

2694M

Neurofibromatosis type 1 and Optic Gliomas. *E. Parkhurst, S. Abboy.* Dept Genetics, Kaiser West Los Angeles, Los Angeles, CA.

Introduction: Neurofibromatosis type 1 (NF1) is a common autosomal dominant disorder that affects approximately 1 in 3,500 people. One of the most common tumors in children with NF1 is optic glioma with reported incidences ranging from 5-30%. Controversy surrounds the recommended frequency of eye exams and neuroimaging in asymptomatic individuals with NF1.

Methods: We queried the Southern California Kaiser Permanente electronic medical record database (Clarity) to find patients diagnosed with NF1 and seen in Opthalmology to determine the incidence of optic glioma. Medical records showing a diagnosis of optic glioma were reviewed.

Results: 708 patients under 21 years-old have a diagnosis of NF1 in the Southern California Kaiser system. 347 of these patients were seen in the opthalmology department and included in the study. Of the 347 with an opthalmology exam, 30 (8.6%) had a diagnosis of optic glioma. Average age of diagnosis was 5 yrs; the youngest was diagnosed at age 18 months and the oldest was diagnosed at 12 yrs. 30% (9/30) of patients with optic glioma required treatment other than corrective lenses; two patients had surgery alone and seven had chemotherapy. Treated patients had an age of diagnosis ranging from 19 months to 10 yrs. The youngest treated patient was asymptomatic and had 10 weeks of chemotherapy before 2 yrs. 60% (18/30) of patients found to have an optic glioma on MRI presented with symptoms (vision loss, proptosis, precocious puberty, etc.) and 40% (12/30) were asymptomatic. 63% (19/30) of the gliomas were bilateral; 23% (7/30) right sided and 13% (4/30) left sided. 53% (16/30) of the gliomas involved the optic chiasm. The four patients diagnosed with precious puberty all had chiasmal gliomas and were treated with chemotherapy.

Discussion: Current standards of care for children with NF1 include annual ophthalmologic examination, especially for children <6 years of age. Baseline MRI to detect asymptomatic optic glioma is not warranted. Our findings show that most optic gliomas are asymptomatic and never require treatment. Children with vision changes or precious puberty are most likely to require treatment such as surgery or chemotherapy, although very young asymptomatic children may have actionable gliomas as well.

2695T

Genetic heterogeneity in Van der Woude syndrome. *P. Kumari[1], A. Ali[2], SK. Singh[3], R. Raman[1].* 1) Cytogenetics Laboratory, Department. of Zoology, Banaras Hindu University; 2) Centre for Genetic Disorders, Banaras Hindu University, Varanasi; 3) G. S. Memorial Plastic Surgery Hospital and Trauma Center, Varanasi, India.

Van der Woude syndrome (VWS; OMIM 119300), an autosomal dominant disorder, is one of the most common syndromic forms of cleft lip and palate with lower lip pits as additional phenotype. *IRF6* (OMIM 607199) is the foremost candidate gene in several populations. *WDR65* (OMIM 614259), *GRHL3* (OMIM 608317) and 17p11.2-11.1 have also been identified to be in linkage with VWS, showing locus heterogeneity. In earlier studies from India on several cases *IRF6* is not found to be associated with VWS. Here we present results from genomic studies in two VWS families having 8 and 11 affected individuals, respectively. Initial analysis of *IRF6* did not reveal any mutation in the total coding region, and no association with the polymorphism, rs642961, in the enhancer region. Genome-wide linkage study (GWLS) using Affymetrix 10K SNP microarray in one of the families resulted in 3 linked loci (chromosome #1: q31.3-q42.2, #16: q23.1-q24.4 and #18: q11.2-q21.2) which had a LOD score of 2.4. Since microsatellite marker-based linkage analysis from the 3 high LOD score regions did not narrow down the mapped region, NGS was performed for whole genome with paired-end read which obtained a novel mutation in *NOL4* (OMIM 603577) that could affect its function by creating target site for a micro RNA. In the second VWS family, a haplotype involving an intron and two regulatory regions of *IRF6* co-segregated with the affected family members. *IRF6* haploinsufficiency being the major cause of VWS, its expression level was checked in blood of the affected and unaffected members of the family. Compared with the normal haplotype, *IRF6* expression was 2.27 fold lower in the risk haplotype. In silico analysis discovered a putative repressive transcription factor that would bind to one of the variants in the risk haplotype. Thus we not only identify a novel candidate gene (*NOL4*) but find a novel haplotype in *IRF6* as causal factor for VWS.

2696S

EXOMIC SEQUENCING AND MOLECULAR ANALYSIS OF IRF6 GENE IN PATIENTS WITH VAN DER WOUDE SYNDROME OR FAMILIAR HISTORY OF CLEFTING IN PATIENTS FROM SMILE BOGOTA COLOMBIA. *L. Patino[1], I. Briceno[2,3], j. martinez[2], D. mosquera[3].* 1) Bogota, Colombia; 2) Universidad de La Sabana; 3) Unmiversidad Javeriana.

The advent of technologies for massively parallel sequencing has resulted in thousands of sequenced genomes. The Van der Woude syndrome (VWS [MIM 119300]).is a skull - facial malformation characterized by the association of dimples or pits of the lower lip with cleft lip and / or palate. It represents the most common form of cleft lip and palate syndrome associated with two percent of the cases. With a prevalence in the general population of about 1 in 35,000 to 1 in 100,000. 3.4 The diagnosis can be confirmed by sequencing the IRF6 gene in 72% of cases. 5 However other mutations in different regions have been associated with the same clinical features. Among these mutations some have been described in the MSX1 gene, TP63 and FGFR120 -6. Moreover, recent studies have correlated clinical findings in patients with mutations in the gene GRHL3. 7.8 The description of the mutations of the Colombian population are of high importance since the Van der Woude syndrome is an entity that must be checked to give appropriate genetic counseling to the patient and family, as a vague genetic counseling will give them less than 50 % risk. Which is the proper genetic counseling due to its autosomal dominant inheritance. 2.3 In our study population coming from a reference center of cleft lip and palate treatment, we will be able to transfer our research to the Colombian population. In this population is where these mutations were sought by exomic sequencing. In patients with clinical suspicion of Van der Woude syndrome. We will transfer this data to a simpler and cheaper method of sequencing selected exons in which we have found a higher number of pathogenic mutations in our community. In individuals with a family history of cleft palate up to the second degree of consanguinity (this broader group represents another clinical spectrum of patients that do not fulfill the diagnostic criteria of Van der Woude syndrome). Performing the IRF6 gene sequencing sample of patients would get only 72 percent of the causal mutations of disease 5. But by exomic sequencing we should get better results to detect new mutations in different genes, which can be compared with the effectiveness of the technique IRF6 gene sequencing. So we can give a sufficient external validity to the results. Depending on the outcome, they will narrow or expand the range of mutations to be considered for the diagnosis of Van der Woude syndrome in our population.

2697M

WNT signalling and eye development disease genes. *I. Prokudin[1], V. Kumar[3], S. Davila[3], R. Jamieson[1,2].* 1) Eye Genetics Research Group, Children's Medical Research Institute; The Children's Hospital at Westmead; Save Sight Institute, Sydney, Australia; 2) Sydney Medical School, University of Sydney, Australia; 3) Genome Institute of Singapore, Singapore.

Microphthalmia and coloboma are rare genetic disorders often associated with other ocular abnormalities including anophthalmia, anterior segment dysgenesis and cataracts. A number of causative variants have been identified in genes including SOX2, OTX2, CH×10, BMP4 and RAX. However, for the majority of patients the disease genes are still not known. Low penetrance and variable expressivity are among the factors contributing to the low detection rate in microphthalmia and coloboma patients. In this study we performed whole exome sequencing on 4 affected individuals from two generations in an Australian family with autosomal dominant microphthalmia and coloboma. Heterozygous variants shared by all four affected family members were selected for the analysis. The variants were filtered based on their population frequency, pathogenicity prediction and conservation scores. Subsequently we prioritised variants in previously investigated animal disease genes and pathways known to be involved in eye disease. The top candidate variant, a frameshift deletion, was in a gene in the WNT signalling pathway. The variant was confirmed by Sanger sequencing and further processed for in vitro analysis using a WNT reporter assay in HEK293 cells. Our results indicate a role for the WNT signalling pathway in the microphthalmia and coloboma phenotype in humans.

2698T

A novel mutation in two patients with Fabry disease. *L. Wong-Ley, J. Medrano Valenzuela, A. Zambrano Parra.* Dept Gen, SSN-UAN, Tepic, Tepic, Nayarit, Mexico.

Fabry disease--a genetic disorder characterized by the accumulation of globotriaosylceramide in cell lysosomes resulting from an X-linked deficiency of α-galactosidase A activity--presents with multiorgan manifestations. We report, mother and daughter with Fabry disease, the mother of 36-year-old is asymptomatic and daughter 1 year old with neurological, renal, cardiac and ophthalmological disorders. At 4 months of age begins with proteinuria and recurrent infection of urinary tract, such as kidney problems. Later are noted cardiac (left ventricular enlargement and leaky heart valves), gastrointestinal (abdominal cramps, frequent bowel movements shortly after eating and diarrhea) and neurological (transient ischemic attacks) disturbances. Was performed gene sequencing (GLA HPLC/Tandem MS lyso-Gb3). The GLA gene was analysed by PCR and sequencing of the entire coding region and the highly conserved exon- intron splice junctions. Deep intronic mutations are not tested. The reference sequence of the GLA gene is: NM_000169.2. The biomarker lyso-Gb3 was measured by HPLC and tandem mass spectrometry. Was detected a previously unreported heterozygous variant in exon 6 of the GLA gene (c.968C>G p.P323R). It is located in a moderately conserved nucleotide and highly conserved amino acid position, with moderate physiochemical differences between the amino acids proline and arginine. Software analyses show inconsistent predictions: Mutation taster indicates this variant is probably damaging, whereas PolyPhen, SIFT and Align- GVGD predict toleration. The concentration of the biomarker lyso-Gb3 was normal. Our findings therefore suggest that the defective gene in the heterozygote has resulted from a new mutation. Genetic counselling is recommended for our patient and other relevant family members, also the detection of other possible asymptomatic carriers of the mutation.

2699S

Rasopathies and RAS/MAPK pathway disorders: Genetic screening of a cohort of 37 Tunisian children. *N. Abdelmoula[1], R. Louati[1], I. Trabelsi[2], S. Kammoun[2], M. Zenker[3], T. Rebai[1].* 1) Lab Histology, Univ Medicine, Sfax, Tunisia; 2) Dep Cardiology , Hedi Chaker Hospital, Sfax, Tunisia; 3) Institute of Human Genetics, University Hospital Magdeburg, Magdeburg, Germany.

Introduction: Recent studies have shown that a group of genetic disorders baptized RASopathies results from genes mutations of the Ras/MAPK pathway. These disorders include Noonan syndrome caused by mutations in PTPN11, SOS1, RAF1, KRAS, BRAF, NRAS and RIT1; Costello syndrome caused by activating mutations in HRAS; CFC syndrome caused by mutations in BRAF, MAP2K1/2 and KRAS; LEOPARD syndrome caused by mutations in PTPN11 and RAF1; Noonan-like syndrome caused by mutations in SHOC2 or CBL; Neurofibromatosis type I caused by haploinsufficiency of neurofibromin; NF-1 like syndrome caused by haploinsufficiency of SPREDI; Hereditary gingival fibromatosis caused by a mutation in SOS1; Capillary malformation-arteriovenous malformation caused by haploinsufficiency of RASA1 (p120 GAP). Because the underlying molecular mechanism for these syndromes is dysregulation of the Ras/MAPK pathway, the RASopathies exhibit numerous overlapping phenotypic features, including reduced growth, characteristic facial features, cardiac defects, cutaneous abnormalities, neurocognitive delay, and a predisposition to neoplasia, both benign and malignant. Materiel and Methods: Using HRM screening and bidirectional sequencing (of the whole gene or hot spot gene mutations) in all genes involved in Rasopathies (PTPN11, SOS1, RAF1, KRAS, BRAF, NRAS, RIT1, HRAS, BRAF, MAP2K1/2, KRAS, RAF1, SHOC2 or CBL), we have investigated a cohort of 37 Tunisian children. Results: A total of 12 mutations have been retrieved: 8 mutations of PTPN11, 2 mutations of SOS1, 2 mutations of Braf. No mutation of the new described gene; RIT1 was detected for negative Noonan patients. Neurobehavioral and dysmorphic features in patients with confirmed RASopathies developmental disorders have been compared to negative patients. Conclusion: Our findings indicate that mutations promoting dysregulation of the RAS-MAPK cascade mark an increased psychopathological risk and highlight that probably other new genes may be involved in Tunisian patients with mild, non-specific or atypical features. Using whole exome sequencing for molecular investigation of our negative patients will be a very powerful tool to infirm of confirm our hypothesis.

2700M

Prenatal and natal findings in a patient with Timothy syndrome type 1. *J.R. Corona-Rivera[1,2], R. Bloise[3], S. Priori[3], C. Napolitano[3], R. Nieto-García[2], E. Barrios-Prieto[4], L. Bobadilla-Morales[1,2], A. Corona-Rivera[1,2], E. Zapata-Aldana[2], C. Peña-Padilla[2], E. Chavana-Naranjo[2].* 1) Instituto de Genetica Humana "Dr. Enrique Corona-Rivera", Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) Centro de Registro e Investigación sobre Anomalías Congénitas (CRIAC), Servicio de Genética y Unidad de Citogenética, División de Pediatría, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca", Hospital-Escuela, Guadalajara, Jalisco, México; 3) Molecular Cardiology Laboratories, IRCCS Fondazione Maugeri, Pavia, Italy; 4) Unidad de Medicina Materno-Fetal, División de Ginecología y Obstetricia, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca", Hospital-Escuela, Guadalajara, Jalisco, México.

Introduction. Timothy syndrome (TS) or long QT (LQT) syndrome 8 (LQTS 8) (MIM #601005), is a rare multisystemic disorder characterized by cardiac defects (corrected QT (QTc) intervals of 480-700 ms, atrioventricular (AV) block, and congenital heart defects), hand and/or foot syndactyly, facial dysmorphism, and neurodevelopmental abnormalities, with almost 25 cases molecularly confirmed of *CACNA1C* gene mutations. Here, we report a propositus with TS type 1 who prenatally showed fetal hydrops due to congenital AV block, as a previously unreported fetal presentation. **Clinical report.** He was the first child of a 23-year-old mother. During the first 4 months of pregnancy the mother was treated with amitriptyline. The ultrasound at 34 weeks of pregnancy found complete AV block, and fetal hydrops (hydrothorax, ascites, and hydrocele). Delivery was carried out via cesarean in the 36th week of gestation. Apgar scores were 2, 7, and 7 at 1, 5, and 10 minutes, respectively. Birth weight was 3,400 g (>90th centile), length was 48 cm (75th centile), and occipitofrontal circumference was 36 cm (>90th centile). Physical examination showed macrocephaly, telecanthus, frontal hemangioma, abdominal distention and bilateral hydrocele. He showed a "mitten" hand with cutaneous syndactyly of the second to fifth fingers, and synonychia; cutaneous syndactyly of toes, and talipes equinovarus of right feet. A chest radiograph in the first day confirmed a bilateral hydrothorax, mild ascites, and 11 pairs of ribs. Echocardiogram showed a PDA and a PFO. The ECG confirms 2:1 AV block, and LQTS with prolongation of the QTc interval of 570 ms. The ECG and echocardiography were normal in both parents. . The parents were negative for targeted mutation analysis on *CACNA1C*. At the age of 3 months developed multiple episodes of ventricular tachycardia and fibrillation related to a gastrointestinal infection, and unfortunately died. **Conclusions**. The fetal hydrops can be attributed to the rythmogenic abnormality inherent to its TS type 1 and probably, they were exacerbated by maternal exposure to amitriptyline, a tricyclic antidepressant that can cause prolongation of QT interval. Despite the fatal outcome for our patient, prenatal diagnosis of TS may help to prevent life-threatening events or early death in future patients, especially in developing countries where availability of therapies such as cardioverter defibrillator is very limited, or required time for its sponsorship.

2701T

Antithrombin deficiency in a founder population: different genetic architectures for types 1 & 2. *P. Salo[1,2,3], M. Puurunen[4], J. Corral[5], S. .Engelbarth[4], K. Javela[4], M. Perola[1,2,6].* 1) Public Health Genomics, Natl Inst Health & Welfare, Helsinki, Finland; 2) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 3) Hospital District of Helsinki and Uusimaa, Finland; 4) Hemostasis Laboratory, Finnish Red Cross Blood Service, Helsinki, Finland; 5) Centro Regional de Hemodonación, Servicio de Hematología y Oncología Médica, University of Murcia, Spain; 6) Estonian Genome Center, University of Tartu, Estonia.

Hereditary thrombophilia due to antithrombin deficiency (ATD) is a rare autosomal dominant disorder. It is caused by mutations in *SERPINC1* coding for antithrombin, a central anticoagulant molecule in the blood. The affected individuals have an approximately 30-fold increase in risk for thrombotic complications such as deep vein thrombosis and pulmonary embolism. ATD is designated as type 1 when both antithrombin levels and its activity are reduced. In type 2 deficiency the level of antithrombin is normal but its activity is diminished. We characterized the genetic background of ATD in a sample containing the majority of patients (N=221) diagnosed with ATD in Finland using capillary sequencing, Sequenom genotyping, and multiplex ligation-dependent probe assays. The two ATD types have remarkably different genetic architectures in the Finnish population. Type 1 is dominated by allelic heterogeneity with 28 distinct mutations explaining the disease in the 36 families with type 1 ATD. Following from the requirement of decreased antithrombin levels for diagnosis with type 1 deficiency, these patients carried either frameshift-causing or exon-spanning indels, or point mutations affecting splice-sites or introducing premature stop codons. In stark contrast, only 5 distinct mutations were found in the 48 families with type 2 deficiency. A single mutation (p.Pro73Leu) explained almost 90% of the type 2 deficiency cases. The genetic structure of the Finnish population is a result of the small number of founding individuals, relative isolation, and slow population growth until a rapid expansion to the current size of 5.4 million inhabitants. This has reduced the allelic heterogeneity of most Mendelian diseases in Finland, where one or a few major disease-causing mutations typically explain most of the cases for a given disease. These major mutations were present in the population at the onset of the population growth and increased in number together with the expanding population. In the case of ATD, this phenomenon is strikingly limited to only type 2 deficiency, possibly indicating historically stronger negative selection against alleles causing type 1 ATD. Given that both types are malignant and their phenotypic differences are not large, this may have interesting implications for association studies of disease traits, including the search for low-frequency variants associated with complex diseases.

2702S

Novel mutation in SCN3A associated with multiple anomalies and encephalopathy. *S. Jhaveri[1], R. Miller[2,4], A. Braxton[3], F. Xia[3], J. Zhang[3], P. Ward[3], R. Dracker[4], R. Lebel[2,4].* 1) William Carey College of Osteopathic Medicine, Massena, NY; 2) SUNY Upstate Medical University, Section of Medical Genetics, Syracuse, NY; 3) Baylor College of Medicine, Houston, TX; 4) SUNY Upstate Medical University, Division of Pediatrics, Syracuse, NY.

Our patient was born at term to an 18-year-old primigravid Caucasian woman after an uncomplicated pregnancy with no suspected teratogen exposures; the union is not known to be consanguineous. Pierre Robin anomaly was apparent at delivery. Respiratory compromise required intubation at 14 days, and she underwent mandibular distraction and tracheostomy. Her course was complicated by cryptogenic pediatric partial seizures and apneic episodes. At age 24 months, she is notable for microcephaly, poor ponderal and longitudinal growth, intellectual disability, optic atrophy and hypotonia. Brain MRI has shown pachygyria, hydrocephalus, and hypoplasia of the corpus callosum. She has had multiple admissions, including for recurrent skin infections and autonomic storming. Nutrition is provided through a G-J tube. Family history is significant for cleft palate in the mother. Oligo-microarray revealed no anomalies. Whole exome analysis was undertaken at Baylor College of Medicine. A heterozygous novel, de novo, missense, likely pathogenic variant was discovered in the SCN3A gene. SIFT and Polyphen 2 predict the variant to be damaging. This gene encodes a voltage-gated sodium channel. There is only one previous report of a patient with cryptogenic pediatric partial epilepsy with a heterozygous missense variant in SCN3A. However, mutations in the voltage gated sodium channels (SCN1A, SCN1B, SCN2A, SCN9A) have a well known role in epileptic encephalopathy. We propose that this novel de novo variant could be the cause of encephalopathy in this patient. Functional studies would help further evaluate this likely pathogenic variant.

2703M

Evidence of germline mosaicism in Fibrodysplasia Ossificans Progressiva post discovery of the ACVR1 gene. *M.B. Alcausin[1,2], F.S. Kaplan[3,4,5], E.M. Shore[3,4,6], M.P. Baluyot[1,2].* 1) Institute of Human Genetics, National Institutes of Health, University of the Philippines-Manila, Manila, Philippines; 2) Department of Pediatrics, Philippine General Hospital, Manila, Philippines; 3) Department of Orthopaedic Surgery, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, Pennsylvania, USA; 4) Center for Research in FOP and Related Disorders, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, Pennsylvania, USA; 5) Department of Medicine, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, Pennsylvania, USA; 6) Department of Genetics, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, Pennsylvania, USA.

Fibrodysplasia ossificans progressiva (FOP) is a rare autosomal dominant disorder of connective tissue characterized by malformed great toes and by progressive heterotopic endochondral ossification in characteristic anatomic patterns. The transformation of soft tissue into bone causes disfiguration and inhibition of normal motor functions, which in most cases, can be severely debilitating. Most affected individuals have FOP as a result of a spontaneous new mutation but the occurrence of germline mosaicism has previously been suggested in a report of a family with two affected siblings born to unaffected parents. We report the occurrence of FOP in a Filipino family with two affected brothers who have the classic FOP mutation ACVR1/ALK2 c.617G>A; R206H. Their parents are phenotypically and genotypically unaffected. This provides further evidence of the occurrence of germline mosaicism in this condition which impacts on recurrence risk assessment. In other autosomal dominant genetic disorders for which germline mosaicism has been established, recurrence risk for future pregnancies has been estimated at 5-6%. This is the first reported family with this genetic test result after the discovery of the causative gene, ACVR1.

2704T

Multi-systemic Involvement in NGLY1-related disorder Caused by Two Novel Mutations. *J. Heeley, M. Shinawi.* Department of Pediatrics, Division of Genetics and Genomic Medicine, Washington University in St. Louis School of Medicine, St. Louis, MO.

Background: NGLY1-related disorder is a newly described autosomal recessive condition characterized by neurological (global developmental delay, severe hypotonia, movement disorder, seizures or EEG, acquired microcephaly, diminished reflexes and nerve conduction abnormalities), hepatic (neonatal jaundice, elevated liver enzymes, fibrosis, intrahepatic cytoplasmic inclusions), ophthalmological (alacrima/hypolacrima, strabismus, chalazion, apraxia) findings and associated with dysmorphic features, constipation and scoliosis. The gene encodes an enzyme, N-glycanase 1 [MIM 615273], involved in deglycosylation of glycoproteins, an essential step in endoplasmic reticulum-associated degradation pathway. It was hypothesized that loss of function of NGLY1 causes accumulation of misfolded glycoproteins in the cytoplasm. The condition has been described so far in 8 patients. Methods: Clinical and molecular characterization of a proband with novel mutations in the NGLY1 gene detected via whole exome sequencing. The findings in the proband are compared with previously reported cases. Results: The proband is a 13-year-old boy with profound hypotonia and elevated transaminases diagnosed in infancy. Liver biopsy at age 1 year showed lipid accumulation with dilated endoplasmic reticulum. The patient exhibits severe global developmental delay and is nonverbal and non-ambulatory. He has severe scoliosis and osteopenia. He has acquired microcephaly after 10 months, involuntary movements, poor weight gain, muscle atrophy, absent reflexes, and seizures. He has had multiple procedures for lacrimal duct stenosis and strabismus and has intractable blephiritis. We also noted persistent hypocholesterolemia, which was not previously reported. Whole exome sequencing revealed two novel variants in the NGLY1 gene: a maternally inherited truncating mutation, c.347C>G (p.S116X), and paternally inherited splicing mutation, c.881+5G (p.IVS5+5G>T), predicted to abolish the splice donor site of exon 5. Neither change was previously reported in NHLBI. Conclusion: Mutations in the NGLY1 gene cause a severe, multisystemic but recognizable phenotype. This study along with previously reported cases suggests that targeted sequencing of the NGLY1 gene should be considered in patients with the typical combination of neurological, hepatic, and ophthalmologic findings. Our data reveal that osteopenia and hypocholesterolemia can be part of the phenotypic spectrum of NGLY1-related disorder.

2705S

Is SMN2 related to severity in Spinal Muscular Atrophy? : a case report. *PM. Hurtado, AI. Sanchez.* Ciencias Básicas de la Salud, Pontificia Univ Javeriana, Cali, Colombia.

Introduction: Spinal muscular atrophy (SMA) is an autosomal recessive neuromuscular disease characterized by progressive degeneration and loss of the anterior horn cells in the spinal cord leading to progressive muscle weakness and atrophy. Incidence is 1 in 6,000 and carrier frequency is 1 in 35. It is caused in 96% of patients by homozygous absence of the survival motor neuron gene (SMN1). Copy number of SMN2 has been proposed to determine severity of the disease. SMA is classified by age of onset and severity of the disease. Case Presentation: This is a 15 years-old (y/o) male patient who was born at first gestation from non-consanguineous parents. Mother was 18 y/o. Patient started to have frequent falls and trouble walking up and down stairs at 4 years. Progressive muscular weakness was evident first in legs and then in arms. After that, he started to have difficulty to solid swallowing. At 6 y/o, patient was not capable of deambulate anymore. He has had presented respiratory failure in several opportunities, requiring hospitalization. He has a 13 y/o brother with similar phenotype but with symptoms starting after 4 y/o. Physical examination revealed generalized muscular hypotrophy, tongue twitching and knee retractions. Muscular biopsy showed muscular dystrophy. Spinal cord MRI evidenced left lumbar scoliosis and paravertebral muscle atrophy. MRI of the brain was normal as well as EKG. CPwK levels were mildly elevated. Polysomnography evidenced obstructive sleep apnea. Leg electromyography showed demyelinating axonal neuropathy. Quantitative PCR evidenced no copies of the SMN1 gene, but ≥ 3 copies of SMN2. Discussion: This individual is affected with SMA type III because onset was after age ten months and no copies of SMN1 were seen at molecular testing. It is known that SMN2 genes are capable to produce a protein identical to that of the SMN1 gene but at reduced capacity (10-20%). As a result, more than 3 copies of SMN2 are related to milder severity of the illness. Despite the anterior fact, muscular weakness in this patient was rapidly progressive and severity was significant. This case is also noteworthy because patient's brother quantitative PCR showed SMN1 exon 7 and 8 homozygous deletion. This SMN1 homozygous deletion is typically detected by demonstrating the absence of exon 7 (≈96% of SMA patients). Conclusions: Further studies are necessary to elucidate mechanisms of severity of the disorder and to promote development of new treatments.

2706M

Stargardt Disease (Juvenile Macular Degeneration), Clinical Analysis in Patients of the Colombian Population. *L.Ma. MORA[1,2], F.J RODRIGUEZ[2], M. VALENCIA[2], M.L TAMAYO[1,2].* 1) CLINICAL GENETICS, PONTIFICIA UNIVERSIDAD JAVERIANA, INSTITUTO DE GENÉTICA HUMANA, BOGOTA, COLOMBIA; 2) CLINICAL GENETICS, FUNDACION OFTALMOLOGICA NACIONAL (FUNDONAL), BOGOTÁ, COLOMBIA. M.D.

INTRODUCTION. Stargardt disease or Juvenile Macular Degeneration, initially described in 1909 by Karl Stargardt, is known as a nosological entity of autosomal recessive inheritance. It is the most frequent cause of macular degeneration in childhood, due to the accumulation of lipofuscin in the endothelium of retinal pigmentation and the limited energy exchange in the photoreceptors; clinically manifested as a progressive deterioration of visual acuity (VA), preserving the peripheral vision. It has been reported a global incidence of 1 per 10.000; however not local epidemiology is available in Colombia. RESULTS. 1. Ten patients had best corrected (VA) major 20/100. 2. 8 of the evaluated are in stage I/II. 3. Inverse correlation was found between foveal thickness and VA (Fig.3). 4. The symptoms start in the first two decades of life for 9 of the evaluated. 5. As the worldwide information, autosomal recessive mechanism of inheritance is the most frequent in the Colombians. 6. Night-blindness and reduced vision are the most common initial symptoms. DISCUSSION. Stargardt disease is a progressive disease, which affects individuals in the first decade of life. The clinical and paraclinical characteristics are unknown in Colombian population. We made a description of ten individuals in collaboration with the Instituto de Genética Humana at the Pontificia Universidad Javeriana and the Fundación Oftalmologica Nacional, it allowed us to establish for our population the most frequent symptoms, night-blindness and low vision with a slow progression. The majority of the patients are in stages I and II, this being compatible with a VA of major 20/100; with an average time of evolution of 20 years. The findings on the Optical coherence tomography (OCT) show reduced thickness and loss of photoreceptors in the macular area, which may be useful in the diagnosis of Stargardt disease, because strong correlation was found between the foveal thickness and visual function in patients in the sample.

2707T

Atrophic skin patches with abnormal elastic fibers, as a presenting sign of MASS phenotype associated with mutation in the Fibrillin-1 gene. *E. Reinstein[1], R. Bergman[2], J. Nevet[2], H. Shosany[2].* 1) Medical Genetics Inst, Beilinson Medical Center, Petach-Tikva, Israel; 2) Dermatology dept, Rambam Medical Center, Haifa, Israel.

Marfan syndrome (MFS) is a dominantly inherited disorder of connective tissue caused by mutations in the Fibrillin-1 (FBN1) gene. The most common skin finding in MFS is striae distensae. Particular individuals referred for suspicion of MFS whom do not completely fulfill the MFS diagnostic criteria are classified as having a MASS phenotype. The acronym represents the phenotype's apparent manifestations: a prolapsed Mitral valve, Myopia, Aortic root enlargement, Skeletal and Skin manifestations. Mutations in FBN1 have been shown to be associated in a few cases with MASS phenotype. Skin manifestations may be an important clue to the diagnosis of these disorders. We studied a case referred for unusual atrophic skin patches on the buttocks. Histopathology and electron microscopy demonstrated markedly abnormal elastic fibers. Consequent medical genetics evaluation led ultimately to the diagnosis of MASS phenotype, and to the discovery of an underlying FBN1 mutation. Though the clinical suspicion and diagnosis of MFS and related disorders are usually established by its main associated clinical features including the eye, skeletal and vascular involvement, clinicians should be aware of the associated skin manifestations, including unusual atrophic patches with abnormal elastic fibers that can sometimes be the first noted sign of the genetic disorder.

2708S

Congenital ophthalmoplegia: Dysinnervation, myasthenia or myopathy? Can genetics have an answer? *S. Shaaban[1,2,4,5,7,8], W. Chan[1,4,5,6,8], C. Andrews[1,4,5, 6,8], E. Engle[1,3,4,5,6,7,8].* 1) Department of Neurology, Boston Children's Hospital; 2) Department of Ophthalmology, Mansoura University, Egypt; 3) Department of Ophthalmology,Boston Children's Hospital; 4) F. B. Kirby Neurobiology Center, Boston Children's Hospital; 5) Program in Genomics, Boston Children's Hospital; 6) Howard Hughes Medical Institute, Chevy Chase, Maryland; 7) Dubai Harvard Foundation for Medical Research; 8) Manton Center for Orphan Disease Research, Boston Children's Hospital.

Congenital cranial dysinnervation disorders (CCDDs) are congenital, non-progressive, sporadic or familial disorders that result from developmental abnormalities of one or more cranial nerves/nuclei and are often characterized by abnormal eye, eyelid, and/or facial movements. Previous work in our lab has identified gene mutations that underlined a set of these CCDDs. In a subset of the patients we studied; the congenital fibrosis of extraocular muscle (CFEOM) was associated with a constellation of findings in other systems (e.g E410K TUBB3 syndrome in which patients in addition to CFEOM and facial palsy had developmental delay, vocal cord paralysis, stridor, tracheomalacia and abnormal EMG/NCS). Interestingly in a subset of syndromic patients presenting to us with CFEOM, facial palsy, gross motor delay, hypotonia, bulbar dysfunction, feeding problems and episodic apnea that were mutation-negative when tested for known CCDDs genes; recent work in our lab using next-generation sequencing (NGS) identified disease-causing mutations in four pedigree. We identified mutations in three genes known to cause congenital myasthenia (*CHRND, CHRNE* and *SCN4A*) in three pedigrees and mutations in *NEB* which causes Nemaline myopathy in the fourth pedigree. The mutations were two novel compound heterozygous missense mutations (c.308T>C; p.L103P and c.991C>T; p.L331F) in *CHRND* (cholinergic receptor, nicotinic, delta), a recurrent frameshift mutation (c.130dupG; p.P45Afs*2) in *CHRN*E (cholinergic receptor, nicotinic, epsilon), a novel missense mutation (c.2890C>T; p.V964I) in *SCN4A* (sodium channel, voltage-gated, type IV, alpha subunit) and two compound heterozygous missense mutations (c.18164T>C; p.M6055T and c.7441A>G; p.R2481G) in *NEB* (Nebulin). The overlap of phenotypes resulting from mutations in some of the CCDDs genes or those causing congenital myasthenia or congenital myopathies makes it challenging to differentiate between them at the clinical level. These results delineate a spectrum of genetic disorders that overlap phenotypically, yet can result from mutations in different genes disrupting any point from the nervous system, to the neuromuscular junctions to muscles' structure and/or function. NGS proves to be a powerful tool that could help physicians and geneticists reach a definite diagnosis when clinical findings are not as conclusive.

2709M

Chromosome deletion 11q13.1 involving deletion of the CLCF1 gene in a female with features of Cold-Induced Sweating Syndrome. *J.D. Weisfeld-Adams, K.E. Brown.* Department of Pediatrics, University of Colorado, Denver, CO.

The CLCF1 gene (cardiotrophin-like cytokine factor 1) is known to be implicated in some cases of autosomal recessive Cold-Induced Sweating Syndrome (CISS). CISS, a designation used synonymously with "Crisponi syndrome", is characterized by paradoxical sweating response, high-arched palate, scoliosis, facial weakness, prominent ears, feeding difficulties in infancy, hyperthermia, and velopharyngeal insufficiency or hypernasal speech. CLCF1 is a member of the glycoprotein 130 cytokine family and encodes cardiotrophin-like cytokine factor 1 (CLCF1). CLCF1 partners with cytokine receptor-like factor 1 (CRLF1) to form the CRLF1/CLCF1 protein complex, with CLCF1 mutations accounting for significantly fewer historical CISS cases than CRLF1mutations. The complex binds to ciliary neurotrophic factor receptor (CNTFR) and activates the Jak-STAT signaling cascade. CLCF1 is a potent neurotrophic factor, B-cell stimulatory agent and neuroendocrine modulator of pituitary corticotrophic function. Mutations in the CLCF1 gene may disrupt CNTFR signaling leading to disruption of sympathetic nervous system development, and resultant dysregulation of temperature homeostasis, manifesting as hyperthermia and pathologic sweating response. The CNTFR pathway is also implicated in motor neuron and skeletal development which may explain facial dysmorphology, facial weakness and skeletal abnormalities observed in some affected individuals. Here we report a 10 year old female with a history of frequent, unexplained fevers, lack of sweating with activity, feeding difficulties, short stature, scoliosis, subtle facial dysmorphism, and hypernasal speech. Chromosomal microarray analysis revealed a heterozygous de novo deletion at 11q13.3, of approximately 0.153 Mb in size and incorporating CLCF1. We believe this patient likely has a point mutation on the non-deleted CLCF1 allele. Research testing with sequencing of CLCF1 is currently underway. To our knowledge, this is the first CISS patient reported with a genomic rearrangement incorporating CLCF1.

2710T

Functional studies of *EZH2* histone methyltransferase activity in Weaver syndrome. *A.S.A. Cohen[1,2], D. B. Yap[3,4], J. Denny[5], S. M.E. Lewis[1,2,6], C. Chijiwa[1,6], M.A. Ramos-Arroyo[7], N. Tkachenko[8], V. Milano[9], M. Fradin[10], C. J.D. Ross[1,2,5], W. B. Dobyns[11,12,13], D. D. Weaver[14], S. Aparício[3,4], W. T. Gibson[1,2].* 1) Dept of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Child and Family Research Institute, Vancouver, BC, Canada; 3) Dept of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada; 4) Dept of Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, BC, Canada; 5) Centre for Molecular Medicine and Therapeutics, Vancouver, BC, Canada; 6) British Columbia Children's and Women's Health Center, Vancouver, BC, Canada; 7) Dept of Medical Genetics, Complejo Hospitalario de Navarra, Pamplona, Spain; 8) Medical Genetics Unit, Centro de Genética Médica Dr Jacinto Magalhães, Oporto Hospital Center, Porto, Portugal; 9) Instituto di Genetica Medica, Università Cattolica del Sacro Cuore, Policlinico Universitario Agostino Gemelli, Roma, Italy; 10) PH Génétique Clinique, Hôpital SUD, Rennes, France; 11) Center for Integrative Brain Research, Seattle Children's Hospital, Seattle, WA, USA; 12) Dept of Pediatrics, University of Washington, Seattle WA, USA; 13) Dept of Neurology, University of Washington, Seattle WA, USA; 14) Dept of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA.

In late 2011, we and others found that constitutional mutations in the epigenetic regulator *EZH2* (enhancer of zeste homolog 2) cause Weaver Syndrome (WS). WS is a rare congenital disorder characterized by generalized overgrowth, advanced bone age, intellectual disability, neuronal migration defects and susceptibility to various cancers. We are now investigating the link between *EZH2* mutations and the clinical features of WS. Normally, EZH2 acts as a histone methyltransferase in the polycomb-repressive complex 2 (PRC2), and silences transcription through methylation of histone H3 lysine 27 (H3K27). Somatic mutations of *EZH2* found in various cancers lead to enhanced trimethylation activity; thus, functional investigations can provide insights into the pathogenesis of WS.

Last year at ASHG (63rd annual meeting, 2013), I provided an update on our cohort, where 9 out of 35 individuals with Weaver-like phenotypes were found to have mutations in *EZH2*. Since then, we have identified a mutation in *EZH2* in one more patient in our cohort (currently 45 patients), as well as 2 mutations in *NSD1*, the Sotos syndrome gene.

At the time of my presentation, I also showed results of an *in vitro* functional assay which supported the hypothesis that WS is caused by loss-of-function mutations in *EZH2*. These findings suggested that EZH2 inhibitors being developed for cancer treatment may not be as effective in rare cancers associated with WS. As such, a more detailed investigation of our WS mutants was required. Our most recent results suggest that WS mutants may be classified into different groups based on variable degrees of loss-of-function, which could explain some of the variability in the observed phenotype and also suggests a need for different therapeutic approaches based on the mutation identified.

2711S

Cantu syndrome: Delineation of cardiovascular abnormalities in six affected individuals evaluated at a research clinic. *D. Grange[1], B. Kozel[1], G. Singh[1], M. Levin[1], P. Stein[2], C. Nichols[3].* 1) Dept Pediatrics, Washington University; 2) Dept Internal Medicine, Washington University; 3) Dept Cell Biology and Physiology, Center for the Investigation of Membrane Excitability Diseases, Washington University, St Louis, MO.

Cantu syndrome (CS) is associated with hypertrichosis, neonatal macrosomia, macrocephaly, craniofacial dysmorphic features, osteochondrodysplasia, edema and cardiovascular abnormalities including PDA, cardiomegaly, pericardial effusion and abnormal cerebral vasculature. CS is caused by heterozygous activating mutations in either ABCC9 or KCNJ8 which encode the regulatory SUR2 and Kir6.1 subunits, respectively, of the ATP-sensitive potassium channels (KATP). KATP channels formed from SUR2 and Kir6.1 subunits are prominent in cardiovascular tissues, and KATP activation in vascular smooth muscle results in decreased vascular contractility. Six individuals with CS were evaluated in the first Cantu Syndrome Research Clinic, included 4 females and 2 males ranging in age from 12-46 years of age. All patients had heterozygous mutations in ABCC9. Participants had cardiovascular testing including ECG, echocardiograms, 24 hour Holter monitoring with heart rate variability (HRV) assessment and pulse wave velocity testing via SphygmoCor. Blood pressures were below the 50th percentile for age for all participants, although none had orthostatic changes with positional change. They had full pulses with normal peripheral perfusion. Echocardiograms revealed cardiomegaly with enlarged ventricles. Left ventricular diastolic volumes were +3 SD from the mean. Cardiac mass was +2.5 SD from the expected values for age, gender and body surface area. Left ventricular systolic strain, a measure of cardiac contractility, was increased compared to age and sex matched normative values. Echocardiogram-derived cardiac output showed increased output when compared to controls. Pulse wave velocity was slow in the younger individuals, but average in the adults, suggesting a trend toward lax vasculature in childhood that becomes stiffer with age. Most participants showed peripheral edema. Although there was no clinical heart failure, overall, the findings were suggestive of a high output, hypercontractile state with low peripheral vascular resistance in the absence of hypertrohic cardiomyopathy. 12-lead ECGs showed a repolarization abnormality with inverted T-wave and flat ST segment in inferior leads in 4/6 patients. The Holter analysis revealed autonomic dysregulation manifested by increased sympathetic activity and decreased vagally modulated HRV in all participants. The oldest participant had extremely depressed HRV values. Cardiac function may deteriorate with age in patients with CS.

2712M

Exome sequencing reveals compound heterozygous mutations in *ATP8B1* in a *JAG1/NOTCH2* mutation-negative patient with clinically diagnosed Alagille syndrome. *C.M. Grochowski[1], R. Rajagopalan[1], A.M. Falsey[1], K.M. Loomes[1], I.D Krantz[1,2], M. Devoto[2,3], N.B. Spinner[1,2].* 1) The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) University La Sapienza, Rome, Italy.

Exome sequencing is helping to revolutionize genetic screening, as it permits unbiased testing, demonstrating erroneous clinical diagnoses. Alagille syndrome (ALGS) is an autosomal dominant disorder characterized by bile duct paucity and the presence of distinct facial features, cardiac findings, musculoskeletal malformations, and ocular abnormalities. Prior to genetic testing, ALGS was diagnosed clinically in individuals with bile duct paucity plus 3 of 4 secondary findings. Currently, 94% of ALGS patients with a clinically confirmed diagnosis have mutations detected in *JAGGED1 (JAG1)* and an additional 2% of patients have mutations in *NOTCH2*. We report a patient who presented with 4/5 features of ALGS including bile duct paucity, a cardiac murmur, posterior embryotoxon and facial features (high broad forehead, triangular face), leading to a clinical diagnosis of the disorder. Genetic testing revealed no mutations in *JAG1* or *NOTCH2*, warranting an expanded search for causal genes. Exome sequencing uncovered compound heterozygous mutations in the gene *ATP8B1*, previously reported to cause Progressive Familial Intrahepatic Cholestasis Type 1 (PFIC1), which has very limited overlapping clinical findings with ALGS. A splice-site mutation (c.3400+2;T>C) inherited maternally and a frameshift mutation (c.1889_1890insGTAAC p.His630fs) inherited paternally were detected and validated through Sanger sequencing. A similar splice site mutation has been previously reported in a patient with a clinical presentation similar to our patient, and a diagnosis of atypical PFIC1. We are investigating further clinical features within this patient to determine the consistency of a PFIC diagnosis versus an ALGS diagnosis. This work helps to delineate the phenotypic spectrum of Alagille syndrome and more precisely characterize clinical diagnostic criteria and avoid confounding diagnoses.

2713T

New dominant mutations in SF3B4 encoding an mRNA spliceosomal protein important in embryonic limb patterning underlie Rodriguez acrofacial dysostosis. *M.D. Irving[1,2], B. Dimitrov[2], D. Chitayat[3], J.I. Rodriguez[4], M.W. Wessels[5], M.A. Simpson[2].* 1) Department of Clinical Genetics, Guy's and St Thomas' NHS Foundation Trust, London, London, United Kingdom; 2) Division of Medical and Molecular Genetics, King's College, London, United Kingdom; 3) The Hospital for Sick Children, Toronto, Canada; 4) Department of Pathology, Hospital Universitario La Paz, Madrid, Spain; 5) Department of Clinical Genetics, Erasmus Medical Centre, Rotterdam, The Netherlands.

Acrofacial dysostosis syndrome of Rodriguez (RADS) (OMIM 201170) is a lethal condition comprising multiple congenital abnormalities. It is characterised by severe mandibular hypoplasia, upper limb phocomelia with oligodactyly, absent fibulae, cleft palate and microtia. Abnormal lung lobulation results in pulmonary hypoplasia and early neonatal demise. It was first reported by Rodriguez et al. who described three siblings with multiple congenital anomalies affecting the limbs, cardiovascular and neurological systems, and severe mandibular hypoplasia, resulting in lethal respiratory insufficiency. Thereafter five additional cases have been described in the literature expanding the clinical phenotype to include abnormal lung lobulation, absent fibulae, cleft palate and arhinencephaly. Given the occurrence in three siblings, RADS has been considered to be an autosomal recessive genetic condition. There are obvious similarities with RADS and other syndromes associated with severe mandibular hypoplasia, such as Nager syndrome (NS)(OMIM 154400) and Miller syndrome (OMIM 263750). Whilst there are clinical features to distinguish these conditions from RADS, such as eyelid colobomata and accessory nipples, there is clear phenotypic overlap, raising the possibility that RADS and NS are indeed allelic. Heterozygous loss-of-function mutations in SF3B4, encoding the mRNA spliceosomal protein Splicing factor 3B, subunit 4 that has a role in bone morphogenetic protein signalling in early embryogenetic limb patterning, have been shown to underlie NS. Recently a de novo mutation in SF3B4 exon 4 that is predicted to disrupt splicing was demonstrated in a patient with RAD. In this study, we obtained DNA samples for whole exome sequencing from four fetuses with RADS, including one originally described by Rodriguez et al., to investigate the underlying molecular basis and to explore the concept that there is both phenotypic and molecular overlap between this and other acrofacial dysostosis conditions. Results demonstrated two heterozygous frameshift mutations of SF3B4 in three of the four cases. In addition, inheritance studies failed to identify mutations in unaffected parents of one fetus, supporting new dominant inheritance and not recessive transmission, which will significantly influence genetic counseling. The clinical and molecular data will be presented, as well as a discussion of the role of SF3B4 mutations in causing this distinctive phenotype.

2714S

New case of a small AFF2 (FMR2) intragenic deletion associated to development delay causing a Fragile X E phenotype. *E. Pipiras[1], Ch. Beldjord[2], L. El Khattabi[3], B. Héron-Longe[4], B. Benzacken[1], A. Delahaye[1].* 1) AP-HP, Hopital Jean Verdier, Laboratoire d'Histologie-Embryologie-Cytogénétique, Hopital Jean Verdier, Bondy, France; 2) AP-HP, Hopital Cochin, Laboratoire de Biochimie et Génétique Moléculaire, Paris, France; 3) AP-HP, Hopital Cochin, Laboratoire de Cytogénétique, Paris, France; 4) AP-HP, Hopital Jean Verdier, Service de Pédiatrie, Bondy, France.

Fragile X E (FRAXE) disorder is an X-linked form of intellectual disability characterized by mild to moderate mental retardation associated with learning difficulties, speech delay, communication deficits, attention problems, hyperactivity and autistic behavior. The typical FRAXE phenotype is due to (CCG)n repeats expansions in the AFF2 gene which leads to silencing of the AFF2 gene. We describe clinical features and molecular characterization of an intragenic deletion of AFF2 gene in a 15 year-old boy born to non consanguineous parents. The patient was referred to genetics department for mild mental retardation, learning difficulties, speech delay and developmental delay. Usual analysis of (CCG)n repeat expansions in the AFF2 gene was normal. Chromosomal microarray analysis using whole genome HumanCyto-12 SNP (Illumina) array revealed a 60.3 kb intragenic deletion within AFF2 on Xq28 band. PCRq analysis confirmed this rearrangement and showed its maternal inheritance (arr[hg19] Xq28(148,004,585-148,064,903)x0 mat). The mother is apparently normal. X-inactivation study is ongoing. We are reporting a new case of small AFF2 (FMR2) intragenic deletion. We studied genotype-phenotype correlations for this patient and the previously reported patients with CCG expansion or intragenic deletion of AFF2.

2715M

Disruption of *HDAC8* gene due to partial duplication in a female with Cornelia de Lange syndrome diagnosed by SNP microarray. *S. Ramanathan, R.D. Clark.* Pediatric Genetics, Loma Linda University, San Bernardino, CA.

The clinical diagnosis of Cornelia de Lange syndrome (CdLS [MIM 122470]), a cohesinopathy, is evolving as its genetic heterogeneity is better understood. To date, mutations in 5 genes, 3 core cohesion subunits (*SMC1A* [MIM 300040], *SMC3* [MIM 606062], *RAD21* [MIM 606462]) and 2 cohesin-regulatory proteins (*NIPBL* [MIM 608667], *HDAC8* [MIM 300269]) have been identified in CdLS. The distinctive features in classic CdLS include synophrys, arched eyebrows, long prominent philtrum, small, upturned nose and upper limb reduction defects. Significant growth and cognitive impairment are typical. Milder, subtle phenotypes have also been reported, with the variability likely due to inter- and intragenic mutation spectrum (Mannini L et al., 2013).

We report a 4-month old female with CdLS due to a partial duplication of the *HDAC8* gene, who presented as a newborn with IUGR, tetralogy of Fallot, bifid uvula and asymmetric crying face due to right facial palsy. Her dysmorphic features did not raise suspicion for CdLS until SNP microarray analysis detected a *de novo* 201 kb duplication on chromosome X at Xq13.1 that disrupted the *HDAC8* gene. An X-inactivation ratio of 100:0 in peripheral blood was highly skewed. At 2 months, she had a wide open anterior fontanel, depressed supraorbital ridges with hooded appearance of the eyes, mild synophrys and long, prominent philtrum with thin upper lip. The 5th fingers were short with a single crease and clinodactyly. She had failure to thrive, dysphagia and G-tube dependency. She did not have hirsutism or arched eyebrows.

Mutations in *HDAC8*, most of which are missense, have been recently identified in CdLS. The facial features can be atypical, with some overlap with that of classic CdLS. Distinguishing features are delayed anterior fontanelle closure, hooding of the eyelids and a broader nose with no gross limb abnormalities (Deardorff MA et al., 2012, Kaiser FJ et al., 2014). Most patients are heterozygous females with marked X-inactivation skewing in peripheral blood, favoring the normal allele. To our knowledge, this is the first patient with CdLS caused by a partial duplication of the *HDAC8* gene. The atypical phenotype associated with *HDAC8*, which was not recognized on clinical evaluation, illustrates the emerging phenotypic variability in CdLS.

2716T

Exome sequencing of individuals with non-deletion Smith-Magenis syndrome reveals potentially causative genic variants. *J.J. White[1], C.R. Beck[1], T. Gambin[1], T. Harel[1], J.R. Lupski[1,2,3].* 1) Molecular & Human Genetics, Baylor College of Medicine, houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 3) Texas Children's Hospital, Houston, TX, USA.

Smith-Magenis syndrome (SMS; MIM #182290) is a developmental disability/multiple congenital anomaly disorder often resulting from haploinsufficiency of RAI1 and manifesting with a wide range of phenotypes. The clinical manifestations of SMS varies widely with roughly 90% of cases manifesting intellectual disability, speech delay and hypotonia, and 70% developing hearing loss, sleep disturbances and stereotypic behaviors including self-hugging and self-injurious behaviors. We investigated a cohort of 154 patients with SMS and subjected DNA samples to a targeted, high-density array comparative genomic hybridization (aCGH) specific for chromosome 17p. Samples lacking a copy number variant (CNV) in the SMS critical region were then sequenced for point mutations in RAI1, of which seven (5%) contained a mutation within the RAI1 coding sequence. On the remaining 16 individuals without a molecular diagnosis, whole exome sequencing (WES) was implemented using extensive filtering of variants common in the Atherosclerosis Risk in Communities (ARIC), Exome Sequencing Project (ESP), and our in-house Exome sequencing databases. We found predicted damaging variants and putative de novo variants in other genes known to cause developmental delay and/or intellectual disability in six patients. Additionally, subsequent genome-wide high exon-targeted aCGH identified a potentially causative 15q11.2 deletion in a singleton case. The initial clinical diagnosis of SMS with each of these patients underscores the challenge of clinical diagnosis for complex disorders with variable phenotypes, and highlights the benefit of WES in patients with normal CMA. Moreover, it suggests that the genes underlying phenotype mimics of SMS may share common networks with RAI1, and presents this as a further venue of research.

2717S

Cervical myelopathy in a patient with metatropic dysplasia caused by a TRPV4 mutation. *E. Zapata-Aldana[1], L. Arnaud-López[1], E.L. Mellín-Sánchez[1], C. Peña-Padilla[1], J. Rivera-Vargas[1], F.J. Martínez-Macías[1], L. Bobadilla-Morales[1,3], A. Corona-Rivera[1,3], A. Superti-Furga[2], J.R. Corona-Rivera[1,3].* 1) Servicio de Genética, División de Pediatría, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca", Guadalajara, Jalisco, México; 2) University of Lausanne, Centre Hospitalier Universitare Vaudois, Lausanne, Switzerland; 3) Instituto de Genética Humana "Dr. Enrique Corona Rivera", CUCS, Universidad de Guadalajara, Jalisco, México.

Introduction. The name of metatropic dysplasia (MD[MIM #156530]) is derived from the Greek "metatropos," (changing patterns), because initially exhibit a long thorax with short limbs and later a short-trunked dwarfism. MD is an autosomal dominant disorder caused by mutations in the *TRPV4* gene (MIM #156530), which encodes for a calcium channel that participates in the ossification of the chondrocytes, however, *TRPV4* produces also other skeletal dysplasias and neurodegenerative disorders. Here, we present the clinical, radiological and molecular findings in a female patient with a classic form of MD, complicated by a cervical myelopathy. **Clinical report.** The *proposita* was born from healthy, and remotely consanguineous parents. She was born after an uncomplicated term pregnancy and a normal vaginal delivery. At birth, she weighed 3.5 Kg (50th percentile). Physical examination at 2.5 years of age showed a weight of 11 Kg (-2.9 SD), height of 81 cm (-3.4 SD), span of 87 cm. She had short trunk, severe kyphoscoliosis, hypermobile fingers, enlarged joints at knees and elbows, and hallux valgus. Radiographies demonstrated thoracolumbar scoliosis, platispondyly, wide intervertebral space, lumbar hyperlordosis, halberd-shaped pelvis, dumbbell-shaped metaphyses in femurs, and wide-irregular metaphyses in fingers and toes. The spine MRI showed C1-C2 subluxation, cervical instability, and spinal cord compression. At the age of 5 years the *proposita* undergo thru a surgical procedure with cervical arthrodesis, unfortunately, she progressed to quadriparesis six months after the surgery. The molecular study of the TRPV4 confirmed the mutation P799L at the heterozygous state. **Conclusions.** The clinical and radiographic features, in addition to the molecular study in our patient confirmed the diagnosis of MD. All of the cases with a P799L heterozygous mutation in the *TRPV4* gene have been reported as MD, in ours with a suitable genotype-phenotype correlation, even considering the cervical myelopathy. Early detection with a careful neurological assessment, spine MR imaging, and appropriate surgical treatment should be carried out in all patients in whom MD is suspected, to prevent the observed neurological sequelae and improve its life quality.

2718M

CNS Involvement in OFD1 syndrome: a Clinical, Molecular, and Neuroimaging study. *B. Franco[1,2], M. Macca[1], F. Imperati[2], A. D'Amico[3], P. Parent[4], L. Pasquier[5], V. Layet[6], S. Lyonnet[7], V. Stamboul-Darmency[8], C. Thauvin Robinet[9,10], E. del Giudice[1], Oral-Facial-Digital Type I (OFD1) Collaborative Group.* 1) TIGEM, Fondazione Telethon, Naples, Italy; 2) Department of Translational medical Sciences, Federico II University of Naples, Italy; 3) Department of Radiological Sciences, Neuroradiology Unit, Federico II University of Naples, Naples, Italy; 4) Service de Pédiatrie et de Génétique, CHU Brest, France; 5) Service de Génétique Médicale - Centre de Référence Maladies rares Labélisé « Anomalies de Développement et Syndromes Malformatifs » de l'Ouest, CHU Rennes, France; 6) Consultation de Génétique, Groupe Hospitalier du Havre, CH Le Havre, France; 7) Département de Génétique, Hôpital Necker-Enfants Malades, APHP, Paris, France; 8) Service de Pédiatrie 1, Hôpital d'Enfants, CHU Dijon, France; 9) Centre de Génétique et Centre de Référence Maladies rares Labélisé « Anomalies de Développement et Syndromes Malformatifs » de l'Est, Hôpital d'Enfants, CHU Dijon, France; 10) EA 4271 GAD « Génétique des Anomalies du développement », IFR 100 - Santé STIC, Université de Bourgogne, Dijon, France.

Background: Oral-facial-digital type 1 syndrome (OFD1; OMIM 311200) belongs to the expanding group of disorders ascribed to ciliary dysfunction. With the aim of contributing to the understanding of the role of primary cilia in the central nervous system (CNS), we performed a thorough characterization of CNS involvement observed in this disorder. Methods: A cohort of 117 molecularly diagnosed OFD type I patients was screened for the presence of neurological symptoms and/or cognitive/behavioral abnormalities on the basis of the available information supplied by the collaborating clinicians. Seventy-one cases showing CNS involvement were further investigated through neuroimaging studies and neuropsychological testing. Results: Seventeen patients were molecularly diagnosed in the course of this study and five of these represent new mutations never reported before. Among patients displaying neurological symptoms and/or cognitive/behavioral abnormalities, we identified brain structural anomalies in 88.7%, cognitive impairment in 68%, and associated neurological disorders and signs in 53% of cases. The most frequently observed brain structural anomalies included agenesis of the corpus callosum and neuronal migration/organisation disorders as well as intracerebral cysts, porencephaly and cerebellar malformations. Conclusions: Our results support recent published findings indicating that CNS involvement in this condition is found in more than 60% of cases. Our findings correlate well with the kind of brain developmental anomalies described in other ciliopathies. Interestingly, we also described specific neuropsychological aspects such as reduced ability in processing verbal information, slow thought process, difficulties in attention and concentration, and notably, long-term memory deficits which may indicate a specific role of OFD1 and/or primary cilia in higher brain functions.

2719T

Neurofibromatosis type 1: Familial case and retroperitoneal neurofibroma. *M.A. Aceves-Aceves[1], I.M. Salazar-Dávalos[1], S.A. Alonso-Barragán[2], R.E. Jiménez-Arredondo[2], D. García-Cruz[1], N.O. Dávalos[1], A.G. Puebla-Mora[3], M. Salazar-Páramo[4], I.P. Dávalos[1,2].* 1) Instituto de Genética, DGH, Universidad de Guadalajara, Guadalajara, Jalisco, México; 2) Doctorado Genética Humana, CUCS, Universidad de Guadalajara, CIBO-IMSS, Guadalajara, México; 3) Servicio de Anatomía Patológica, UMAE, HP, CMNO, IMSS, Guadalajara, México; 4) Depto. Fisiología, CUCS, Universidad de Guadalajara, Div. Investigación, UMAE, HE, CMNO, IMSS, Guadalajara, México.

INTRODUCTION: Neurofibromatosis type I (NF1), OMIM #162200, is an autosomal dominant disorder characterized by café-au-lait spots, Lisch nodules (iris hamartoma), fibromatous tumors of the skin and development of benign and malignant tumors. It is caused by mutation in the neurofibromin gene on chromosome 17q11.2. The incidence of NF1 is 1 in 3000 individuals. OBJECTIVE: To present a familial case with NF1 and retroperitoneal neurofibroma in the propositus. CASE REPORT: Family data: Patient 1 (propositus): 9 year-old, product of the 4th pregnancy, obtained by vaginal delivery; from non-consanguineous parents, mother (40 years), father (50 years, patient 3), sister (14 years, patient 2) and a healthy sister (13 years). Physical examination: weight 22kg (pc10), height 127cm (pc25), head circumference 53cm (pc90), café-au-lait spots in thorax (>30), axillary freckles, post extirpation scar of plexiform neurofibroma on right costal region, congenital pseudoarthrosis of the left tibia, treated with intramedullary nail placement. Abdominal ultrasonography studies reported retroperitoneal mass, confirmed by CT as a solid lobulated mass, from the upper border of the pancreas extending to retroperitoneum, embracing aorta and left renal vessels, displacing the left kidney, lymph nodes in large vessels; the histopathological study reported intraneural neurofibroma. Patient 2: (sister) 14 year-old female, product of 2nd pregnancy, obtained at 40 weeks. Physical examination: weight 51 Kg (pc25), height 152cm (pc10), head circumference 59cm (pc>97), café-au-lait spots in thorax and abdomen, axillary freckles, surgically removed subcutaneous nodule in left arm, histopathological study reported plexiform neurofibroma. Pelvic ultrasonography revealed right polycystic ovary. Patient 3: (father) 50 year-old male with a history of surgical resection of nodule on right arm 20 years ago. Physical examination: weight 65kg, height 162 cm, café-au-lait spots in thorax and abdomen and axillary freckles. CONCLUSIONS: According to the NIH Consensus Development Conference established criteria we present a NF1 family case. The propositus presented a retroperitoneal tumor and the histopathological analysis demonstrated an intraneural neurofibroma, an uncommon finding in NF1. Early diagnosis of NF1, close clinical evaluations, periodical imaging studies and a multidisciplinary approach allows an opportune detection and treatment of complications improving patient's prognosis.

2720S

Clinical-pathological features in a female infant with Pfeiffer syndrome type 3 negative for FGFR2 mutation. *AK. SANDOVAL TALAMANTES[1], C. PEÑA-PADILLA[1], E.L. MELLIN-SANCHEZ[1], J. RIVERA-VARGAS[1], E. ZAPATA-ALDANA[1], L. VIRAMONTES-AGUILAR[1], J. TAVARES-MACIAS[1], L. BOBADILLA-MORALES[1], M. CUNNINGHAM[2], S. PARK[2], J.R. CORONA-RIVERA[1].* 1) Centro de Registro e Investigación sobre Anomalías Congénitas (CRIAC), Servicio de Genética y Unidad de Citogenética, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca·, Guadalajara, Jalisco, Mexico; 2) Craniofacial Center, Seattle Children's Hospital, Washington, USA.

Introduction. Pfeiffer syndrome (PS) is a genetic disorder characterized by craniosynostosis, broad thumbs and big toes, and variable syndactyly (OMIM # 101600). Cohen classified three types of PS: Type 1, have a better prognosis, usually familiar and could have normal intelligence. Type 2, is lethal with cloverleaf skull, severe ocular proptosis, elbow ankylosis and visceral anomalies. Type 3, share the same features that type 2, without cloverleaf skull, and can have abdominal wall defects. All of the cases with PS types 2 and 3 are caused by different mutations in *FGFR2* gene, 80% of them in exons 8 and 10. Here, we describe the clinical-pathological features found in a female infant with PS type 3. **Clinical report.** The proposita was the product of the first pregnancy of a healthy 21-year-old mother and a 25-year-old father, who were non consanguineous. Prenatal ultrasonography reported omphalocele and craniosynostosis. The mother has positive serology for parvovirus B19. Vaginal delivery was at the 40th week of gestation. Apgar scores were 5, 3, and 1 at 1, 5 and 10 min, respectively. Birth weight was 3200 g, length 50 cm, and occipitofrontal circumference 31.3 cm (<3rd percentile). Physical examination showed turribrachycephaly, ocular proptosis, midfacial hypoplasia, down-slanting palpebral fissures, low-set ears, hemangioma on dorsum of the nose, prominent abdomen with prune-belly appearance, supraumbilical midline omphalocele; limitation of elbow extension, and square shaped first toes. He died at 30 min after birth. Autopsy was performed adding: brain microgyria, laryngotracheal stenosis, patent ductus arteriosus, omphalocele, hepatomegaly, partial intestinal malrotation with a subhepatic cecal appendix. Karyotype with G-bands was normal 46,XY, at a 550 band-level resolution. Sequencing of the exons 3, 7, 8, 10, 11, 14, 16, 17 of the *FGFR2* gene, was negative for mutation. **Conclusions**. Turribrachycephaly, broad and square shaped big toes, limitation of elbow extension, abdominal wall defects, visceral anomalies, and early death supports the clinical diagnosis of PS type 3. Four previous patients with PS and prune belly have been reported -two with cloverleaf skull, however, molecular study was not performed in none of them. The negative result obtained for the search of common mutations in *FGFR2* gene, probably implies a new or atypical mutation for the severe PS phenotype in our patient and consequently, further investigations are necessary.

2721M

Identification of structural alterations in the *CX50* gene in patients with congenital cataracts. *A.L. Araujo[1], E.S. Figueiredo[2], P.R.S. Cruz[1], B.B. Souza[1], C.E.L. Arieta[2], M.B. Melo[1].* 1) Center of Molecular Biology and Genetic Engineering (CBMEG), University of Campinas, Campinas, São Paulo, Brazil; 2) Department of Ophthalmology, Faculty of Medical Sciences (FCM), University of Campinas, Campinas, São Paulo, Brazil.

Congenital cataract is the leading cause of reversible blindness in childhood, with a prevalence of one to five cases per 10,000 live births. The hereditary form represents about 50% of these cases. The most common mode of inheritance is autosomal dominant with high penetrance, although autosomal recessive and X-linked inheritance are also observed. Genetic alterations responsible for non-syndromic congenital cataracts lead to changes in lens proteins, such as crystallins, membrane transport proteins (connexins, aquaporin) and cytoskeletal proteins (BFSP2). Connexins are structural proteins that compose an extensive network of intercellular channels of low resistance (gap junctions) present between the lens fibers. These proteins allow ions, metabolites and second messengers flow between cells, maintaining the balance of intercellular communication and ensuring the transparency of the lens. Connexins are encoded by *CX46* and *CX50* genes (among others), both previously linked to congenital cataract. In order to better understand the molecular basis of the disease, structural alterations in the *CX50* gene were evaluated in patients with nuclear and lamellar bilateral congenital cataract with hereditary etiology by means of PCR and direct sequencing. Patients' recruitment and blood collection were approved by the Ethics Committee of FCM-UNICAMP (Campinas-SP, Brazil). Twenty seven patients from ten families were analyzed to date, sixteen affected and eleven unaffected. We observed eight point mutations, among which two were not yet described. There were alterations common to both affected and unaffected individuals (two), unique for affected individuals (five, including the novel ones) and unique for unaffected individuals (one) as well. Five alterations are located in the coding region, being three missense variations, including both novel ones, and two synonymous. All three missense variations were predicted to be probably damaging by SIFT, PolyPhen-2 and MutPred softwares. Noteworthy, the missense variation already described was observed exclusively in both affected members of a family of three individuals, being absent in the unaffected member. This might suggest a possible segregation with the disease, although the family size restricts the analysis power. The remaining members of the families are being analyzed to allow further conclusions about the alterations. Financial Support: Capes and FAPESP.

2722T

Female with a Complex Phenotype Associated with Variants in Two Neurodegenative Genes Detected by Whole Exome Sequencing: Diagnostic and Counseling dilemma. *M. Khalifa[1], L. Naffaa[2].* 1) Medical Genetics and Genomics, Akron Childrn Hospital, Akron, Ohio., USA; 2) Lena Naffaa. Department Radiology, Akron Children's Hospital, Akron, Ohio, USA.

We report on a female child with severe intellectual disability, aphasia, short stature, ataxia, and structural brain abnormalities. A brain MRI obtained in infancy showed hypomyelination involving the central periventricular white matter and thinning of the corpus callosum. Initial reporting of Whole Exome Sequencing (WES) identified three POLR3A 3 missense heterozygous variants: two, a maternal variant (c.1724A>T) and a paternal variant (1745G>A) on exon13, and a novel maternally inherited variant (c.346A>G) on exon 4. These variants are likely damaging. The patient's clinical features, early MRI findings, and WES results supported the diagnosis of hypomyelinating leukodystrophy type 7, and the family was counseled accordingly. An updated WES report revealed a novel WDR45 deleterious frameshift mutation in Exon 9 (c.587-588del). This, with later brain MRIs that showed progressive iron accumulation in the globus pallidus and substania nigra, raised the possibility of Neurodegeneration with Brain Iron Accumulation (NBIA). Both WDR45 and POLR3A are newly recognized genes; each is associated with a distinct neurodegenerative disease. WDR45 is an X-linked gene associated with a dominant form of NBIA, manifested by progressive neuropsychiatric abnormalities, dystonia, cognitive decline, spastic paraplegia, and iron deposition in the basal ganglia. POLR3A, on the other hand, is an autosomal gene causing a recessive form of a hypomyelination with leukodystrophy disease, also known as 4H syndrome. This syndrome is characterized by variable manifestations including congenital Hypomyelination with thinning of the corpus callosum, Hypodontia and Hypogonadotropic Hypogonadism. This patient's complex clinical presentation and mixed brain neuroimaging findings might be attributed to the confounding effects of the expression of these two genes. Alternative, they may represent an expansion of the phenotype of either condition. In any event, these made the diagnosis and genetic counseling of the family challenging.

2723S

Whole exome sequencing allows the identification of a novel large deletion in *PRPF31* in a family with autosomal dominant retinitis pigmentosa. *B. Almoguera[1], P. Fernandez-San Jose[2,3], Y. Liu[1], J. Liang[4], R. Golhar[1], M. March[1], R. Pellegrino[1], M. Corton[2,3], F. Blanco-Kelly[2,3], MI. Lopez-Molina[3,5], B. Garcia-Sandoval[3,5], Y. Guo[1], L. Tian[1], X. Liu[4], L. Guan[4], B. Keating[1], X. Xu[4], C. Ayuso[2,3], H. Hakonarson[1].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Genetics and Genomics, IIS-Fundacion Jimenez Diaz, 28040, Madrid, Spain; 3) Center for Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, Spain; 4) BGI-Shenzhen, Shenzhen 518083, China; 5) Department of Ophthalmology, Fundacion Jimenez Diaz, 28040, Madrid, Spain.

With all types of Mendelian inheritance patterns described and an extreme phenotypic, clinical, genetic and allelic heterogeneity, diagnosis of patients with retinitis pigmentosa (RP) is a complex task. Next generation sequencing (NGS) is increasingly being applied to the diagnosis of highly heterogeneous monogenic diseases, such as RP, as a rapid and affordable alternative to the former methods. However, one of the main limitations of NGS is the detection of copy number variants (CNV), which can account for a significant number of RP cases. In this study, we identified a large deletion in *PRPF31* using whole exome sequencing (WES) data in a family with autosomal dominant RP. Eight members of the family were enrolled in the study and two affected and two unaffected members were subjected to WES using Illumina Hiseq 2000 instruments. The Burrows-Wheeler alignment tool, Genome Analysis Tool Kit and ANNOVAR were used for mapping, genomic variant detection and variant functional annotation, respectively. Single nucleotide variant (SNV) analysis was performed searching among all potentially functional variants and for CNV analysis the standard Exome Hidden Markov Model (XHMM) was used. CNV validation was performed using TaqMan predesigned probes (Life Technologies) and the Universal Probe Library (Roche). SNV analysis did not yield any candidate variant. A heterozygous deletion in the genomic region chr19: 54600186-54628017, spanning the entire *PRPF31* gene, was identified by XHMM in the two affected members of the family. Both CNV assays used confirmed that the entire coding sequence of *PRPF31* was hemizygously deleted in the two affecteds and not in the unaffecteds. With this study we demonstrate the feasibility of detecting CNVs in RP genes using WES, thus expanding the potential of this tool in the diagnosis of this disease.

2724M

Whole gene duplication and partial duplication and triplication of *OPHN1*. *J.G. Pappas[1], E. Ward[1], P.R. Papenhausen[2].* 1) Dept Pediatrics, Clin Genetic Scvs, New York Univ, Sch Med, New York, NY; 2) Cytogenetics, Laboratory Corporation of America, Research Triangle Park, NC.

Inactivating mutations and intagenic deletions in *OPHN1* cause X-linked mental retardation with cerebellar hypoplasia, ataxia, seizures and distinctive facial appearance (Al-Owain M et al 2011). Duplication X(q12q13.3) containing *OPHN1* along with other OMIM annotated genes has been associated with developmental delay, microcephaly, short stature and autism in males (Bedeschi, MF et al 2008, Kaya N et al 2012, Prontera P et al 2012). We present a 14 year old boy (case 1) with dup(X)(q12q13) that includes *OPHN1* and *AR* and a 3 year old boy (case 2) with duplication and triplication in Xq12q13.1 involving *OPHN1* only. Our cases are unrelated and presented with mild to moderate mental retardation. Case 1 was born full term male, small for gestational age with microcephaly. He continued to grow below the 3rd percentile. Routine karyotype was normal. His cognitive development was delayed throughout childhood. Our evaluation at age 14 revealed mild to moderate mental retardation, short stature, microcephaly, elongated face, hypotelorism, narrow palpebral fissures and long nose with bulbous tip. Case 2 was born full-term without complications and normal anthropometrics. His growth followed the 50th centile for height, 95th for weight and 25th for head circumference. Developmental delays and strabismus were noted in infancy. He started walking at 17 months. At 28 months, he was speaking a few words; his face was triangular and elongated with prominent forehead, hypotelorism and bulbous tip of the nose. Blood from case 1 was analyzed by SNP microarray (Affymetrix 6.0) which showed a 983 KB interstitial duplication of X(q12q13): arrXq12q13.1(66,720,159-67,702,721)x2. This de novo copy number change was confirmed by the RP11-963N10 region specific BAC. Case 2 was tested with a higher resolution microarray (Affymetrix Cytoscan) which revealed a maternally inherited duplication Xq12(67,295,132-67,404,398) and triplication Xq12q13.1(67,501,664-67,908,008). The mother was phenotypically normal. The duplication in case 1 shares only *OPHN1* with the previously reported duplications and it contributes to the association of the phenotype to *OPHN1* only. Case 2 is unique because it lacks ataxia and seizures which are common in cases with disruption of *OPHN1*. Our cases have common facial features with cases reported with duplications, intragenic deletions and inactivating mutations in *OPHN1* and this observation may help elucidating the role of this gene during development.

2725T

A Unique Family with Progressive Pseudorheumatoid Arthropathy of Childhood. *A. Neogi[1], V. Kimonis[1], J. Soni[2], E. Chao[1].* 1) Division of Genetics and Genomic Medicine, Department of Pediatrics, University of California, Irvine, CA; 2) Gandhi Lincoln Hospital, Deesa, Banas Kanta, Gujarat, India.

We report a family in India with five affected female siblings with an unusual manifestation of *progressive pseudorheumatoid arthropathy of childhood.* Progressive pseudorheumatoid arthropathy of childhood is an autosomal recessive condition that affects the bones and cartilage of joints with a childhood onset. Major symptoms include stiffness of joints, bony swellings of toes and fingers, weakened muscles, fatigue and bowed legs. The disorder is progressive and occurs due to mutations in the WISP3 gene encoding for a signaling protein that is essential for normal postnatal skeletal growth and cartilage homeostasis. In the affected girls, symptoms typically started around 3-5 years of age and lead to crouching followed by a squatting gait. They developed severe joint contractures of the elbows, knees, ankles and hips, with motor weakness and difficulty sitting and walking. The girls developed joint swellings. They do not have pain but have severe restriction of their mobility, which has been documented in a video clip. Skeletal X-Rays revealed enlargement and cystic malformations of the epiphyses and metaphyses and joint erosions. We performed **exome sequencing** through Ambry Genetics for the affected individuals. Two novel mutations in the WISP3 gene were reported after assessment. These c.172A>T, p.K58X and c737_738del, p. L246LfxX32 mutations are considered most likely contributory to the progressive arthropathy. The girls have an unaffected brother and two unaffected sisters and consanguinity was denied in the family. **Confirmatory testing** is in progress in the parents in order to confirm phase (cis/trans) of the variants and provide further support for the molecular etiology of the disorder in this family. Future research of this pathway could lead to promising therapies to modify this progressive disease.

2726S

Mutation in the EZH2 gene in a Brazilian family - Complex clinical findings. *D.L. Polla[1], N.M. Kokitsu-Nakata[2], M.C.B. Silva[1], I.C.C. Cardoso[1], A. Richieri-Costa[2], R. Pogue[1].* 1) Catholic University of Brasilia, Brasilia, Distrito Federal, Brazil; 2) Hospital for Rehabilitation of Craniofacial Anomalies - University of Sao Paulo, Bauru, Sao Paulo, Brazil.

Weaver syndrome (WS) is a multi-systemic pre- and post- natal overgrowth syndrome associated with variable intellectual disability. Most of the reported cases have been described as sporadic, however in a few instances vertical transmission has been documented indicating autosomal dominant inheritance. Here we describe a family with 3 affected patients (mother, son and daughter), and an unusual clinical presentation within the family. There was a wide variation in clinical manifestations ranging from severe vestibule-cochlear anomalies in the proposita to behavioral disturbance, delayed language acquisition and learning difficulties in her brother. The original diagnostic hypothesis was Sotos syndrome. However, sequencing was carried out using an NGS (Illumina) skeletal dysplasia-specific panel which included the Sotos syndrome genes (NDS1 and NFIX) and the WS gene, EZH2. The affected individuals shared a c.149A>G; p.50L>S variant in EZH2, confirming a diagnosis of WS. Panel sequencing was therefore useful and important in determining the definitive diagnosis. Unknown modifier loci may be responsible for the clinical heterogeneity within the family.

2727M

A Novel Mutation, p. (Lys1474*), in a Female adds Seizures and Ptosis to Clinical Findings in MED13L Haploinsufficiency Syndrome. *M.M. Ali[1], N. Smaoui[2], A.M. Slavotinek[1].* 1) Genetics, UCSF, San Francisco, CA; 2) GeneDx, Gaithersburg, Maryland, USA.

We describe a 12 year old female with a history of seizures and epileptiform activity on EEG, choreiform movements, moderate developmental delays, macrocephaly (OFC >97th centile), central obesity and dysmorphic findings, including a prominent and square forehead, left ptosis, epicanthic folds, a right ear pit, prominent lower jaw with underbite, thin and tapered fingers, cubitus valgus and an increased lumbar lordosis. An MRI showed T2 hyperintense foci in the periventricular and subcortical white matter with low white matter volume and a dysmorphic craniocervical junction. Whole exome sequencing demonstrated heterozygosity for a de novo, nonsense mutation in the Mediator Complex Subunit 13-like (MED13L) gene, p.(Lys1474*), that has not been reported previously and is predicted to cause loss of normal protein function. The MED13L protein is a subunit of the Mediator complex, a large complex of proteins that functions as a transcriptional coactivator for most of the RNA polymerase II-transcribed genes. MED13L is highly expressed in early development of the heart and brain. Heterozygous, missense mutations in MED13L were first described in patients with transposition of the great arteries, dextro-looped (dTGA) and one patient with a translocation disrupting MED13L between exons 1 and 2 had intellectual disability (ID) and dTGA. Subsequent reports have described mutations and translocations hypothesized to disrupt MED13L expression in patient with phenotypes including non-syndromic ID, hypotonia, conotruncal and septal heart defects, facial anomalies and a 'Noonan-like' phenotype with hypotonia and facial anomalies. MED13L dosage is important and mutations have been non-penetrant, heterozygous or homozygous; a de novo triplication of chromosome 12q24.2 including MED13L and MAP1LC3B2 has also been reported to result in a milder phenotype. Our case is a further example of MED13L haploinsufficiency syndrome and adds seizures, choreiform movements and ptosis as possible clinical findings in this condition.

2728T

Case report: Acromesomelic dysplasia with primary congenital glaucoma, and a distinct pattern of brachydactyly in a Brazilian patient. *W.A.R. Baratela[1], G.L. Yamamoto[1,2], C.A. Kim[1], D.R. Bertola[1,2].* 1) Unidade de Genética, Instituto da Criança - Hospital das Clínicas - Universidade de Sao Paulo, Sao Paulo, Sao Paulo, Brazil; 2) Centro de Estudos do Genoma Humano - Departamento de Genética e Biologia Evolutiva - Universidade de Sao Paulo, Sao Paulo, Sao Paulo, Brazil.

The (acro)mesomelic skeletal dysplasias, classified as n # 16 and 17 under the current skeletal dysplasias nosology, comprise a heterogeneous group of disorders, ranging from mildly affected Leri-Weill, Robinow dominant, Maroteaux type to more severe phenotypes such as Langer, Kantaputra, Nievergelt, and Savarirayan types. We report on a 14 year-old girl with disproportionate acromesomelic short stature and bilateral congenital glaucoma. Product of an uneventful pregnancy, she was born by cesarian section at term, small for gestational age, BW: 2320g, BL: 38,5cm (below 5th%ile), OFC: 34,5cm, Apgar score 4/9. Parents are nonconsanguineous, with a negative family history. Primary congenital glaucoma surgery was performed at 1mo. She evolved with normal developmental milestones and cognitive function. Echocardiogram, abdominal US, brain CT scan, and chromosome analysis were all normal. At 12 y 3mo she presented with H: 98.5cm (below 5th%ile), and normal OFC. Mesomelic shortening of the upper and, especially, lower limbs, brachydactyly with hypoplastic nails and an increased distance between the 1st and the 2nd toes were observed. Radiological findings included a mild deformity of the forearm, hypoplastic middle and distal phalanx and carpal bone fusions, and a more prominent deformity of the lower limbs, with short and broad tibiae-fibulae, brachymetatarsalia with fusion of the 2nd metatarsal and the intermediate cuneiform bone. Small iliac wings with coxa vara were also found. The skeletal findings presented by our patient seem distinct from the previous described acromesomelic disorders, although the mesomelic involvement resembles the one described in Savarirayan type. Besides, the eye abnormality is rarely reported in skeletal dysplasias. The etiological genetic background for some of the acromesomelic dysplasias is still unknown. NGS of the exome for the case here described is underway, with preliminary results. This could soon unmask the causative gene, and whether this is a unique entity or part of the spectrum of a known genetic defect.

2729S

An autosomal recessive microcephaly syndrome with primordial growth failure and pigmentation changes is caused by mutations in the gene *ANKLE2*. *R. Clark[1], C. Curry[2], W. Dobyns[3], J. Lupski[4], H. Bellen[4], M. Wangler[4].* 1) Division of Medical Genetics/Peds, Loma Linda University Medical Center, Loma Linda, CA; 2) UCSF/Genetic Medicine Central California, Fresno, CA; 3) Center for Integrative Brain Research, University of Washington, Seattle, WA; 4) Dept. of Human and Molecular Genetics, Baylor College of Medicine, Houston, TX.

We report a non-consanguineous Hispanic family with two children affected with severe congenital microcephaly, primordial growth failure, pigmentation changes and novel mutations in *ANKLE2* identified through whole exome sequencing (WES). The proband, now aged 8, presented with a history of mildly low birth weight (2.67 kg) and severe microcephaly. His growth parameters have been severely reduced with weight at -4 SD, height at -6 SD and OFC at -9 SD. His development is at the <6 mos level. He has had unexplained anemia, glaucoma, seizures and feeding issues requiring a G tube. His examination is notable for hyperpigmented and hypopigmented macules over his entire body, a sloping forehead, ptosis, micropenis and generalized spasticity. Brain MRI revealed a simplified gyral pattern, absence of the corpus callosum, thick cortex and small frontal horns of the lateral ventricles with mildly enlarged posterior horns. Normal evaluations included chromosomes in blood and skin, a microarray, and studies for the Njimegen breakage syndrome and Fanconi Anemia. His affected sister had a similar phenotype and died within hours of birth of progressive cardiac failure without a pathological etiology. Her OFC at 37 wks GA was 26.5 cm and she had diffuse hyperpigmented and hypopigmented skin macules. A severe autosomal recessive microcephaly was suspected in this family and given the skin findings and severity of the brain abnormality, this appeared to be a unique phenotype.

DNA from both siblings and parents was sent for WES, and no mutations were found in previously published microcephaly loci. Among the candidates that met Mendelian expectation in this family, four genes were conserved and expressed in the CNS. One of these genes had been flagged as a potential novel Mendelian disease gene from a forward mutagenesis screen of the Drosophila X-chromosome which identified 165 fly genes. A subsequent study of variants in the human homologs from 1,929 exomes performed at Baylor led to prioritization of *ANKLE2*. See abstracts by Wangler et al. and Charng et al. Compound heterozygous variants in *ANKLE2* are the best candidate for disease-causing mutations in this family because of the unique combination of the human genetic evidence and the consistent neurologic phenotype in mutant flies with small brain size, and loss of neurons and support cells of the fly sensory organs.

2730M

A new craniofacial syndrome caused by localized mutations in *TWIST1*. *A.L. Fenwick[1], H. Cox[2], E. Sweeney[3], T. Lester[4], H. Lord[4], A.O.M. Wilkie[1].* 1) WIMM, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom; 2) Birmingham Women's NHS Foundation Trust, Mindelsohn Way, Edgbaston, Birmingham, UK; 3) Royal Liverpool Children's Hospital, Alder Hey, Liverpool, UK; 4) Genetics Laboratories, Oxford University Hospitals NHS Trust, Churchill Hospital, Oxford, UK.

Background: The association of *TWIST1* haploinsufficiency with Saethre-Chotzen syndrome (SCS [MIM 101400]; coronal synostosis with ptosis, small ears, and minor limb anomalies) is well known. Here we present evidence that mutations at a specific residue of *TWIST1* lead to a radically different craniofacial phenotype, possibly through a dominant-negative mechanism. **Patients and Results:** The proband was a 3-year old male with hypertelorism, a wide anterior fontanelle, upper eyelid colobomas, deficient bony orbits with pseudoproptosis, small low-set dysplastic cupped ears, syndactyly of fingers, bilateral talipes, bilateral undescended testes, imperforate anus and hypertrichosis. Parent/child trio-based exome sequencing revealed a *de novo* c.350A>T (p.Glu117Val) substitution in *TWIST1*. As the patient did not resemble typical SCS, the remaining exome was analysed but no other convincing pathogenic changes were identified. Presentation of the findings at the UK Dysmorphology group prompted *TWIST1* sequencing of a female subject with similar, but more severe features including bilateral eyelid colobomas, hypertelorism, choanal atresia, camptodactyly and abnormal hair distribution. This revealed a different heterozygous mutation at the identical nucleotide (c.350A>G; p.Glu117Gly). Of note, mutations at TWIST1-Glu117 have never been reported in association with SCS. This residue lies in the DNA binding domain, is highly conserved and makes specific contacts with bases involved in DNA binding sequence specificity. A heterozygous mutation at the equivalent residue (p.Glu29Lys) in the *C.elegans* homolog *hlh-8* shows dominant inheritance whereas null alleles are recessive (Corsi *et al*, *Development* **129**:2761;2002). We are currently investigating the transactivating activity of the above mutations, as well as previously reported SCS mutations, on E-box-luciferase reporter constructs. **Discussion:** On the basis of the similarity of the patients' clinical features, and their dissimilarity to SCS, we propose that heterozygous substitutions at Glu117 of TWIST1 are associated with a unique, severe phenotype. Whereas in SCS, haploinsufficency of *TWIST1* causes a change in homo/heterodimer balance in the coronal suture leading to craniosynostosis, in the cases presented here, a dominant-negative mechanism may act at an earlier stage of development to disrupt facial tissues derived from the cranial neural crest.

2731T

Autosomal Dominant Opitz GBBB Syndrome. *PS. Kruszka[1], BD. Solomon[2], RK. Iyer[2], M. Ahmad[2], DC. Thach[2], D. Bodian[2], WS. Wong[2], RA. Hart[1], AF. Martinez[1], NH. Robin[3], JM. Opitz[4], JG. Vockley[2], J. Niederhuber[2], M. Muenke[1].* 1) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Inova Translational Medicine Institute (ITIMI), Falls Church, VA; 3) Department of Genetics, University of Alabama Birmingham, AL; 4) Department of Pediatrics, University of Utah, Salt Lake City, UT.

Background: Opitz GBBB syndrome is a well-described genetic syndrome characterized by hypertelorism, hypospadias, dysphagia, and congenital heart disease. Opitz GBBB syndrome has been linked to Xp22 (XLOS) and 22q11 (ADOS) previously by our group. The gene XLOS has been identified, *MID1*; however, the gene for ADOS has not been identified. A number of case reports have found 22q11.2 deletions in patients described as having Opitz GBBB syndrome. Phenotypic differences have been described between XLOS and ADOS by our group in the past; however, with newer genetic information available, further discernible genotype-phenotype relationships can be made. We present 24 families with ADOS for phenotyping and genotyping. **Methods:** 24 families with Opitz GBBB syndrome were identified and family history and linkage analysis were completed as previously described using highly polymorphic microsatellite markers. All families were tested for the *MID1* mutation. Probands negative for *MID1* mutations underwent chromosomal microarray analysis, and are currently undergoing whole exome sequencing for chromosome 22. **Results:** 24 families were found to be autosomal dominant based on linkage to 22q11, male-to-male transmission in their pedigrees, and/or negative for mutations in *MID1*. The 19 probands that underwent chromosomal microarray analysis tested negative for copy number variations in the 22q11 deletion syndrome critical region. **Conclusion:** We have shown that ADOS is a clinical entity distinct from XLOS and from 22q11.2 deletion syndrome. Whole exome sequencing of chromosome 22 will likely identify the causative gene for this syndrome, and gene discovery for ADOS will have a significant impact on counseling affected families about transmission risks and in their decision for reproductive options. Additionally, finding the responsible gene including functional studies will add scientific insight into molecular pathways involved in normal and abnormal midline craniofacial, cardiac and urogenital development.

2732S

Clinical and pathological features of an infant with concurrence of C syndrome and renal-hepatic-pancreatic dysplasia. *C. Peña-Padilla[1], G. Razo-Jiménez[2], G. Tavares-Macías[2], L. Arnaud-López[1], L. Bobadilla-Morales[1,3], E. Zapata-Aldana[1], E.L. Mellín-Sánchez[1], J.R. Corona-Rivera[1,3].* 1) Centro de Registro e Investigación sobre Anomalías Congénitas (CRIAC), Servicio de Genética y Unidad de Citogenética, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca", Guadalajara, Jalisco, México; 2) Anatomía Patológica, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca", Guadalajara, Jalisco, Mexico; 3) Instituto de Genética Humana "Dr. Enrique Corona Rivera", CUCS, Universidad de Guadalajara, Jalisco, México.

Introduction. C syndrome [(CS) MIM 211750] is an autosomal recessive (AR) entity manifested by trigonocephaly, mental retardation, congenital cardiac defects, facial dysmorphism, postaxial polydactyly, CNS and renal malformations. CS is caused by mutation in CD96, mapped on 3q13. Otherwise, renal-hepatic-pancreatic dysplasia [(RHPD) MIM 208540] is a rare AR ciliophatie characterized by renal cystic dysplasia (RCD), hepatic fibrosis (HF), hepatic cyst (HC), and pancreatic cyst (PC) or dysplasia, due to mutations in NPHP3, whose locus was at 3q22.1. The concurrence of these two entities in a same patient has not been reported previously. Here we present a male infant with clinical pathological findings of CS and RHPD, supporting a common pathogenesis as both entities have the same mode of inheritance and chromosome localization. **Clinical report.** The *propositus* was the product of the third pregnancy of nonconsanguineous, young, and healthy parents. He was born after uncomplicated pregnancy and normal vaginal delivery at 38 weeks of gestation. Birth weight was 2340 g (below 10th centile); length was 48 cm (25th centile), and occipitofrontal circumference (OFC) was 34 cm (below 10th centile). Physical examination showed trigonocephaly, metopic ridge, upward slant palpebral fissures, *oral frenula*; postaxial polydactyly, and brachydactyly of both hands, and syndactyly of 2-3 toes. The CT brain scan showed corpus callosum hypoplasia and colpocephaly. 3D CT skull reconstructions confirm craniosynostosis of the metopic suture. He showed a 46,XY karyotype at the 550-band level. The patient died of renal failure and sepsis at 1 month of age. Autopsy results found congenital HF, bilateral RCD and multiple PC, and additionally a right double ureter. using the AmpFLSTR Identifiler PCR Amplification kit we found homozygosity for allele 15 in *locus* D3S1358. **Conclusions.** Clinical and pathologic findings confirm the diagnoses of CS and RHPD in our patient. Although RCD has been occasionally reported in CS, other obligate features of RHPD such as PC or HF, are not parts of the CS. This is the first time that the combination of CS and RHPD is reported. Since uniparental isodisomy poses a theoretical risk for AR disorders, as in the case of the CS and the RHPD, both located in 3q, a complete or partial chromosome 3 isodisomy is proposed to explain this unusual concurrence.

2733M

Molecular genetic study of 75 patients with X-linked alpha-thalassemia and mental retardation (ATR-X) syndrome in Japan. *H. Shimbo[1], K. Kurosawa[1], N. Okamoto[2], S. Ninomiya[3], T. Wada[4].* 1) Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan; 2) Osaka Medical Center and Research Institute for Maternal and Child Health; 3) Department of Clinical Genetics, Kurashiki Central Hospital, Kurashiki, Japan; 4) Department of Medical Ethics and Medical Genetics, Kyoto University Graduate School of Medicine, Kyoto, Japan.

[Introduction] ATR-X syndrome (MIM#603040) is an X-linked intellectual disability syndrome caused by mutations in the ATRX gene. The syndrome is characterized by severe intellectual disability, dysmorphic facies, hypotonia, genital and skeletal abnormalities, and the presence of α-thalassemia. More than 200 patients have been diagnosed worldwide; more than 80 reside in Japan. We have been studying ATR-X syndrome for more than a decade and have diagnosed almost all of the Japanese patients molecular genetically. In order to delineate molecular and clinical characteristics, we reviewed identified mutations in the *ATRX* gene of 75 patients from 62 families in Japan. [Methods] We analyzed genomic DNA and/or cDNA synthesized from RNA extracted from the leucocytes of patients. PCR was performed using 40 and 12 pairs of primers for genomic DNA and cDNA, respectively, followed by sequencing by Sanger method. [Results] Most of the mutations are clustered in two functionally important regions: 35 out of 75 mutations (47%) reside in ADD (ATRX-DNMT3-DNMT3L) domain, 27 (36%) in helicase domain, and the other 13 (17%) are outside of these domains. The types of mutations are as follows: 55 missense mutations, 2 nonsense mutations, 5 small deletions, 5 frame shifts, 3 intragenic deletions, 2 nucleotide exchanges in introns, 1 large insertion in an intron, 2 nucleotide exchanges in 5'- and 3'-UTR, and 1 large deletion. [Discussion] We did not find a clear correlation between genotype and phenotype. However, patients with typically severe clinical features tended to have mutations in ADD domain. We identified several cases with atypical clinical features. (1) A case with intractable seizures with an intragenic deletion of exon 1 of the 35 exons in the *ATRX* and exons 2-10 in the neighboring gene MAGT1, which would result in a premature translation stop codon. (2) Two siblings presented with clinically quite different severity with the intragenic deletion involving exons 2-5 in the *ATRX,* resulting in a prematurely truncated protein, [p.Ser7-Argfs*13]. A null mutation of *ATRX* is lethal because the *Atrx* knockout mouse cannot be viable. Proteins of ATR-X patients should have some residual functions. However, the proteins of these patients are considered to have no functions, because their stop codons reside near the 5' end. This may suggest the existence of a mechanism of re-initiation at the downstream ATG codon, or the possibility of other mechanisms.

2734T

Impact of Plexiform Neurofibromas on Adult Patients with Neurofibromatosis type 1. *S. Stueber, R.J. Hopkin, E. Schorry, L. Martin, K. Wusik.* Cincinnati Children's Hospital, 3333 Burnett Ave, Cincinnati, OH.

Neurofibromatosis type 1 (NF1) is one of the most common autosomal dominant genetic conditions diagnosed today. Up to 50% of patients with NF1 will develop a specific type of neurofibroma tumor called plexiform neurofibromas (PNFs), which develop along the nerve endings or along multiple nerve systems. Though typically benign, these tumors can cause significant complications. Some of the more common morbidities reported to be associated with PNFs are pain, loss of function, disfigurement, risk of malignancy development, mortality, and risk of bleeding complications. We performed a retrospective chart review of 69 adult patients that investigated rates of surgery and development of malignancy in patients with NF1 and PNFs. Results demonstrated that 50.7% of patients required surgery, and patients with greater numbers of PNFs were more likely to have surgery ($p = 0.02$). PNFs located in the head/neck and paraspinal region had a higher frequency of surgeries compared to those located in the pelvis or extremities ($p = 0.026$). In our patient population malignancies arising from PNFs were significantly more likely to occur in a core location as opposed to in the extremities or head/neck ($P<0.05$). While pain was not consistently documented in patient charts, in the 35 patients where pain could be reported, 60% had chronic pain. Taken together these results suggest that the adult population has significant tumor morbidity associated with PNFs in relation to surgery rates and malignancies. Currently, there is not a lot of information about how adult patients are experiencing PNFs, or how to manage this population. This study highlights the need of adult NF1 patients with PNF to have ongoing monitoring of their tumors due to significant morbidities. Future studies will be needed to determine modality and intervals of screening for best outcomes.

2735S

Early Onset Epileptic Encephalopathy Caused by *de novo SCN8A* Mutations. *H. Saitsu[1], C. Ohaba[1,2], M. Kato[3], S. Takahashi[4], T. Lerman-Sagie[5], D. Lev[6], H. Terashima[7], M. Kubota[7], H. Kawawaki[8], M. Matsufuji[9], Y. Kojima[10], A. Tateno[10], H. Goldberg-Stern[11], R. Straussberg[11], D. Marom[11], E. Leshinsky-Silver[12], M. Nakashima[1], K. Nishiyama[1], Y. Tsurusaki[1], N. Miyake[1], F. Tanaka[1], N. Matsumoto[1].* 1) Dept Human Gen, Grad Sch Med, Yokohama City Univ, Yokohama, Japan; 2) Dept Clin Neurol & Stroke Med, Yokohama City Univ, Yokohama, Japan; 3) Dept Pediatrics, Yamagata Univ Sch Med, Yamagata, Japan; 4) Dept Pediatrics, Asahikawa Med Univ, Asahikawa, Japan; 5) Ped Neurol Unit, Wolfson Medical Center, Holon, Israel; 6) Metabolic Neurogenetic Clinic, Holon, Wolfson Medical Center, Holon, Israel; 7) Div of Neurol, Nat Center for Child Health and Dev, Tokyo, Japan; 8) Dept of Pediatric Neurol, Osaka City General Hosp, Osaka, Japan; 9) Dept Pediatrics, Japan Community Health Care Org Kyusyu Hosp, Kitakyushu, Japan; 10) Dept Pediatrics, Toho Univ Sakura Med Center, Chiba, Japan; 11) Dept of Neurogenetics, Schneider's Children Med Center, Petah Tiqwa, Israel; 12) Institute of Med Genetics, Wolfson Medical Center, Holon, Israel.

Purpose: *De novo SCN8A* mutations have been reported in patients with epileptic encephalopathy. Here we report 7 patients with *de novo* heterozygous *SCN8A* mutations, which were found in our comprehensive genetic analysis (target capture or whole exome sequencing) for early-onset epileptic encephalopathies (EOEEs).

Methods: A total of 163 patients with EOEEs without mutations in known genes, including 6 with malignant migrating partial seizures in infancy (MMPSI), and 60 with unclassified EOEEs, were analyzed by target capture (28 samples) or whole exome sequencing (135 samples).

Results: We identified *de novo SCN8A* mutations in 7 patients: 6 of 60 unclassified EOEEs (10.0%), and 1 of 6 MMPSI cases (16.6%). The mutations were scattered through the entire gene: 4 mutations were located in linker regions, 2 in the fourth transmembrane segments, and 1 in the C-terminal domain. The type of the initial seizures was variable including generalized tonic-clonic, atypical absence, partial, apneic attack, febrile convulsion, and loss of tone and consciousness. Onset of seizures was during the neonatal period in 2 patients, and between 3 and 7 months of age in 5 patients. Brain MRI showed cerebellar and cerebral atrophy in 1 and 6 patients, respectively. All patients with *SCN8A* missense mutations showed initially uncontrollable seizures by any drugs, but eventually one was seizure-free and 3 were controlled at the last examination. All patients showed developmental delay or regression in infancy, resulting in severe intellectual disability. Our data reveals that *SCN8A* mutations can cause variable phenotypes, most of which can be diagnosed as unclassified EOEEs, and rarely as MMPSI. Together with previous reports, our study further indicates that genetic testing of *SCN8A* should be considered in children with unclassified severe epilepsy.

2736M

Recessive *TBC1D24* mutations cause early-onset epileptic encephalopathy and sensorineural hearing loss. *K. Writzl[1], D. Neubauer[2], D. Paro Panjan[3], B. Gnidovec Stražišar[2].* 1) Institute of Medical Genetics, University Medical Centre, Ljubljana, Slovenia; 2) Department of Child, Adolescent & Developmental Neurology, University Children's Hospital, Ljubljana, Slovenia; 3) Department of Neonatology, University Children's Hospital, Ljubljana, Slovenia.

Recent studies have shown that recessive mutations in the *TBC1D24* gene cause a variety of epilepsy syndromes (familial infantile myoclonic epilepsy, early infantile epileptic encephalopathy 16), nonsyndromic deafness and DOORS syndrome (Deafness, Onychodystrophy, Osteodystrophy, Mental retardation and Seizures). We report on two siblings with early-onset epileptic encephalopathy and deafness. The patients presented with clonic and myoclonic jerks within one hour after birth. The seizures were resistant to treatment. They had severe neurological impartments with axial hypotonia, upper limb dystonia, dyskinetic limb movements and poor eye contact. The older sibling died at the age of two years due to respiratory failure following severe respiratory infection. Audiologic examination showed bilateral sensorineural hearing loss in both siblings. The clinical and radiological investigations didn't show any fingers or toes abnormalities. Genetic analysis using next-generation sequencing panel for epileptic encephalopathy revealed compound heterozygous mutations in the *TBC1D24* gene: a novel missense mutation c.32A>G (p.Asp11Gly) in exon 2 and a frameshift mutation c.1008delT (p.His336Glnfs*12) in exon 4. Interestingly, the frameshift mutation was previously reported as unique to DOORS syndrome, yet our siblings did not meet the DOORS syndrome criteria. This report supports previous observations that mutations in *TBC1D24* cause diverse phenotypes. We recommend testing all infants with early onset epileptic encephalopathy and sensorineural hearing loss for *TBC1D24* mutations.

2737T

Kindler Syndrome: Novel and Recurrent *FERMT1* Mutations in Iranian Families. *L. Youssefian[1,2], H. Vahidnezhad[1,2,3], M. Barzegar[4], Q. Li[2], S. Sotoudeh[1], A. Yazdanfar[5], AH. Ehsani[1], AM. Kajbafzadeh[1], F. Agha-hosseini[1], S. Zeinali[3], M. Tabrizi[1], J. Uitto[2].* 1) Tehran University of Medical Sciences, Tehran, Iran; 2) Thomas Jefferson University, Philadelphia, PA., USA; 3) Pasteur Institute of Iran, Tehran, Iran; 4) Shahid Beheshti University of Medical Sciences, Tehran, Iran; 5) Hamedan University of Medical Sciences, Hamedan, Iran.

The Kindler syndrome (KS) is a rare autosomal recessive genodermatosis characterized by diffuse poikiloderma, cutaneous atrophy, acrokeratosis, trauma-induced blisters and photosensitivity. Mutations have been previously disclosed in the *FERMT1* gene which encodes kindlin-1, a component of keratinocyte focal adhesions. The incidence of KS is expected to be high in Iran due to high rate of consanguineous marriages. We identified a total of 13 Iranian families with KS, including ~40 affected individuals. Pedigree analysis of these families suggested high degree of consanguinity. The patients demonstrated characteristic clinical features, and the diagnosis was confirmed by skin biopsy in one of the cases while the parents, obligate heterozygote carriers, were clinically unaffected. Mutation analysis of *FERMT1* by amplification of all 15 exons and flanking intronic sequences, followed by bidirectional sequencing, revealed homozygous nonsense mutations in 7 Iranian families, both recurrent and novel, including c.1176T>G, c.1383C>A, c.550_551insA, c994_995delCA, and c.910G>T. In other 2 families three novel mutations including one missense and two splice site mutations [c.889 A>G, c.1139+2 T>C], c.957+1G>A were detected. There was no mutation in *FERMT1* gene in the other remaining families; promoter analysis and further studies are ongoing. These families, some with atypical clinical features, may present new entities with clinical overlapping features with the Kindler syndrome. These previously unpublished mutations in the *FERMT1* gene expand the spectrum of the mutations underlying KS. The identified mutation database forms the basis to confirm the clinical diagnosis by genetic testing and to identify heterozygous carriers, coupled with genetic counseling to reduce the burden of KS in Iran.

2738S

A Case report of disorders of sex development: Female patient whit novel mutation in AR gene and 45,X[5]/46,XY[95] karyotype. *J. Prieto, G. Giraldo, LP. Barragan-Osorio.* Inst de Gen Humana, Univ Javeriana, Bogota Cundinama, Colombia.

Introduction: Disorders of sex development (DSD) are a group of rare conditions usually presented in early infancy with external and/or internal reproductive organs abnormalities usually caused by gonadal disorders, adrenal or hormonal function impairment. DSD overall birth prevalence is estimated to be higher than 1 in 300. Androgen insensitivity syndrome (AIS) is typically characterized by evidence of feminization (i.e., undermasculinization) of the external genitalia at birth, abnormal secondary sexual development in puberty and infertility in individuals with a 46,XY karyotype. Here we report a novel mutation in AR in a female patient with 45,X[5]/46,XY[95] karyotype. Case Report: 26-year-old female patient who consulted by primary amenorrhea. At physical exam she presented a female phenotype with breast and pubic hair tanner of I-II/V. She had not body or axillary hair. Also she presented an adhered mass to the deep planes in the right inguinal region. The pelvic ultrasound did not show uterus neither ovarian follicles. We ordered an MRI which displayed an appropriate differentiation of the urethra, a short vagina that ends in blind pouch, defect in the formation of Muller and Wolf duct derivatives, a gonadal tissue in right inguinal region and a lower left cavity with solid appearance without cysts. The breast ultrasound evidenced heterogeneous breast tissue without masses. Her karyotype was 45,X[5]/46,XY[95]. With all of these findings we decided to order AR gene sequencing which reported a deleterious mutation c.202C>T (p.Q68X). Discussion: Here we presented a clinical chart consistent in AIS, female phenotype and 45,X[5]/46,XY[95] karyotype. The sequence analysis identified a c.202C>T change in the AR gene, which predicts an amino acid substitution of Glutamine to a premature termination codon at residue 68(p.Q68X). This mutation has not been previously reported but is interpreted as disease-causing in databases.

2739M

Complex Genomic Presentation in the NICU. *A. Khromykh[1], B.D. Solomon[1,3], D.L. Bodian[1], R.K. Iyer[1], R. Baveja[2], S.W. Wong[1], K.C. Huddleston[1], E.Z. Klein[1], D. Ascher[3], J.G. Vockley[1], J. Niederhuber[1].* 1) Inova Translational Medicine Institute, Falls Church, VA; 2) Fairfax Neonatal Associates, Falls Church, VA; 3) Inova Children's Hospital, Falls Church, VA.

Congenital anomalies are a leading cause of infant morbidity and mortality. The goal of our IRB-approved 'Impact of Genetic Disorders' study is to use trio-based whole genome sequencing (WGS) to elucidate the genetic and genomic factors that contribute to the etiologies of congenital anomalies, as well as assess these methods in real-time clinical practice, evaluating feasibility and cost-benefit ratio vs. conventionally available clinical genetic tests. Analysis was done through our ongoing trio-based (mother, father, proband) study (n=80+) that focuses on the subset of the NICU patient population suspected of having an underlying etiological genetic component and who underwent a comprehensive but unrevealing clinical/genetic workup. Our research utilizes trio-based WGS, RNA expression, methylation, and miRNA characterization. Clinical data are obtained from the electronic health records as well as parental input on nutrition, stress, behavior and environmental exposure via study-specific surveys. A variety of principal component, familial-based, pathway and genomic network analyses are utilized to identify the underlying genomic explanation of the observed clinical pesentation. Our analyses are bolstered by our database of >5,000 WGS (along with other biological data and clinical information) derived from multiple ongoing trio-based genomic studies. Data generated from an initial subset of 30 probands reveals that not only were we able to bioinformatically identify the causative genetic/genomic mutations in multiple probands in our cohort utilizing WGS including molecular evidence for novel findings as well as known conditions such as IMAGE syndrome, DiGeorge syndrome, Rubinstein-Taybi II, D-bifunctional protein deficiency, CHARGE syndrome, GATA4, none of which were diagnosed through standard means, but that the current cost of clinical genetic/etiological testing was ~$3,500 per patient, roughly double the cost of single proband WGS in our research institute (and approximately 2/3 the cost of trio-based WGS). Our study demonstrates the potential of WGS analysis to identify molecular etiologies for previously undiagnosed patients in an efficient and increasingly cost-effective way. The establishment of our analysis pipeline can allow effective, timely analysis of hundreds of cases per year. Our genomic counseling infrastructure provides physicians and patients with a wide range of services, making WGS a usable diagnostic tool.

2740T

Deep sequencing detects very low-grade somatic mosaicism in the unaffected mother of siblings with nemaline myopathy. *E. Koshimizu[1], S. Miyatake[1], Y.K. Hayashi[2,5], K. Miya[3], M. Shiina[4], M. Nakashima[1], Y. Tsurusaki[1], N. Miyake[1], H. Saitsu[1], K. Ogata[4], I. Nishino[5], N. Matsumoto[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Department of Neurophysiology, Tokyo Medical University, Tokyo, Japan; 3) Department of Pediatrics, Faculty of Medicine, University of Toyama, Toyama, Japan; 4) Department of Biochemistry, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 5) Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, Japan.

When an expected mutation in a particular disease-causing gene is not identified in a suspected carrier, it is usually assumed to be due to germline mosaicism. We report here very-lowgrade somatic mosaicism in ACTA1 in an unaffected mother of two siblings affected with a neonatal form of nemaline myopathy. The mosaicism was detected by deep resequencing using a next-generation sequencer. We identified a novel heterozygous mutation in ACTA1, c.448A>G (p.Thr150Ala), in the affected siblings. To explore the effect of this mutation, we mapped the mutation onto reported crystal structures. Thr150 is located near the polymerization/interaction interfaces between actin monomers and between actin and its interacting proteins. Thus, p.Thr150Ala may affect polymerization and/or the interactions of actin with other proteins. In this family, we expected autosomal dominant inheritance with either parent demonstrating germline or somatic mosaicism. Sanger sequencing identified no mutation. However, further deep resequencing of this mutation on a next-generation sequencer identified very-low-grade somatic mosaicism. The total read depth at c.448A in ACTA1 was 131495× to 425933×. Very-low-grade somatic mosaicism was confirmed in the mother: 0.4%, 1.1%, and 8.3% in the saliva, blood leukocytes, and nails, respectively. We used allele-specific PCR to confirm the presence of the mutation in the mother. Both the wild-type and mutant alleles were amplified in the proband and the affected sister at a similar level. Both alleles could also be amplified in the mother, but the wild-type allele was amplified at a much greater level than the mutant allele. The wild-type allele only was amplified in the father. We performed High Resolution Melting (HRM) analysis using DNAs from normal controls, the affected siblings, the father (all DNA derived from blood), and the mother (DNA derived from the nails, which showed the highest rate of mosaicism (8.3%)). The melting curves of both affected siblings were aberrant and were called mutant, but those of the father and mother were called normal. Our study demonstrates the possibility of very-low-grade somatic mosaicism in suspected carriers, rather than germline mosaicism.

2741S

Identification of *RIT1* mutations in patients with RASopathies by clinical whole exome sequencing. *P. Liu[1], F. Xia[1], W. He[1], L. Potocki[1], F. Scaglia[1], P. Magoulas[1], D.M. Muzny[2], A.L. Beaudet[1], R.A. Gibbs[1,2], C.M. Eng[1], Y. Yang[1].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

RASopathies including Noonan syndrome (NS) are a group of autosomal dominant developmental syndromes characterized by short stature, distinctive facial features, and congenital heart defects. These disorders are caused by germline mutations that alter the activity of the RAS/MAPK signaling pathway and the number of causal genes has increased to 14 over the years. Recently, potential gain-of-function mutations in *RIT1*, which encodes a small GTPase that belongs to the Ras superfamily of low molecular weight GTP-binding proteins, were reported in nine unrelated Japanese patients with NS, indicating *RIT1* as a new RASopathy gene (Aoki *et al., AJHG* 2013 *93*:173). Seventy percent of the reported patients with *RIT1* mutations presented with hypertrophic cardiomyopathy. Here, we report additional patients with potentially gain-of-function mutations in *RIT1* identified by clinical whole exome sequencing (WES). Patient 1 is a 2-year-3-month old Latin American female with developmental delay, dysmorphic features including low set ears, broad and flat face, malar hypoplasia, downturned corners of the mouth, crowded dentition, widened neck posteriorly, congenital cardiovascular disease with pulmonary valve stenosis, infundibular stenosis, and pulmonary regurgitation, history of cystic hygroma *in utero*, history of bronchopulmonary dysplasia, prematurity, history of bilateral moderate hearing loss. She was found to have a *de novo* novel heterozygous c.229G>T (p.A77S) mutation in the fourth exon of *RIT1*. Patient 2 is an 8-month-old Latin American male with developmental delay, dysmorphic features including upslanting palpebral fissures and clinodactyly of fifth fingers bilaterally, failure to thrive, aortic stenosis and pulmonic stenosis. He was found to have a *de novo* heterozygous c.104G>C (p.S35T) mutation in the second exon of *RIT1*. The same mutation was also reported previously in two unrelated probands with NS. Both of these two mutations are located within the evolutionarily conserved GTPase domain in *RIT1*. Additionally, we identified the recurrent c.104G>C (p.S35T) mutation in a third patient who has a clinical diagnosis of NS, demonstrating the c.104G position as a mutation hotspot and p.S35T as a common mutation for *RIT1* related RASopathy. Our results expand the molecular and phenotypic spectrums of RASopathy caused by *RIT1* mutations, and demonstrate the advantage of WES in diagnosing newly delineated genetic disorders.

2742M

Splicing mutation in *IQSEC2* gene modulating the phenotype in three siblings with intellectual disability. *I. Madrigal[1,2], J. Rosell[3], L. Rodriguez-Revenga[1,2], MI. Alvarez-Mora[1,2], O. Karlberg[4], D. Elurbe[1,2], A. Bayes[5], S. Sauer[4], AC. Syvänen[6], M. Mila[1,2].* 1) Hospital Clinic Barcelona and IDIBAPS, Barcelona, Spain; 2) Centre for Biomedical Research on Rare Diseases (CIBERER), Spain; 3) Genetics Section, Hospital Universitari Son Espases, Palma de Mallorca, Spain; 4) Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 5) Molecular Physiology of the Synapse Laboratory IIB, Sant Pau, Barcelona, Spain; 6) Max-Planck Institute for Molecular Genetics Ihnestrasse 63-73 D-14195 Berlin Germany.

The *IQSEC2* (IQ motif and Sec7 domain 2) gene is located on chromosome Xp11.22 and encodes a guanine nucleotide exchange factor for the ADP-ribosylation factor family of small GTPases. This gene is known to play a significant role on cytoskeletal organization, dendritic spine morphology and synaptic organisation. Mutations in *IQSEC2* cause moderate to severe intellectual disability (ID) in all affected males and a variable phenotype in females due to this gene escapes X chromosome inactivation. Here, we report on a novel splicing mutation in *IQSEC2* gene (g.83032_83033delCA) that cosegregates with the disease in a family diagnosed with a nonsyndromic form of X-linked ID. The mutation is a deletion of 2bp (c.3116-2delCA) that activates an intraexonic splice acceptor site resulting in 78 nucleotides deletion in exon 12 (c.3116_3194del). The activation of this acceptor site abolishes 26 amino acids from the highly conserved PH domain of *IQSEC2* and creates a premature stop codon 36 amino acids latter in exon 13. Two separate PCR products were obtained when *IQSEC2* mRNA from exons 9 to 13 was analyzed. Interestingly, the percentage of aberrant splicing in each patient correlates with the severity of their clinical features. As far as we are concerned, we are describing the first splicing mutation affecting *IQSEC2* gene that, moreover, modulates the severity of phenotype in these patients. Acknowledgements: This study was supported as a transnational access project at the European Sequenicng and Genotyping Infrastructure (ESGI) which has received funding from the Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 262055. Exome sequencing was performed by the SNP&SEQ Technology Platform (www.sequencing.se), Department of Medical Sciences, Science for Life Laboratory at Uppsala University, a national infrastructure supported by the Swedish Research Council (VR-RFI) and the Knut and Alice Wallenberg Foundation. We also thank AGAUR from the Autonomous Catalan Government (2009-SGR1337) and Fundación Agrupación Mutua (Premio Discapacidad 2012). The CIBER of Enfermedades Raras is an initiative of the ISCIII.

2743T

A NOVEL MUTATION IN GAPO SYNDROME. *S. Sestito[1], I. Mascaro, M. Gris[1], M. Grisolia[1], F. Falvo, S. , Lenka[1], L. Piherova[2], P. Strisciulio[3], S. Kmoch[2], D. Concolino[1].* 1) Pediatrics, University "Magna Graecia", Catanzaro, Italy; 2) 1Institute for Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague, 120 00 Prague 2, Czech Republic; 3) Department of Pediatrics, University Federico II, Naples, Italy.

GAPO syndrome is a rare autosomal recessive disease and is an acronym composed of growth retardation, alopecia, pseudoanodontia, optic atrophy. Only few cases are reported in literature and the clinical spectrum of this rare disease has been expanded . Recently we described on a new patient with GAPO syndrome associated with hypoacusis and bilateral vestibular hypofunction (1) and the details of the progression of the clinical findings of GAPO syndrome in the girl from her birthday to 10 years of life. The genetic cause of GAPO syndrome has now been identified and mutation in ANTXR gene1, which encodes anthrax toxin receptor 1 has been reported by Stránecký et al. (2) in four ethnically unrelated affected individuals. We performed a molecular analysis in our case and homozygous deletion potentially affecting splicing of ANTXR1 (NM_032208.2:c.704-2_720del19) has been identified. Both parents are heterozygous carriers of the mutation. This mutation is not reported in previously described cases. 1 Concolino D et. al, GAPO syndrome associated with vestibular dysfunction and hearing loss. Am J Med Genet A. 2013 Aug;161A(8):2102-4. 2 Stránecký V et al., Mutations in ANTXR1 cause GAPO syndrome, Am J Hum Genet. 2013 May 2;92(5):792-9.

2744S

Mutation identification in New Zealand populations: a pilot study in neurodevelopmental disorders. *J. Jacobsen[1], S. Robertson[2], C. Wilson[3], B. Swan[1], E. Glamuzina[3], J. Taylor[4], R. Hill[5], D. Love[6], K. Lehnert[1], R. Snell[1].* 1) Centre for Brain Research and School of Biological Sciences, The University of Auckland, New Zealand; 2) Dunedin School of Medicine, The University of Otago, New Zealand; 3) Adult and Paediatric National Metabolic Service, Auckland City Hospital, New Zealand; 4) Genetic Health Service New Zealand, Auckland City Hospital, New Zealand; 5) Department of Neurology, Auckland City Hospital, New Zealand; 6) Diagnostic Genetics, LabPLUS, Auckland City Hospital, New Zealand.

New Zealand has a unique and diverse population for genetic studies. We are currently investigating the genetic underpinnings of genetically undiagnosed neurodevelopmental disorders in the New Zealand population. In a pilot study, using an exome sequencing approach, we have sequenced two affected individuals from each of four families and have searched for mutations assuming a recessive mode of inheritance. We have found putative mutations in three of the families, two of which are novel and are presented here.

The first family consists of two affected female siblings who presented with gait ataxia at 12 and 18 months of age. MRI revealed vermian hypoplasia and increased spaces between the folia. WES identified compound heterozygous inheritance of two novel mutations (non-synonymous and splice site) in the aarF domain containing kinase 3 (*ADCK3*) gene in both individuals. Interestingly, mutations in this gene result in coenzyme Q_{10} deficiency, with the possibility of treatment with oral Q_{10} supplementation.

The second family consists of two male siblings who presented with profound muscular hypotonia and developmental delay. Investigations revealed mildly low CSF neurotransmitter metabolites (HVA and 5_HIAA). The children were a result of a consanguineous union, and regions of homozygosity were identified by array CGH. Using WES we identified a missense mutation in a conserved transmembrane domain of the monoamine transporter *SLC18A2* gene (solute carrier family 18, member 2; VMAT2), which was present in both boys and located within a region of homozygosity. Successful treatment of a previously reported family with an *SLC18A2* gene mutation and similar clinical presentation has initiated a treatment strategy with the dopamine agonist pramipexole.

The success of this pilot sequencing study is due to the close cooperation between the research team, physicians and a strong genetic hypothesis. As the study expands in scope it has the potential to identify mutations unique to our New Zealand population, and deliver a genetic diagnosis for the participating families.

2745M

Whole exome sequencing identified a novel RAB3GAP1 mutation in Turkish patient with Micro Warburg Syndrome. *B. Yuceturk[1,2], H. Ulucan[1], A. Koparir[1], E. Kirat[1], O.F. Karatas[1,3], E. Fenercioglu[1], M. Seven[1], A. Yuksel[1,4], M. Ozen[1,5].* 1) Istanbul University, Cerrahpasa Medical School, Medical Genetics Department, Istanbul, Turkey; 2) TUBITAK, Kocaeli, Turkey; 3) Erzurum Technical University, Faculty of Science, Molecular Biology and Genetics Department, Erzurum, Turkey; 4) Biruni University, Istanbul, Turkey; 5) Baylor Collage of Medicine, Department of Pathology & Immunology, Houston, TX, USA.

Micro Warburg syndrome (#600118) is a rare autosomal recessive inherited disorder characterized by facial dysmorphism, eye abnormalities including cataract, microphthalmia, microcornea, brain abnormalities and endocrine abnormalities. Three genes, RAB3GAP1, RABGAP2 and RAB18 have been identified to be related with this syndrome. Martsolf syndrome is an allelic and milder form of Micro Warburg syndrome. Here we present a female with Micro Warburg syndrome. She had facial dysmorphism, severe motor-mental retardation, congenital cataract, microcephaly and corpus callosum hypoplasia. Since the associated genes are various and quite long, we utilized whole-exome sequencing (WES) as a diagnostic tool for identifying the molecular basis of Micro Warburg syndrome and all variants were confirmed by Sanger sequencing. WES is an affordable, efficient and powerful method and it is commonly applied as a diagnostic tool for identifying the molecular basis of genetically heterogeneous disorders. As a result a novel homozygous c. 2607-1G>C splice site mutation was detected in exon 23 of RAB3GAP1 and confirmed by Sanger sequencing. Parents also have been demonstrated to be heterozygous for this mutation.

2746T

Molecular Diagnosis of Congenital Limb Defect Syndromes by Next Generation Sequencing. *G. Mendiratta-Vij, J. Zhang, Y. Kasai, T. Brandt, L. Mehta, R. Kornreich, L. Edelmann.* Icahn School of Medicine at Mount Sinai, NYC, NY, USA.

Congenital Limb Defects (CLD) are etiologically and genetically heterogeneous and can occur as an isolated or syndromic defects. Prevalence of CLD is estimated to be at 1 in 1500 - 1 in 3000 live births. Common presentations include polydactyly, syndactyly, limb reduction or long bone deficiencies, and multiple anomalies or generalized skeletal syndromes. Genetic syndromes including CLD may be caused by germline mutations in genes involved in cell fate regulation and patterning during embryogenesis. Between 70% and 100% of syndromic CLD cases with a Mendelian etiology have a known causative mutation. Based on multiple reports, subject matter experts, and detection rates, we considered Townes-Brocks syndrome, Holt-Oram syndrome, Duane-radial ray syndrome, Greig cephalopolysyndactyly syndrome and Pallister-Hall syndrome, Brachydactyly-syndactyly syndrome, Robinow syndrome (RRS), Al-Awadi/Raas-Rothschild/Schinzel phocomelia syndrome (AARRS) and Werner mesomelic syndrome in the development of a next-generation sequencing (NGS) panel. The 7 well-established CLD genes to be interrogated include: SALL4, SALL1, TBX5, GLI3, HOXD13, ROR2, WNT7A, and a cis-regulatory sequence for SHH: ZRS. Agilent custom SureSelect target enrichment combined with the Illumina HiSeq platform was employed to capture coding exons plus 50 bp of upstream and downstream sequence to ensure capture of splicing regions. Analysis of paired-end reads was performed using NextGENe NGS analysis software. Our panel has been fully validated on different specimen types containing a variety of sequence variants including prenatal samples: direct chorionic villi, amniocytes of direct amniotic fluid, cultured villi, and amniocytes. We obtained 100% capture coverage with all bases in the coding and splicing regions enriched and sequenced to sufficient depth (~400X) in all 20 validation samples. The test sensitivity is 99% for detection of substitutions and 95% for detection of small insertions and deletions. The panel was also validated using the Rapid Run mode on the Illumina HiSeq and MiSeq to achieve faster turn-around-times for prenatal testing. Our panel provides a comprehensive, rapid, accurate and cost effective molecular diagnostic test for patients with CLD and prenatal cases with abnormal ultrasound findings and should improve the diagnostic yield in these situations.

2747S

Mutation spectrum and report of 4 novel mutations in *IDS* gene in Indian patients with Hunter syndrome. *G. Verma[1], M. Kabra[1], N. Gupta[1], P. Kaur[2], P. Mishra[1], S. Shastri[1], S. Sapra[1], M. Roychowdhury[1], S. Gulati[3].* 1) Pediatrics (Genetics), All India Institute of Medical Sciences, New Delhi, Delhi, India; 2) Biophysics, All India Institute of Medical Sciences, New Delhi, Delhi, India; 3) Pediatrics (Neurology), All India Institute of Medical Sciences, New Delhi, Delhi, India.

Introduction: Mucopolysaccharidosis type-II (MPS II, Hunter syndrome) is an uncommon X-linked recessive multisystem disorder with significant variability in both age of onset and rate of progression. It is caused by deficiency in the activity of the lysosomal enzyme, iduronate-2-sulfatase, synthesized by the 24 kb long *IDS* gene, located at Xq28. **Objective:** To identify the mutation spectrum of *IDS* gene in Indian patients with MPS-II. **Method:** Twenty two patients were recruited from 19 unrelated families after institutional ethics committee approval and informed consent based on the deficient leucocyte iduronate-2-sulfatase levels. Genomic DNA was amplified by PCR using 9 sets of primers that covered the entire coding region of the *IDS* gene, flanking intronic-splice regions and 5' and 3' UTRs. Bidirectional Sanger-sequencing was performed using ABI 3130 Genetic Analyzer. Sequences were analyzed using NCBI BLAST. **Results and Discussion:** Based upon the phenotypic analysis of 22 patients, 18 patients were classified as severe MPS II. Mutations were identified in 10 families out of 19. Missense (3/19, 16%), nonsense (2/19, 10%), splice-site (2/19, 10%), small insertion, small duplication and indel (1/19, 5%) each were identified. Four mutations c.1248C>T, c.205_206insAAAACTGGCAT, c.1435_1456dupAAGCCGAGTTTAAAAGATATAA and c.1442delGinsTC were novel and predicted to be pathogenic. Nonsense mutations were associated with severe phenotype in two patients. Prenatal diagnosis was done in one family and the fetus was found to be a carrier. Majority of the mutations (4/19, 21%), though nonrecurrent, were in exon 9 and its flanking splice-site region. *IDS/IDS2* recombination has not yet been tested and this could be the main reason for not finding mutations in the remaining 9 families. **Conclusion:** Mutation analysis of IDS gene is helpful in making a specific diagnosis, carrier testing and prenatal diagnosis. No specific genotype phenotype correlation was observed. Exon 9 appears to be the mutational hot spot in our Indian patients.

2748M

The first two AUTS2 mutations on the nucleotide level causing AUTS2 syndrome. *G. Beunders[1], S.A. Munnik de[2], J. Morton[3], P. Vasudevan[4], E. Voorhoeve[1], A.J. Groffen[1], W.M. Nillesen[2], H. Meijers-Heijboer[1], H.G. Yntema[2], E.A. Sistermans[1].* 1) clinical Gen, VU Med Ctr, Amsterdam, Netherlands; 2) Department of Human Genetics, Radboud university medical centre, Nijmegen, The Netherlands; 3) Department of Clinical Genetics, Birmingham Women's Hospital, Edgbaston, Birmingham, UK; 4) Department of Clinical Genetics, University Hospitals of Leicester, Leicester, UK.

Background: We recently described a new ID syndrome, AUTS2 syndrome that is characterised by low birth weight, feeding difficulties, intellectual disability, microcephaly and mild dysmorphic features. All cases thus far were caused by chromosomal rearrangements. Mutations at the base pair level disrupting AUTS2 have not yet been described.

Methods: Here we present the full clinical description of the first two cases with AUTS2 syndrome caused by mutations at the nucleotide level, found by diagnostic exome sequencing.

Results: The phenotypic features of both cases include: intellectual disability, microcephaly, feeding difficulties, dysmorphic features and mild contractures. Both cases have AUTS2 mutations in exon 7 (one nonsense mutation and one two-basepair deletion), both causing a premature stop of the full length transcript and without an predicted effect on the shorter 3' transcript starting in exon 9 that is also expressed in human brain.

Conclusions: The similarities between the phenotypes of these two cases and formerly described patients with AUTS2 syndrome, further confirm that AUTS2 syndrome is a single gene disorder with a recognizable phenotype that can be caused by haploinsufficiency of the full length transcript of AUTS2.

2749T

Novel mutations and clinical outcomes of copper-histidine therapy in Menkes disease patients. *G. Kim[1], J. Kim[2], J. Choi[2], B. Lee[1,2], H. Yoo[1,2].* 1) Medical Genetics Center, Asan Med Ctr, Seoul, South Korea; 2) Dept. Pediatrics, Asan Med Ctr, Seoul, South Korea.

Menkes disease is a very rare X-linked copper metabolism disorder caused by mutations in the ATP7A gene. With the advent of subcutaneous copper-histidine therapy, the early diagnosis of Menkes disease becomes of utmost importance for patients' prognosis. The clinical characteristics of 12 Korean patients with Menkes disease (11 males and 1 female from 11 unrelated families) were studied along with the mutation spectrum. Only 2 males were diagnosed in the neonatal period, and the other male patients were diagnosed at age 4.3 ± 1.9 months. The presenting signs included depigmented kinky hair, neurologic deficits, and hypotonia. Serum copper and ceruloplasmin levels were markedly decreased. Intracranial vessels were dilated with tortuosity and accompanied by regional cerebral infarctions, even at an early age. Notably, a female patient who presented with developmental delay, was diagnosed at age 18 months by characteristic MRA findings, biochemical profiles, and genetic analysis. CAG repeats in the AR gene at Xq12 were investigated to evaluate the X-inactivation pattern, and the patient revealed a 76% inactivated normal X-chromosome. A total of 11 mutations in ATP7A were identified, including five novel mutations: p.L825fs*1, c.121-930_2626+488del (exons 3-12 deletion), c.3744_4123+576del (exons 19-21 deletion), p.N686fs*1, and p.H1086Pfs*3. Most mutations were truncated (except 1 missense mutation), including 3 frameshift, 2 nonsense, 3 large deletion, and 2 splice-site variants. The age at commencement of copper-histidine treatment was 7.3 ± 7.5 (0.5 - 27) months. Despite treatment, seven patients died before age 5 years, and the remaining patients were severely retarded in neurodevelopment. The poor outcomes of our patients might be related to delayed therapy as well as truncated nature of ATP7A mutations.

2750S

Genetic analysis of an atypical case of Branchio-Oto-Renal (BOR) Syndrome. *R. Birkenhager, T. Jakob, E. Prera, S. Arndt, W. Maier, R. Laszig, A. Aschendorff.* Department of Otorhinolaryngology and Head and Neck Surgery, University Medical Center Freiburg, Killianstrasse 5, D-79106 Freiburg, Germany.

Introduction: Branchio-oto-renal (BOR) (MIM 113650) and Branchio-otic (BO) (MIM 602588) syndromes are one of the most common forms of autosomal dominant syndromic hearing loss. BOR/BO syndromes are genetically heterogeneous and caused by mutations in EYA1, SIX1 and SIX5 genes. The phenotypes of BOR syndrome are highly variable, with common clinical features including hearing impairment (HI), malformations of the pinnae, the presence of branchial fistulae, and various renal abnormalities, collecting system duplications; renal hypoplasia, dysplasia, and agenesis, additional features include inner ear anomalies such as dysplasia or hypoplasia of the cochlea, enlarged vestibular aqueduct or Mondini malformation and facial asymmetry. Middle ear defects described in BOR syndrome include hypodysplastic stapes, malleus-incus fusion and closed oval windows. BOR syndrome is clinically heterogeneous and has reduced penetrance. The prevalence of BOR syndrome was reported to be 1:40,000. In contrast Branchio-otic (BO) syndrome, patients suffer from inner ear and bronchial defects without renal abnormalities. In this study, clinical and genetic analyses were performed in a family with BOR syndrome focusing on auditory characterization and rehabilitation. Methods: Hearing testing BERA/Electrocochleography was only done in the Patient other family members were not tested because they have no hearing impairment or clinical abnormalities. Radiological a high-resolution CT scan was made. Genetic analysis was performed for EYA1, SIX1, and SIX5 genes of the affected individual and first degree relatives using direct sequencing of the coding exon and intron transitions of the genes, including deletion analysis. The modality and outcome of auditory rehabilitation were evaluated. The genotype-phenotype correlation was investigated. Results: One novel SIX1 mutation, a small deletion (c.376_378delGAG, p.Glu126del) was identified in the family with BOR syndrome. EYA1 and SIX5 mutations were not detected in the present study. No renal and branchial defects were observed in the patients. Cochlear implantation performed in one patient resulted in significant hearing improvement. Conclusions: Successful outcome can be expected with cochlear implantation in patients with BOR syndrome who cannot benefit from hearing aids. The novel SIX1 mutations may add to the genotypic and phenotypic spectrum of BOR syndrome in the European population.

2751M

Mutations in ERF gene as a new genetic cause of craniosynostosis - enabling parents and clinicians to understand why a child is affected. *A. Chaudhry[1, 2], P. Sabatini[2,3], L. Han[3], P. Ray[2,3], C. Forrest[2,4], S. Bowdin[1,2].* 1) Division of Clinical and Metabolic Genetics. Department of Paediatrics. The Hospital for Sick Children. 525 University Avenue. Toronto. ON. M5G2L3; 2) University of Toronto. Toronto. ON; 3) Molecular Genetics Division, Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto; 4) Division of Plastic & Reconstructive Surgery, Centre for Craniofacial Care & Research, The Hospital for Sick Children.

Craniosynostosis is a genetically heterogeneous condition resulting from the premature fusion of one or more cranial sutures. Several genes have been identified as a cause of craniosynostosis including FGFR1, FGFR2, FGFR3, TWIST, EFNB1, FAM20C and LMX1B. Twigg et al. recently published mutations in the ERF gene as a cause of 'ERF-related craniosynostosis' syndrome in about 6% of population they screened. The main clinical features included multiple-suture synostosis, craniofacial dysmorphism (hypertelorism, shortening and/or vertical displacement of the nose, prominent orbits and forehead), Chiari malformation, language delay and behavioral issues. These patients presented after infancy and did not need repeated cranial surgeries. ERF encodes an inhibitory ETS transcription factor directly bound by ERK/2(refs. 2-7). With the help of mouse models, the authors were able to show that reduced dosage of the ERF gene causes complex craniosynostosis in humans and mice. In our retrospective study, we sequenced the coding regions of the ERF gene in 43 patients who tested negative for mutations in FGFR1, FGFR2, FGFR3 and TWIST as well as by cytogenetics testing. In our cohort of patients with multi-suture or sagittal suture synostosis, we identified heterozygous pathogenic mutations in the ERF gene in five individuals (11.6%). The review of medical records revealed that three of our patients (60%) presented later in infancy. Subject A1 presented at 6 and half years of age with headaches and visual loss secondary to oxycephaly. 2/5 (40%) of our patients had other abnormalities on head imaging including Chiari I malformation. The dysmorphic features in our cohort include hypertelorism, proptosis, depressed nasal bridge and retrognathia in addition to the abnormal head shape as a result of craniosynostosis. These patients did not have developmental delays other than speech delays in one patient. They did not need repeated cranial surgeries in general. In conclusion, ERF-related craniosynostosis should be suspected in patients who present later in infancy with multiple suture or sagittal synostosis and who have other cranial malformations like Chiari I malformation. For the purposes of counseling, it is important to note that most of these individuals do not have developmental delays, but few might have speech delay. In general, they do not need repeated cranial surgeries.

2752T

COlobomatous Microphthalmia, Macrocephaly, Albinism & Deafness (COMMAD syndrome), a new syndrome caused by biallelic mutation of MITF: clinical characterization and molecular analysis. *A. GEORGE[1], M. RIUS[1], R.K.P ALUR[1], D.J. ZAND[3], S. YURI[1], K. BHARTI[2], B.P. BROOKS[1].* 1) Ophthalmic Genetics and Visual Function Branch, National Eye Institute, Bethesda, MD; 2) Section on Epithelial and Retinal Physiology and Disease, National Eye Institute, Bethesda, MD; 3) Dept. of Genetics, Children's National Medical Center, Washington, DC.

The MITF gene encodes for microphthalmia-associated transcription factor, which is a basic helix-loop-helix protein critical for development of neural crest-derived melanocyte, and neuro-ectoderm derived retinal pigmented epithelium. Mutations in MITF are associated with Waardenburg syndrome (WS). In this study, we report a rare family comprised of previously undiagnosed WS parents with disparate clinical manifestations due to heterozygous MITF mutations and a child with compound heterozygous mutations of MITF. This child exhibited extreme colobomatous microphthalmia, a complete lack of melanin pigment in the skin/hair/eyes, profound hearing loss, macrocephaly and low tone. DNA sequencing revealed a paternally inherited c.952_954delAGA MITF allele and a maternally-inherited c.G921C (p.K307N) allele the latter being previously unreported. For analysis of the two mutations at molecular level, GFP and FLAG peptide tagged wild type and mutant MITF-A isoforms were generated by PCR mediated cloning and site directed mutagenesis and were named as MITF-wt, MITF-K307N and MITF-R318del, respectively. Upon transfection of HEK293 cells with MITF expression constructs, MITF-wt was localized primarily in the nucleus. In contrast, MITF-R318del was observed in both nuclear and cytoplasmic compartments. The K307N mutant localization pattern was similar to MITF-wt. The localization pattern was further confirmed by Western blot analysis on isolated nuclear and cytoplasmic fraction of the transfected HEK cells. MITF-wt activated the M-Box, Tyrosinase related protein (TRP) 1, TRP-2, Tyrosinase (TYR) and Best1 promoters by dual luciferase assay. Interestingly MITF-K307N mutant also activated all of the above-mentioned promoters, while MITF-R318del mutant was unable to activate any of the tested promoters. Current studies are focused on how mutant MITF molecules may interact with the wild type protein, with each other, and with proteins important in ocular and other tissues. These results describe a novel syndrome caused by biallelic mutations in MITF (COlobomatous Microphthalmia, Macrocephaly, Albinism and Deafness (COMMAD syndrome) and a novel "activating" mutation in MITF gene.

2753S

A novel missense mutation of ryanodine receptor 1 (*RYR1*) in a Japanese idiopathic hyper CK-emia family. *K. Sano[1,2], S. Miura[1,2], T. Fujiwara[2], R. Fujioka[2], A. Yorita[1], K. Noda[1], H. Kida[1], K. Azuma[1], S. Kaieda[1], K. Yamamoto[3], T. Taniwaki[1], Y. Fukumaki[2], H. Shibata[2].* 1) Department of Respirology, Neurology and Rheumatology, Department of Medicine, Kurume University School of Medicine, Kurume, Fukuoka; 2) Division of Genomics, Medical Institute of Bioregulation, Kyushu University, Fukuoka; 3) Department of Medical Chemistry, Kurume University School of Medicine, Kurume, Fukuoka.

Persistent elevation of serum creatine kinase (CK) in individuals without any symptoms has been called idiopathic hyper CK-emia. We studied a four-generation pedigree of a Japanese family with idiopathic hyper CK-emia to elucidate the genetic basis of this disease. Nine members of the family including 5 affected individuals were enrolled in this study. Four affected members do not have any neurological abnormalities. Two affected members experienced continuous mild hyperthermia. In one affected member, slightly muscular weakness in bilateral iliopsoas, absence of patellar tendon reflexes, and impaired vibration sense in the lower limbs were observed. No significant distinguishing features on muscle biopsy (HE staining). The multipoint linkage analysis using SNP markers showed seven clear peaks of logarithm of odds scores (>1.4). We identified 65,180 patient-specific single nucleotide variants (SNVs) by exome analysis upon one patient and one unaffected relative. By the filtering processes using the information of the public SNP database, the result of our linkage analysis and the 40 candidate genes for disorders with possibly associated with idiopathic hyper CK-emia, we identified only one novel heterozygous non-synonymous SNV, c.7053 G>C, p.S2345T in the *RYR1* gene (MIM 180901). We confirmed cosegregation of the SNV with the phenotype in the family using the Sanger sequencing method. The PolyPhen-2 and PANTHER subPSEC scores of p.S2345T are 0.911 (possibly damaging) and -3.56 (probably damaging), respectively, indicating the SNV might cause the functional alteration. This variation was not detected in 420 control subjects. Immunochemical staining using the RYR1 antibody of the muscle biopsy specimen of one affected member showed weak staining in some muscle fibers. Moreover, western blotting revealed a substantial decrease of the RYR1 protein level. We also found that the mRNA expression of *RYR1* was reduced in the muscle of the affected member. These results suggest that the substitution of Ser with Thr at the position of 2,345 of *RYR1* may have a dominant negative effect. The *RYR1* gene is a causative gene of malignant hyperthermia or central core disease. In our study, the novel missense mutation of *RYR1* might cause idiopathic hyper CK-emia. In conclusion, when clinicians meet patients with idiopathic hyper CK-emia, it is important to consider *RYR1* mutations. This study was approved by the Ethics Committees of Kurume Univ. and Kyushu Univ..

2754M

***De novo* mutation in *SOX18* causes a novel form of Hypotrichosis-Lymphedema-Telangiectasia with severe vascular defects.** *F. Wünnemann[1,2], C. Preuss[1], V. Kokta[3], S. Leclerc[1], M. Thibault[1], J.C. Grenier[1], G. Andelfinger[1].* 1) CHU Sainte Justine Research Center , Montréal, Quebec, Canada; 2) Department of Biology, University of Münster, Germany; 3) Department of Pathology, CHU Sainte Justine, Université de Montréal, Quebec, Canada.

Hypotrichosis-Lymphedema-Telangiectasia (HLTS) is an ultra-rare, congenital condition that is characterized by blood and lymph vessel dysfunction, as well as hair follicle defects (OMIM #607823). Mutations in the transcription factor SRY (sex determining region Y)-box 18 (*SOX18*) have been implicated in HLTS for less than 10 pediatric cases globally. Here, we describe a 13 year old, female patient with anomalies in hair growth, subcutaneous teleangiectasia and dilation of the ascending aorta. First symptoms were noted from the age of six months on, with salient features being absence of eyebrows, hypertension and skin discolorations. Using a whole-exome sequencing approach, we now identified a novel *de novo* mutation in *SOX18* (c.481C>T) that results in a premature STOP codon (p.Q161*). The mutation in the highly conserved site (GERP score > 3) is located directly after the HMG box domain, truncating the protein of its transactivation domain, likely disrupting transactivation activity. We validated the heterozygous mutation by bi-directional Sanger sequencing and performed immunohistochemical staining of skin biopsies to reveal vascular identity of dilated vessels. These studies showed extensive dilation of subcutaneous capillary vessels, without any anomalies of the lymphatic system. Notably, the vascular defects in the presented case resemble the recently described double knockout Sox18/Vegfd mouse model, suggesting a role of vascular endothelial growth factor D (*VEGFD*) in the progression of HLTS. We highlight the use of whole exome sequencing as a valuable tool for the rapid identification of rare alleles involved in cardiovascular defects in a clinical setting. Taken together, our findings indicate a novel form of HLTS with absence of lymphatic malformations.

2755T

A novel case of Epidermolysis Bullosa with Pyloric Atresia due to homozygous mutation of c.600delC in integrin β4 gene: Clinical, Immunohistological and Molecular Diagnosis. *S. Yilmaz[1], I. Mungan Akin[2], C. Chiaverini[4], A. Charlesworth[4], D. Buyukkayhan[3], J.P. Lacour[4], I. Akalin[1].* 1) Medical Genetics, Istanbul Medeniyet University, Department of medical Genetics, Istanbul, Turkey; 2) Neonatology, Istanbul Medeniyet University, Istanbul, Turkey; 3) Neonatology, Goztepe Training and Research Hospital ,Istanbul Medeniyet University, Istanbul, Turkey; 4) Reference Centre for EBH, University Hospital of Nice, France.

Epidermolysis Bullosa (EB) with pyloric atresia (EB-PA) is a hemidesmosomal subtype of EB characterized by blister formation following minor trauma and pyloric atresia with/without renal disorders. The main pathology in EB-PA is basement membrane disruption due to abnormal expression of integrin and/or plectin proteins that are responsible for adherence of epidermis to lower layers in dermoepidermal junction. The genetic etiology of hemidesmosomal EB is homozygous or compound heterozygous mutations of PLEC, ITGA6 or ITGB4 genes. The proband was a male newborn with the findings of bilateral hydronephrosis, pyloric atresia and large aplasia cutis on four extremities. The parents were 2nd degree cousins and they had another newborn baby with the same clinical findings who died at 1st week. The patient was clinically diagnosed as EB-PA. Immunohistology of the biopsy specimen showed ITGB4 deficiency. Molecular analysis revealed a homozygous c.600delC mutation in exon 7 of ITGB4 gene. To our knowledge there is only one previous case of EB-PA with homozygous c.600delC mutation reported in the literature.

2756S

Increased Suseptibility to Attention Deficit Hyperactivity Disorder risk in Marfan Syndrome and Other Connective Tissue Disorders. *A. Hall, M. Tucker, L. Escobar.* Medical Genetics and Neurodevelopmental Center, St. Vincent Hospital, Indianapolis, IN.

Marfan syndrome is an autosomal dominant connective tissue disorder that involves the cardiovascular, ocular and skeletal systems and is caused by mutations in *FBN1*. Based on clinical observations, an increased susceptibility to difficulties with sustained visual attention and visuoconstructive skills may also be present in Marfan syndrome. We report a population of patients clinically diagnosed with a connective tissue disorder, including Marfan syndrome, seen in our medical genetics and neurodevelopment clinic who had a high prevalence of attention deficit hyperactivity disorder (ADHD). Patients were referred to our clinic for genetic evaluation of various connective tissue disorders, including Marfan syndrome, between 2009-2014. Six out of 20 or 30% of patients with a clinical diagnosis of Marfan syndrome had comorbid ADHD with 3 out of 12 or 25% of patients with a molecular confirmation of Marfan syndrome having a clinical diagnosis of ADHD. In addition, 10 out of 32 or 34% of patients with a clinical diagnosis of a connective tissue disorder were also clinically diagnosed with ADHD. Our clinic observations suggest a possible link between connective tissue disorders, including Marfan syndrome, and an increased susceptibility to ADHD as has been previously reported in the literature. This possible link should be further investigated with increased patient populations.

2757M

A novel mutation in *OPN1MW* in a brazilian patient with x-linked retinal cone dystrophy type 5. *A.CV. Castro[1], A.T. Rassi[1], A.AN. Rocha[1], L.FOB. Chaves[1], L.SM. Mendonca[1], T. Oliveira[1], M.P. Avila[1], I.MM. Silva[1], J. Chiang[2], L.AR. Gabriel[1].* 1) Ophthalmic Genetics Department, Federal University of Goias, UFG , Goiania, Goiania, Brazil; 2) Molecular Diagnostic Laboratory, Casey Eye Institute, Portland, Oregon, USA.

Purpose: To report a novel mutation in the OPN1MW gene in a patient with x-linked retinal cone dystrophy type 5 presenting severe photophobia and tritanopia. Methods: Eye examination was performed. Best corrected visual acuity (BCVA) was done (Snellen chart). Color vision was measured with Farnsworth D-15 Dichromatous Color Blindness Test. Retinography, (TRC50DX Topcon) and optical coherence tomography-OCT (Spectralis Heidelberg Engineering Inc.) were done. Full-field electroretinogram (FF-ERG) was performed (Roland Consult RETIport Science 6.12.5). Direct testing for mutations in the GUCA1A gene was performed by PCR amplification and bidirectional DNA sequencing of all coding exons and exon/intron boundaries in a CLIA certified laboratory. Additionally, whole exome sequencing (WES) was performed. The results were analyzed and compared to their NCBI reference sequences. The bioinformatic tool Mutation Taster was used to predict the effect of the mutation on the protein. Protein domains were identified with the bioinformatic tool SMART. Results: Examination revealed a 65-year-old brazilian male with severe photophobia, obligating the patient to use dark lenses sunglasses all the time and pilocarpine eyedrops in order to obtain some relief. BCVA was 20/20 at the right eye (OD) and 20/40 at the left eye (OS). The Farnsworth D15 test showed trinatopia. Retinography revealed moderate arteriolar vasoconstriction, severe venular vasoconstriction, retinal pigment epithelium atrophy, peridiscal atrophy, and retinal temporal atrophy with dark pigment clumping. The macula was relatively spared. The OCT showed a decreased foveal thickness on both eyes: 174μm in OD and 205μm in OS. The FF-ERG showed an important decrease in cones response (30%) and, in contrast, a discreet decrease in rods response (90%) on both eyes. The GUCA1A sequencing didn't reveal any mutations. WES identified the novel missense mutation c.538G>T:p.Ala180Ser on the OPN1MW gene. This novel mutation was predicted as disease causing by Mutation Taster. It affects the 4th transmembrane helix region of the OPN1MW protein according to SMART. Conclusions: Herein we've described a novel mutation in OPN1MW (c.538G>T:p.Ala180Ser) in a patient with severe photophobia and tritanopia.

2758T

Cerebrofaciothoracic dysplasia: a case report with molecular search for *TMCO1* mutation. *J. Rivera[1], P. Campeau[2], E.L. Mellín-Sánchez[1], F.J. Martínez-Macías[1], L. Bobadilla-Morales[1,3], A. Corona-Rivera[1,3], J.R. Corona-Rivera[1,3].* 1) Centro de Registro e Investigación sobre Anomalías Congénitas (CRIAC), Servicio de Genética y Unidad de Citogenética, Hospital Civil de Guadalajara "Dr. Juan I Menchaca", Guadalajara Jalisco, México, Guadalajara Jalisco, México; 2) Sainte-Justine University Hospital Research Center, University of Montreal, Canada; 3) Instituto de Genética Humana "Dr. Enrique Corona Rivera", Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara, México.

Introduction. The cerebrofaciothoracic dysplasia (CFTD), also known as Pascual Castroviejo syndrome type 1 (MIM 213980) belongs to the group of bone dysplasias with spondylocostal/thoracic dysostosis. The CFTD is characterized by macrocephaly, hypoplasia of corpus callosum, elongated septum pellucidum, flat facial profile, hypertelorism, cleft lip and palate, low-set ears, costovertebral abnormalities and intellectual disability. The CFTD is caused by a homozygous nonsense mutation in the human transmembrane and coiled-coil domains protein 1 (*TMCO1*), mapped on 1q23.3-q24.1. We report a Mexican infant with CFTD with molecular analysis negative for *TMCO1* mutation and review the literature regarding to the CFTD. **Clinical Report.** The *propositus* was the product of the first uncomplicated pregnancy from non-consanguineous and healthy parents. Family data included two maternal uncles with hydrocephalus, and two cousins with cleft lip -one maternal and one paternal, respectively. Prenatal ultrasounds reported macrocephaly and suspected of skeletal dysplasia. Delivery was carried out via cesarean in the 36th week of gestation. Apgar scores were 7 and 8 at 1 and 5 min, respectively. The birth weight was 2280 g (10th centile), the length was 44 cm (< 3rd centile), and the infant had an occipitofrontal circumference of 38 cm (> 97th centile). Clinical examination showed broad forehead, flat nasal bridge, posteriorly rotated ears, flat facial profile, short nose, synophrys, short neck, single umbilical artery, and bilateral single palmar crease, clinodactyly, and single crease on both fifth fingers. Echocardiogram revealed a patent ductus arteriosus. On radiographs showed platyspondyly, wedged vertebrae, and multiple thoracic hemivertebras. MRI of the brain exhibited cerebellar atrophy, posterior fossa cyst, corpus callosum hypoplasia, and cavum septum vergae. On fluoroscopy examination with liquid barium, a nasopharyngeal reflux was reported. Karyotype with G-bands was normal 46,XY, at a 550 band-level resolution. The molecular analysis was negative for a *TMCO1* mutation. **Conclusions.** Clinical and radiographics features in our patient were in concordance with the main anomalies reported in the 34 previous reported cases with CFTD up to date. Of them, 16/17 molecularly studied have mutations in the gene *TMCO1*. Since mutation was not found in our patient, further strategies are required in the CFTD for its complete molecular elucidation.

2759S

CFTR: p.I1023R is a rare but recurrent disease-causing mutation found in Chinese patients with Cystic Fibrosis. *B.H.Y. Chung[1,2], K.C. Leung[2], C.C.Y. Mak[1], C.S.K. Chow[1], J.D. Ying[1], Y.W.Y. Chu[1], W.L. Yang[1], Y.L. Lau[1], K.Y.K. Chan[2,3], S.L. Lee[1].* 1) Department of Paediatrics and Adolescent Medicine, Queen Mary Hospital, Li Ka Shing Faculty of Medicine, The University of Hong Kong, HKSAR, China; 2) Department of Obstetrics & Gynaecology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, HKSAR, China; 3) Prenatal Diagnostic and Counseling Department, Tsan Yuk Hospital, HKSAR, China.

Cystic fibrosis (CF [MIM 219700]) is a common autosomal recessive disorder in Caucasians, yet less than 20 Chinese patients with molecularly confirmed CF have been reported in both English and Chinese Biomedical journals. Our department is the only center that offers sweat test in Hong Kong, a city of 7.3 million in southern China. We estimate that the incidence of CF is around 1 in 300,000 live births in Chinese population in Hong Kong. Here, we report 6 unrelated Chinese patients with CF with detailed description of their clinical manifestations. With a molecular approach combining NGS, Sanger sequencing and MLPA, we screened the entire protein coding regions of CFTR and selected deep intronic sites for known disease-causing mutations, deletions and duplications. Molecular diagnosis of 4 patients was confirmed, and interestingly 3 patients were found to carry a missense mutation, p.I1023R (CFTR: NM_000492.3: c.T3068G), reported previously in 2 Taiwanese siblings with CF. All 3 patients inherited the mutation from either parent. Bioinformatics tools predict I1023R is possibly damaging to the protein. The clinical manifestations of patients carrying this recurrent mutation are typical CF features including Pseudomonas aeuroginosa pneumonia, bronchiectasis and meconium ileus. From our search in literature and CFTR mutation database, p.I1023R has not been reported in CF patients of other ethnicities. Furthermore, it is not found in our own exome database (~200 exomes), ESP6500 and 1000GP, suggesting a low allele frequency. Linkage analysis is currently underway to study whether I1023R is a founder mutation in Han Chinese. In conclusion, we propose that I1023R is a rare but recurrent disease-causing CFTR mutation important in Chinese CF patients. p.I1023R is not included in the CFTR panel recommended by the ACMG. This finding has implications in the design of mutation panels/ analysis of NGS for molecular diagnosis of CF in Chinese population.

2760M

Identification of a novel *ERCC8* mutation in a 10 year old brazilian female with Cockayne Syndrome type 1. *L.FOB. Chaves[1], A.CV. Castro[1], A.AN. Rocha[1], J. Chiang[2], A.T. Rassi[1], T. Oliveira[1], M.P. Avila[1], L.AR. Gabriel[1].* 1) Ophthalmic Genetics Department, Federal University of Goias, UFG, Goiania, Goias, Brazil; 2) Molecular Diagnostic Laboratory, Casey Eye Institute, Portland, Oregon, USA.

Purpose: To present a case report of a female patient with Cockayne Syndrome type I presenting two mutations on the ERCC8 gene, one being an already known mutation and the other one a novel missense mutation. Methods: Clinical examination was performed. Direct testing for mutations in the ERCC6 and ERCC8 gene was performed by PCR amplification and bidirectional DNA sequencing of all coding exons and exon/intron boundaries in a CLIA certified laboratory. The results were analyzed and compared to the NCBI reference sequences NM_000124.3 (ERCC6) and NM_000082.3 (ERCC8). The bioinformatic tools PolyPhen-2 and Mutation Taster were used to predict the effect of the novel mutation on the resulting protein. Protein domain identification was done with the bioinformatic tool SMART. Results: Clinical examination revealed a female patient born with 38 weeks of gestation by elective cesarean delivery without any maternal or obstetric complications. At birth she presented APGAR (1st and 5th minutes) 8 and 9 respectively, weight of 2850g, height of 48cm, head circumference of 33.5 cm and chest girth of 32cm. She was developing properly until 3 months when started failure of growth associated with developmental deterioration, microcephaly with a final head circumference of 45 cm (08 years old), mental retardation, ataxia, tremors, cutaneous photosensitivity, thin and dry hair. The ocular findings were sunken eyeballs, strabismus (exotropia), pigmentary retinopathy and a persistent miosis. The ERCC8 sequencing revealed a reported mutation IVS7-1G>A and a novel mutation c.1055G>A:p.Gly352-Asp. This novel mutation was predicted as probably damaging by PolyPhen-2 and disease causing by Mutation Taster. The novel mutation affected the 7th WD-40 repeat domain of the ERCC8 protein. Conclusions: Herein, we reported a patient with typical Cockayne Syndrome type I presenting a novel mutation on ERCC8 affecting the 7th WD-40 repeat domain of the ERCC8 protein. Additionally this is the first mutation in the 7th WD-40 repeat domain causing Cockayne Syndrome, since the only mutation on this domain was reported in a patient with UV-sensitive syndrome.

2761T

Identification of a novel *RP1L1* mutation in a 38-year-old brazilian female with retinal cone dystrophy. *A.AN. Rocha[1], A.CV. Jordao[1], I.MM. Silva[1], A.CV. Castro[1], L. Lando[1], J. Chiang[2], M.P. Avila[1], L.AR. Gabriel[1].* 1) Opthalmic Genetics Department, Federal University of Goias, UFG, Goiania, Goias, Brazil; 2) Molecular Diagnostic Laboratory, Casey Eye Institute, Portland, Oregon, USA.

Purpose: To present a case report of a female patient with cone dystrophy presenting a novel missense mutation in the RP1L1 gene. Methods: Eye examination was performed. Best-corrected visual acuity (BCVA) was done using the Snellen chart. Color test was done with Farnsworth D-15 dichromatous color blindness test, visual field was tested using Goldmann semiautomatic kinetic perimetry (Octopus 900, Haag-Streit), and retinography and fluorescein retinal angiography was done with TRC50DX Topcon. Whole exome sequencing (WES) was performed. All the genes accountable for retinal cone dystrophies were analysed and compared to their NCBI reference sequences. The bioinformatic tools PolyPhen-2 and SIFT were used to predict the effect of the novel mutations on the resulting proteins. Protein domain identification was done with the bioinformatic tool SMART. Results: Clinical examination revealed a 38-year-old female patient born from non-consaguineous parents, with slowly progressive decline of her BCVA on both eyes iniciated at 12 years old, having a eye fundus with temporal pallor of the optic disc and diffuse and faint yellow dots between the retinal vessels arcades on both eyes. The fluorescein angiography highlighted these eye fundus alterations, with discreet mottling of the posterior pole. Farnsworth D-15 test demonstrated color blindness, specially deuteranomaly. Goldmann semiautomatic kinetic perimetry demonstrated central scotoma with normal peripheral fields. WES revealed the novel heterozygous c.4004G>T:p.Gly1-335Val RP1L1 mutation. This novel mutation was predicted as probably damaging by PolyPhen-2 and damaging by SIFT. It caused a substitution of a glycine to a valine in an unknown region of the protein between a low complexity region and the first coiled coil region of the protein (two amino acids before the beginning of the next coiled coil domain). Conclusions: Herein we describe a novel mutation on RP1L1 along with a novel phenotype, once this gene was up to now exclusively related to either dominant retinitis pigmentosa, or occult macular dystrophy.

2762S

CRANIOFACIAL SYNDROMES AND GENETIC VARIABILITY IN A PEDIATRIC HOSPITAL IN MEXICO. *W. San Martin-Brieke[1], J.M. Aparicio-Rodriguez[2, 3], L. Vázquez de Lara[3], A. Peral García[3], S. Chatelain-Mercado[4].* 1) Maxilofacial Surgery; 2) Genetics, Hosp para el Nino Poblano, Puebla; 3) Estomatology, Benemérita Universidad Autónoma de Puebla; 4) Biothecnology, Universidad Metropolitana de México.

It has been observed that a 60 to 70% of congenital malformations, there is not a definite cause. Within the causes that are known to exist: alterations chromosome 3-5%, 20% genetic mutations, environmental agents, radiation 1%, 2-3-% infections, metabolic disorders maternal 1-3%, drugs and agents chemical 2-3% of the remainder the cause is unknown. The understanding of many of the development and growth disorders affecting craniofacial structures Pinto (1979); Alfaro et al. (1994); Gorlin (1985); Witkop (1975); Slavkin (1996) is achieved through knowledge of embryology, genetics and histology of these structures. The true etiologic factors present in several alterations in the development and growth of the oral cavity, maxillary and various soft tissues. Witkop in 1975 and Gorlin in 1983 stressed that in certain craniofacial diseases; genetic and hereditary factors may be decisive or just contribute to the emergence of a specific disease. Most of craniofacial malformations are of unknown etiology, and, as a result, the classification is mainly based on features of form and structure. There are many types of anomalies relating to the shape, number and structure and it has a hereditary origin. The nature of the abnormalities depends mainly by genetic factors. The frequency which these problems may arise depends on the form of inheritance and other laws of probability. Some anomalies may occur as the only apparent hereditary alteration. Others are presented as part of a much more complex genetic problem. Specifically congenital agenesis of tissues that may also be the only problem of hereditary origin. However, Agenesis and malformations can be part of a syndrome and be related to alterations in other ectodermal tissues such as hair, skin and mucus membranes, assessed as syndromes in this study: Moebius, Goldenhar, Cockayne, Opitz G, Cornelia de Lange, Criduchat, Patau, Edwards, Down, Klinefelter and Turner. In Mexico was established in 1978 the "registration and surveillance epidemiological of the malformations congenital external" (RYVEMCE), generating preventive information programmes targeting the population at risk. For their study, malformations were divided as genetic or congenital abnormalities observed in this study into five groups according to the structural defect of the same as well as the genetic, congenital, and multifactorial cause.

2763M

Exploring somatic mosaicism in uterovaginal aplasia. *X. Bonilla[1], P. Makrythanasis[1], F.A Santoni[1], A. Pellet[3], K. Rall[4], S. Eisenbeis[4], M. Guipponi[2], C. Gherig[2], C. Rosenberg[5], S. Lyonnet[3], S.E. Antonarakis[1,6].* 1) Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) Department of Genetic Medicine and Laboratories, University Hospitals of Geneva, Switzerland; 3) Department of Genetics and Embryology of Congenital Malformations, Hôpital Necker, Paris France; 4) Department of Obstetrics and Gynecology, Tübingen University Hospital, Tübingen, Germany; 5) Department of Genetics and Evolutionary Biology, Human Genome Research Center, Sao Paulo University, Brazil; 6) iGE3, Institute of Genetics and Genomics of Geneva, Switzerland.

The molecular cause of congenital uterovaginal aplasia, or Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome in unknown. The existence of families with more than one affected individual supports the possibility of a genetic cause; however, the majority of cases are sporadic. The lack of success in identifying the genetic origin of MRKH syndrome so far, points to a complex etiology. This could be due to a reduced penetrance or sex-specific expression of germline mutations or to somatic mosaicism. The objective of our study is to identify candidate genes for the MRKH syndrome through the investigation of genetic somatic differences between monozygotic twins discordant for this disease. We performed whole genome high-throughput sequencing of blood DNA of two pairs of monozygotic twins discordant for MRKH syndrome to a coverage of 35x for >90% of the genome. Variants were called with samtools and Pindel. About 4.5 million variants were identified per individual, 95% of them have been previously seen in the human population and are found in dbSNP. The identified variants were queried with Variant Master (Santoni et al., 2014) in order to detect differences among the individuals from both twinships. After variant filtering, the number of private variants per individual ranged between 34 and 162. After Sanger validation, one of the variants was a true difference between the individuals of a twinship. The SNV causes a non-synonymous change, p.P84L, in the first exon of PI15, a protease inhibitor gene from the CAP family. Validation of this candidate gene in a sample set of matching blood and uterine tissue remnants of patients with MRKH is ongoing in order to assess the role of early somatic mosaicism limited to Müllerian structures in the etiology of this disease. Another validation set is a cohort of more than 40 MRKH individuals' blood DNA. Using targeted high-depth high-throughput sequencing (coverage >1,500x) we will search for low levels of mosaicism in blood, down to 1%. SANTONI, F. A., MAKRYTHANASIS, P., NIKOLAEV, S., GUIPPONI, M., ROBYR, D., BOTTANI, A. & ANTONARAKIS, S. E. 2014. Simultaneous identification and prioritization of variants in familial, de novo, and somatic genetic disorders with VariantMaster. Genome Res.

2764T

An extended Turkish family with FBN1 mutation and variable clinical phenotype. *S. Temel[1,2], O. Tetik[3], O. Rodoplu[4], M. Melek[5], L. Van Laer[6], B. Loeys[6].* 1) Dept Histol & Embryol, Fac Medicine, Univ Uludag, Bursa, Turkey; 2) Dept Histol & Embryol, Fac Medicine, Near East University, Lefkosia, North Cyprus; 3) Dep Cardiothorasic Surgery, Fac Medicine, Celal Bayar University, Manisa,Turkey; 4) Bursa Yuksek Ihtisas Hospital, Dept Cardiothorasic Surgery, Bursa, Turkey; 5) Bursa Yuksek Ihtisas Hospital, Dept Cardiology Surgery, Bursa, Turkey; 6) Center for Medical Genetics, Antwerp University Hospital and University of Antwerp, Antwerp, Belgium.

Marfan syndrome is a multi-systemic autosomal dominant condition caused by mutations in the fibrillin-1 gene (FBN1). Here we report on an extended family presenting with skeletal, ocular and cardiovascular clinical features. The 37 year old male propositus was referred to our clinic due to chest pain, dyspnea and shortness of breath. The main findings of the propositus included severe myopia and mild marfanoid facial features. Echocardiography and computerized tomography showed chronic type I aortic dissection, ascending aorta aneurysm (90 mm) and severe aortic valve insuffiency. His family history was significant for sudden death of his father and his older sister, and similar phenotypes for three generation in the family. The patient's 41 year old brother's main findings included chest abnormalities (pectus carinatum), myopia-astigmatism, and lens subluxation. Echocardiography and computerized tomography showed also chronic type I aortic dissection, ascending aortic aneurysm (70 mm) and severe aortic valve insuffiency. The two brothers underwent immediate Bentall Procedure. FBN1 sequencing of the proband, brother and available affected family members revealed the presence of a pathogenic variant (c.7828G>C, p.Glu2610Gln). None of the unaffected family members did reveal this variant. Because of its complete segregation and due to the fact that Glu2610 residue is part of the conserved DINE motif found at the beginning of each cbEGF domain of FBN1, it is considered a pathogenic variant. This variant has been reported before in the FBN1-UMD database by personal communication but so far no phenotype has been reported with this variant. The presented family demonstrated variable clinical expression of the Marfan phenotype.

2765S

Combination Biotin Responsive Encephalopathy and Hemiplegic Migraine Disorder presenting as Autism and episodic limb dysfunction/ seizures in a 10 year old girl. *P. Benke.* Genetic Division, Joe Dimaggio Children's Hosp, Hollywood, FL.

A 10 year old girl presented to Genetic Clinic with findings of autism. She had all the features of autism, including developmental delay/late language and a paucity of verbal expression, poor understanding, poor social skills, poor eye contact and oppositional behaviors. Some improvement in her symptoms had been accomplished with a low gluten, dairy free diet, but she was still far behind an age appropriate level. Physical examination did not yield any clues, but when asked whether or not her hair and nails grow, her mother said that they did not. The mother was asked to observe hair and nail growth for 3 weeks; when it was observed that they did not grow, she was instructed to start 10 mg/day biotin, which resulted in good growth of both. In addition, there was immediate improvement in school performance. The child began to talk better and her memory dramatically improved. She was thought to have had partial seizures from age 2-3 yrs, with changes in her EEG. Antiepileptic medicines and the change in diet decreased, but did not completely eliminate the episodes, during which one or another limb would lose function, and she developed a concurrent inability to talk or communicate. She would recover after a few hours time. She improved and was free of the latter symptoms when she vacationed in Colorado, and took Diamox. She was then treated for presumptive hemiplegic migraine with Diamox, and dramatically improved. Her EEG normalized, and she began to demonstrate normal, even advanced artistic skills. Her school work improved further, with normal reading and math skills, and a gain in IQ scores. An Exon genome scan (GeneDx) demonstrated a new mutation (p.T364M, c.1091 C>T) in the ATP1A2 associated Hemiplegic Migraine gene, not found in either parent, but did not show a mutation in a biotin dependent gene process. This study shows that 2 genetically unrelated disorders together can lead to a diagnosis of autism, and symptoms can improve dramatically when specific gene appropriate measures are employed.

2766M

New cases of patients with developmental delay and incidental findings of chromothripsis. *O. Caluseriu[1], J. Chernos[2], N.J. Leonard[1], C.M. Klotz[1], B. Argiropoulos[2].* 1) Dept Med Gen, University Hospital, Edmonton, AB, Canada; 2) Dept Med Gen, Alberta Children's Hospital, Calgary, AB, Canada.

Chromothripsis refers to a phenomenon of a local apparent shattering event of chromosomes and subsequent reassembly, observed in 2-3% of cancers (Stephens et al, 2011). It has also been identified in the germline of a subset of individuals with congenital developmental disorders (Kloosterman et al, 2011). While the understanding of the complex chromosomal rearrangement (CCR) resulting from chromothripsis has started to be unveiled (Kloosterman et al, 2012), the relation to developmental problems and long term consequences for the patient's health are unknown. We are presenting two paediatric patients with incidental findings of chromothripsis, their clinical phenotypes and potential correlations with their genotype known to date. PATIENT 1 was evaluated at 7 years of age for short stature (H -3SD), relative macrocephaly (OFC 75-90%), non-dysmorphic features, mild global developmental delay, hypothyroidism, breath holding spells with normal EEG, and a query paracentric inversion 13q21-q31. A microarray showed a CCR within chromosome 13q involving 7 copy number imbalances. Five duplications (size 216 kb-5.2 Mb) and two deletions (size 496 kb-5.2 Mb) were identified, and may suggest potential predisposition to cancer and neurobehavioral problems. PATIENT 2 presented at 21 months of age for query Neurofibromatosis type 1 (NF1). No other NF1 signs were seen aside from 15 cafe-au-lait macules, but there were subtle dysmorphic features, normal growth parameters, significant global developmental delay, and a diagnosis of autism. A microarray identified a 1.15 Mb interstitial deletion that partially overlapped the critical region for Potocki-Schaffer syndrome, and a full karyotype revealed at least 8 apparently copy-neutral breaks involving chromosomes 9, 10, and 11. Pair-end next generation sequencing (NGS) studies were undertaken and results pending at this time will better characterize the rearrangement and could provide prognostic information. Recent technologies (NGS) contribute to the understanding of the driving role of chromothripsis, and may well provide information for the management and counseling of a subset of patients with developmental delays.

2767T

PEDIATRIC PATIENTS CLINICAL EVOLUTION WITH POSTOPERATIVE NASAL RETAINER FOR BILATERAL LIP AND CLEFT PALATE (BLCP). *J. Marin-Melo[1], J.M. Aparicio-Rodriguez[2,3], S. Ochoa-Caceres[1].* 1) Estomatology; 2) Genetics, Hospital para el Nino Poblano, Puebla, Puebla; 3) estomatology, Benemérita Universidad Autónoma de Puebla, México.

Treatment of patients with lip and cleft palate requires multidisciplinary management, focused on providing optimal structural, functional, and aesthetic conditions so the patient can develop and achieve a better quality of life. The goal of the estomatologic treatment is to repair altered structures to reach a growth potential and suitable development improving results in surgical times, minimizing adverse effects. One of the treatments for nasal deformity in patients with lip and cleft palate is the use of forming, which are attachments that work the nasal deformity using functional orthopedics; these help avoiding to collapse and relapse the nasal ala, have more permeable airways and eliminate surgical adhesions by secretions or contraction of healing. The objective of this study is to present the case of a female patient of 9-year-old with bilateral lip and cleft palate presenting nasal disorders such as; short nasal projection. It indicated the use of the nasal retainer as postsurgical orthopedic device to improve the structural, functional and esthetic soft tissue deficiencies.

2768S

Clinical Aspects associated with Syndromic forms of Orofacial Clefts in Colombia. *I. Briceno[1,3], jc. martinez[1], a. collins[2], I. arias[3].* 1) Bioscience, U Niversidad de La Sabana, Chia, Colombia; 2) University of Southampton. Southampton . UK; 3) universidad Javeriana. Bogota. Colombia.

Abstract Objectives: 1) to present descriptive epidemiology of Orofacial Clefts (OFCs) and;) to determine the association of syndromic forms with antenatal high-risk conditions, preterm birth, and comorbidities, in a group of patients from Operation Smile Colombia. Design: A cross-sectional study was conducted. Frequencies of cleft type, associated congenital anomalies, syndromic, non-syndromic and multiple malformation forms, and distribution of OFCs according to gender and affected-side were determined. Odds ratios were calculated as measures of association between syndromic forms and antenatal high risk conditions, preterm birth and comorbidities. Setting: Operation Smile Colombia. Participants: A total of three hundred and eleven patients with OFCs treated in a 12-13 month period. Results: The most frequent type of OFC was cleft lip and palate (CLP). CLP was more frequent in males, whereas cleft palate (CP) occurred more often in females. The most common cases occurred as non-syndromic forms. Aarskog-Scott syndrome (ASS) showed the highest frequency. Hypertensive disorders in pregnancy, developmental dysplasia of the hip, central nervous diseases and respiratory failure showed significant statistical associations ($p < 0.05$) with syndromic forms. Conclusions: These data provide an epidemiological reference for OFCs in Colombia. Novel associations between syndromic forms and clinical variables are determined. In order investigate causality relationships between these variables further studies must be carried out.

2769M

Distribution of the AKT1 p.Glu17Lys mutation in a patient with Proteus syndrome. *M.j. Lindhurst[1], M. Doucet[2], H.M. Bloomhardt[1], M.R. Yourick[1], K. Moroz[2], L.G. Biesecker[1].* 1) MGMGB, NHGRI/NIH, Bethesda, MD; 2) Department of Pathology and Laboratory Medicine, Tulane University School of Medicine, New Orleans, LA.

Proteus syndrome (PS) is a rare disorder that occurs sporadically and is characterized by progressive, disproportionate, segmental overgrowth that can affect any tissue in the body. It is caused by a post-zygotic activating mutation, c.49G>A, p.Glu17Lys in *AKT1*, resulting in individuals who have both mutation positive and mutation negative cells. Because the mutation occurs in a somatic cell, each individual has a unique constellation of manifestations making this disorder extremely heterogeneous. Recently, we had the opportunity to assess the mutation level in multiple tissues due to the unfortunate death of a 21-year old woman with PS. We sampled 25 tissues that were visibly affected and 13 tissues that were apparently normal. In most cases, the specimens were subdivided and multiple DNA extractions were performed. The *AKT1* p.Glu17Lys mutation was measured using a custom PCR-based RFLP assay and the levels were averaged for each specimen. Of the 13 unaffected samples, six had detectable levels of the PS mutation (2 - 29%) although four of the mutation positive samples had no significant histopathologic abnormalities. Of the seven unaffected samples that had no mutation, only three were histologically normal. The mutation level in the affected samples ranged from 3 to 35% and all but the kidneys with mutation levels of 15 and 19% had abnormal histopathology. Interestingly, the highest mutation level was in a bone sample from the second digit of the right foot while the lowest levels were found in the soft tissue surrounding that bone, as well as in an omental fat nodule. The ovaries, vagina, breast, as well as several bones and pieces of skin were mutation positive, while the spleen, liver and lungs were mutation negative. Samples from a uterine leiomyoma, splenic hemangioma, intracranial osteoma and an intracranial meningioma, were all mutation positive. In addition, a papillary carcinoma was found in the right lobe of the thyroid that was positive for both the *AKT1* PS and *BRAF* p.Val600GLu mutations whereas uninvolved thyroid was mutation negative. This case has given us a rare opportunity to correlate mutation burden with histopathology in many tissues typically unfeasible to sample.

2770T

Utility Of Genetic Testing In Patients With Suspected Fetal Alcohol Spectrum Disorder. *S.S. Jamuar[1,2], J.M. Stoler[2].* 1) Dept of Paediatric Medicine, KK Women's and Children's Hospital, Singapore, Singapore; 2) Division of Genetics, Boston Children's Hospital, Boston, MA.

BACKGROUND: Fetal alcohol spectrum disorder (FASD) is one of the most common developmental disorders with an estimated prevalence of 1 to 2% in the United States. It is a clinical diagnosis and is based on the presence of growth retardation, microcephaly, neurologic dysfunction and distinct craniofacial features. Genetic testing is not routinely offered. However, in a recent study, 9% of individuals referred to the Genetics clinic for evaluation for FASD were given an alternate diagnosis based on examination and testing, including chromosomal microarray analysis (CMA) (Douzgou et al. ADC 2012). We aim to study the utility of genetic evaluation and testing in patients with suspected FASD referred to our Genetics clinic. METHOD: We performed a retrospective chart review of all patients referred for evaluation for suspected FASD to the Genetics clinic at Boston Children's Hospital between the periods January 2006 to January 2013. Records of all patients were reviewed to obtain the following information: demographic data, referral characteristics, birth history, history of prenatal alcohol exposure, family history, medical history, examination findings and investigations, including genetic testing (karyotype, Fragile X and CMA). RESULTS: Thirty-four patients were evaluated for possible FASD during the study period, at a median age of 4 years (range 0 to 17 years). Definite prenatal exposure was documented in 23 (67%) patients. Polysubstance abuse was noted in 16 (47%) patients. Speech delay and behavioral problems were common complaints seen in >50% of the patients. Facial dysmorphism was noted in 18 (53%) patients. Microcephaly was present in 14 (41%) patients. Three patients did not fulfill clinical criteria for FASD. Of the remaining patients, 15, 13 and 21 patients underwent karyotype, Fragile X and CMA, respectively. Karyotype and Fragile X were normal for all individuals tested. On the other hand, CMA detected 16 copy number variants (CNVs) in 12 patients (average of 1.3 CNV per patient). Pathogenic and potentially pathogenic CNVs were detected in 4 (2q37del, 15q26del, 22q11.22dup and 4q31.21del syndromes) and 2 (2p25.3dup and AOH over Xp11.22q21.1) patients, respectively, giving a total diagnostic yield of 28%. CONCLUSION: Genetic testing, especially CMA, should be routinely offered to patients referred for evaluation of FASD as a significant proportion (>25%) have a clinically significant CNV even when they fulfill diagnostic criteria for FASD.

2771S

Is a computer-based Facial Dysmorphology Novel Analysis ready for the clinic? *L. Basel-Vanagaite[1,2,3], L. Wolf[2,3].* 1) Schneider Children's Medical Center of Israel, Rabin Medical Center, and Felsenstein Medical Research Center, Petah Tikva, Israel,; 2) Tel Aviv University, Tel Aviv, Israel; 3) FDNA Ltd., Herzlyia, Israel.

Introduction: Previously, we were able to demonstrate that the facial dysmorphology novel analysis technology was successful in recognizing the dysmorphology associated with selected syndromes by processing 2D facial images. The computer-generated analyses were able to produce results comparable with those of human experts. In this study we investigated the performance of the system by analyzing a random set of images of dysmorphic individuals affected with a random variety of syndromes. Methods: The images were submitted by more than a hundred medical geneticists using the Face2Gene mobile application. For quality assurance, 350 images were chosen randomly to be reviewed (without any personal information included) independently by a single human geneticist experienced in dysmorphology (LBV) and compared to matches found by the facial recognition software. Images of individuals being affected with rare chromosomal imbalances were excluded. A match was considered positive where the syndrome determined by the geneticist was also listed among the ten best matches suggested by software (either by using FDNA's facial gestalt analysis component or FDNA's proprietary textual search engine, based on a list of HPO terms representing facial features automatically detected by the system and additional features added by the users). Results: In 52/350 cases (15%), the human expert was able to clearly recognize the presence of a specific genetic syndrome, based on gestalt only. In 44 of these cases (85%), there was a positive match between the system and the human expert, of which 38 cases were suggested by the gestalt analysis and 13 cases were suggested by the feature-based search engine (7 cases were suggested by both modalities). Only 8/52 (15%) cases were recognized by the human expert, but not by the system. It is unknown in how many cases the system recognized the "true" syndrome when the human expert did not, since the majority of the cases are submitted without molecular data, other than 2 cases in which a molecular confirmation was indicated, and the system was able to suggest the correct syndrome, while the expert could not. Conclusions: We conclude that computer-based facial recognition system can successfully assist medical professionals in the research of genetic syndromes characterized by dysmorphic features. Possible future applications may include usage of facial analysis software to complement molecular studies, such as whole exome sequencing.

2772M

Phenotypic characterization of Microtia in Bogota, Colombia. *LP. Barragan Osorio[1], M. Garcia[1], I. Zarante[1,2], DV. Luquetti[3].* 1) Instituto de Genetica Humana, Pontificia Universidad Javeriana, Bogota, Colombia; 2) Secretaria Distrital de Salud, Bogotá, Colombia; 3) Seattle Children's Craniofacial Center. University of Washington, Seattle, US.

Introduction: Microtia is a congenital anomaly of the shape and size of the ear with a wide range of severity which is often associated with hearing loss in 80-90%; of cases, and can occurs as an isolated condition. The prevalence in South America is considered to be higher, especially in the Andean population ranging from 8 to 18 per 10.000 newborns. The pathogenesis of this congenital malformation is still unclear, strong evidence supports the environmental role as well as genetic causes of microtia. The aim of this study is to characterize the phenotype and risk factors of a population with isolated microtia in Bogota. Methodology: All potential participants with isolated microtia who were registered in the Bogota congenital malformations surveillance program (BCMSP) database, as well as the patients who assisted to otolaryngology visit or were born in any of following up hospitals, were invited to participated in this study. Each patient was cited to a single clinical valuation during which they were asked to undergo a physical examination, photographs taken and provide blood specimen for later DNA extraction. In addition, the medical history, results of hearing tests and ear computed tomography were collected. Participant characteristics were analyzed using descriptive statistics. Results: 27 patients with isolated microtia were recruited during September 2012 to July 2013. Microtia occurs most frequently in male patients (74%), children of multiparous mothers, 74% had unilateral microtia in whom right ear was mostly affected (80%) and 85% of them had atresia of the external auditory canal regardless the microtia grade. Most patients (11/18) presented hearing loss (moderate, severe or profound) that was not correlated with the type of microtia. All individuals had prenatal ultrasounds with a mean of 4 per patient, however no prenatal diagnosis was made. 63% of the mothers had acute infections during pregnancy. Maternal exposure to physical factors (e.g.: insecticides, organophosphates) was presented in 37% mostly during the first trimester. Discussion: A higher percentage of risk factors exposition was observed as: acute illness and physical risk factors expositions. Also the severity grade and middle ear compromise was higher than previously reported in the literature. These findings could suggest a possible environmental factor association in our population, but is necessary to increase our sample to calculate and extrapolate our results.

2773T

Congenital eye malformations and associated clinical aspects in two Colombian cities (Bogotá- Cali) between 2011-2013. *A.M. Garcia[1], A.I. Sanchez[2], P.M. Hurtado[2,3], I. Zarante[1].* 1) Instituto de Genetica Humana, Pontificia Universidad Javeriana Bogota, Bogota, Colombia; 2) Pontificia Universidad Javeriana Cali; 3) Centro Medico Imbanaco.

Introduction: Congenital malformations of the eye and anexa are known to occur in isolation or as a part of systemic malformation syndromes, and lead to severe visual impairment in newborns. Although less frequent than other birth defects, the global prevalence has been estimated in medical literature in 3-6 per 10,000 births. The causes of more than 70% of all congenital malformations are still unknown, however those for congenital eye malformations (CEM) are known to be defects in genes, which regulate normal eye development (monogenic cause or chromosome abnormality). The most common CEM are Anophtalmia/Microphtalmia, congenital cataract and coloboma, affecting annually ≈2 per 10,000 newborns worldwide. Objective: The aim of this study was to determine the prevalence of CEM and to describe some characteristics associated with the occurrence of these defects in a population from Bogotá and Cali, Colombia between 2011-2013. Materials and Methods: Data regarding CEM from the Latin American Collaborative Study of Congenital Malformations (ECLAMC) was obtained in a period between 2011-2013 from Bogotá and Cali and was analyzed to describe frequencies and clinical aspects. Results: There were 231,035 consecutive births in this period. Fifteen newborns had CEM, for a global prevalence of 6,49 per 100,000 births. The most common CEM were anophthalmia/microphtalmia (2,6*100000 births), Dacryostenosis (1,3 *100000 births), congenital cataract and congenital glaucoma (0,87 *100000 births). Other less common CEM included eyelid defects and corneal opacity. Systemic anomalies coexisting with CEM were seen in 8 newborns (53,3% of all cases of CEM). The most common associated systemic anomalies were auricular/facial defects (87,5%) and limb defects (50%). Conclusions: CEM are structural defects that can be recognizable at birth and are an important cause of visual disability worldwide (4%). We found in our study that prevalence of these defects is lower than that reported in medical literature, however this could be explained through the fact that some CEM appear later in life and are not evident during our examination period (first day of life). We observed specific associations of eye malformations with other birth defects, being limb and craniofacial anomalies the most common associations as literature says. Early diagnosis is an important fact in reducing childhood blindness.

2774S

Phenotypic diversity in patients diagnosed with VACTERL association. *M. Husain[1,2], M. Wencel[2], B. Lemieux[1,2], V. Kimonis[2], B. Solomon[3,4].* 1) UC Irvine, School of Medicine, Irvine, CA; 2) Division of Genetics and Genomic Medicine, Department of Pediatrics, UC Irvine Medical Center, Irvine, CA; 3) National Human Genome Research Institute, NIH, Bethesda, MD; 4) Inova Translational Medicine Institute, Inova Health System, Falls Church, VA.

The combination of vertebral, anal, cardiac, tracheo-esophageal, renal and limb anomalies termed VACTERL association is a clinical descriptor and a diagnosis of exclusion for a specific set of phenotypic manifestations observed to co-occur non-randomly and more frequently than expected by chance. Because VACTERL is clinically heterogeneous with an elusive etiology, we investigated the frequency and variety of these anomalies, other co-occurring manifestations and underlying causes of defects in embryogenesis or genetic mutations that elucidate VACTERL's etiology. Unique cases in our cohort include the first Trisomy 18 male with an associated Dandy-Walker syndrome exhibiting the VACTERL phenotype, another with a 498.59 kb microdeletion in the 16p11.2 region, a 215 kb duplication in the 3p25.2 region with no known clinical phenotypes, and a genotypic female with dysmorphic facies and ambiguous genitalia. These are 4 among the 36 patients described in our study. Phenotypically, the most common clinical feature was vertebral anomalies, ranging from vertebral body/sacral malformations to spinal cord/rib anomalies, reported in 30 (83.3%) patients. 29 (80.6%) had cardiac defects ranging from ASD/VSDs to dextrocardia and atrial septum aneurysms. 24 (66.7%) had tracheo-esophageal anomalies ranging from TEF with esophageal atresia to Sandifer syndrome. 23 (62.9%) had renal anomalies ranging from agenesis/duplex kidneys to multicystic dysplastic kidneys. 21 (58.3%) had anal anomalies ranging from imperforate anus/anal atresia to a variety of GI/GU fistulas and atresias. 19 (52.3%) had limb anomalies encompassing both upper and lower extremities. Lastly, 14 (38.9%) also had urogenital anomalies. Indubitably, VACTERL is more complicated than previously thought. Although we found no evidence of teratogenic exposure of the mother during pregnancy or other contributing factors, our findings are promising nonetheless. The 16p11.2 microdeletions notably identified in patients with mental retardation, autism and learning/speech problems, have also been associated with malformations commonly found in VACTERL. Additionally, the 3p25.2 duplication with no previously known clinical phenotypes can now be scrutinized further, along with other emerging microarray abnormalities, to identify candidate genes responsible for VACTERL association. We hope these findings enlarge upon the current understanding of VACTERL and guide research aimed at exploring its etiologies.

2775M

Brooke-Spiegler syndrome: A rare association of thichoepithelioma, cylindroma and spiradenoma. Report of a familial mexican case. *N.O. DAVALOS[1,4], M.E. SANCHEZ-CASTELLANOS[2], I.M. SALAZAR-DAVALOS[1], M.A. ACEVES-ACEVES[1], J.O. HIGAREDA-GONZALEZ[4], M. HERNANDEZ-TORRES M[4], S.A. RAMIREZ GARCIA[3], M. VAN-DICK[4], I.P. DAVALOS[1], D. GARCIA-CRUZ[1], C. GARCIA-SILVA[2].* 1) INSTITUTO DE GENETICA, CUCS, BIOL.MOL., UNIVERSIDAD DE GUADALAJARA, GUADALAJARA, GUADALAJARA, Mexico; 2) INSTITUTO DERMATOLOGICO DE JALISCO, SSJ, ZAPOPAN, MEXICO; 3) UNIVERSIDAD DE LA SIERRA SUR, OAXACA MEXICO; 4) HOSPITAL GENERAL DE OCCIDENTE, SECRETARIA DE SALUD JALISCO, GUADALAJARA, MEXICO.

INTRODUCTION Brooke H (1892) and Spiegler E (1899), independiently described an ephithelioma adenoides cysticum and skin endotelioma, as distinct entities. Brooke-Spiegler syndrome (BSS, OMIM # 605041) characterized by benign adnexal neoplasia. The predominating tumor such as trichoepithelioma, cylindroma and spiradenoma appear in late childhood and early adolescence. BSS is an autosomal dominant entity. The tumors located on head and neck, and increase throughout life with overlapping clinical features. Clinically trichoepithelioma showed flesh-colored papules specially on nasolabial folds; cylindromatosis lesions presents with multiple erythematous nodules arise on the scalp and is associated with alopecia; and blue-colored-painful lesions on breast and back suggested spiradenomas. The pathogenesis in BSS considered as a defect in the diferentiation of folliculo-sebaceous-apocrine unit. Mutations have been identified in CYLD tummor suppressor gene, mapped to chormosome 16q12-q13. CASE REPORT: Case I: The proposita 9 years-old female. Normal psychomotor development. Between 5-6 years-old displayed flesh-colored papules on nose. The physical examination showed clinical and hystopathological features indicating trichoepithelioma, with keratinizing cystic spaces. Case 2:42 years-old female showed since childhood, multiple flesh-colored-pink papules on her scalp, face and trunk. The lesions were asymptomatic. Three biopsies for the patients were reviewed, the first one on right preauricular region revealed trichoepithelioma, the second one on scalp conclusive cylindroma, and the third on lumbar región conclusive ecrine spiradenoma. Patients are being given this cryotherapy treatment and cosmetic surgery with good results. DISCUSSION: BSS an autosomal dominant disease. Female are commonly affected, with a female to male ratio of 3:1. BSS include mutiple skin appendage tumors such cylindroma (OMIM# 132700), thichoepithelioma (OMIM# 601606) and spiradenoma, that share a common genetic basis. It may be associated with other skin disorders such, basal cell adenomas, basal cell carcinomas. Histopathology studies are necessary to diagnosis. The treatment included excision, dermabrasion, cardon dioxide laser, cryotherapy and in some cases radiotherapy. We present a Mexican family, the mother and daugther affected, presenting the clasical clinical features of the disease. Molecular studies are needed to understand the genetic bases of the disease.

2776T

Disruption of the osteogenic niche signaling in craniosynostosis: primary cilium and prostanoid pathways crosstalks. *W. Lattanzi[1, 2], M. Barba[1], C. Cicione[1], L. Massimi[3], G. Di Taranto[1], P. Frassanito[3], F. Pignotti[3], M. Baranzini[1], M. Caldarelli[3], F. Michetti[1,2], C. Di Rocco[3], G. Tamburrini[3], C. Bernardini[1].* 1) INSTITUTE OF ANATOMY AND CELL BIOLOGY, Università Cattolica S. Cuore, Rome, Rome, Italy; 2) Latium Musculoskeletal Tissue Bank, Rome, Italy; 3) Pediatric Neurosurgery Unit, Università Cattolica S. Cuore, Rome, Rome, Italy.

Nonsyndromic craniosynostosis (NSC) is a highly prevalent craniofacial malformation, with a strong and heterogeneous genetic background. A possible involvement of genes involved in the primary cilium signaling has been proposed in recent studies, proposing brand new candidate genes, with variable genotype/phenotype correlations. Through microarray genome-wide expression profiling and in silico bioinformatics analysis we have shown an altered expression of cilium-associated genes, and of genes involved in prostaglandin metabolism, in calvarial tissues and cells of NSC patients. We have developed and in vitro model for the functional assays aimed at demonstrating the functional role of the selected molecular pathways in the aberrant osteogenic process occurring at the site of premature suture closure. The expression on cilium-related genes (BBS9, GLI3, SMO) and of osteospecific genes (BMP2, RUNX2, OSX, OC) was significantly upregulated in cells isolated from fused sutures (syn-cells) compared to cells isolated from matched patent sutures (control cells), and increased upon 5 days of osteogenic induction. Confocal microscopy showed that: syn-cells produced less primary cilia compared to control cells; BBS9 expression was spread throughout the cytoplasm in syn-cells, while appeared organized in polarized structures surrounding the cilium basal body in control cells. Upon BBS9 silencing in syn-cells, the expression of osteo-specific transcription factors (RUNX2 and OSX), and of SMO (key molecule of the hedgehog pathway), were significantly down-regulated. The osteogenic potential of BBS9-silenced syn-cells decreased and reverted to the physiological behavior observed in controls. Also, syn-cells displayed reduced HPGD expression, leading to reduced catabolic degradation of the osteoinductive prostaglandin E2. Interestingly, PGE2 administration induced differential effects in cells overexpressing BBS9, suggesting an unpredicted crosstalk between the primary cilium and prostanoid metabolism. Our data may provide an original insight into the events occurring in the calvarial osteogenic niche and leading to premature suture closure, which may be useful to: 1. clarify unclear etiopathogenic processes, 2 assist prioritization of new genomic variants, 3 pave the way to further translational outcomes.

2777S

Genotype-phenotype correlation in 12 patients with Oculoauriculovertebral spectrum. *S. Bragagnolo[1], M.E.S Colovati[1], R.S. Guilherme[1], A.G. Dantas[1], C.A. Kim[2], M.I. Melaragno[1], A.B. Perez[1].* 1) Universidade Federal de São Paulo, SÃO PAULO, Brazil; 2) Genetics Unit, Instituto da Criança, Universidade de São Paulo, Brazil.

Oculoauriculovertebral spectrum (OAVS [MIM 164210]) is a clinically heterogeneous condition among the syndromes of first and second branchial archs related to embryonic craniofacial development. Most cases are sporadic with probably multifactorial inheritance, but rare familial cases and several chromosomal abnormalities have been associated with OAVS. The minimal criteria according to Tasse et al. [2005] are unilateral or bilateral asymmetric ear anomalies and hemifacial microsomia. Ocular defects, conductive and/or sensorineural hearing loss, vertebral malformations, and more rarely, cardiac, renal and cerebral malformations with intellectual impairment may appear. The identification of the OAVS genes/genomic position is important to the better understanding of the molecular pathways involved in craniofacial development and clarify clinic heterogeneity of the syndrome. Considering a sample of 62 patients that met Tasse et al. [2005] OAVS criteria (Eur J Med Genet 48:397:411), we identified 12 potentially pathogenic copy number variations (CNV). Genomic arrays identified: 7 deletions (4p16.3p15.33, 4q13.3q21.1, 8q13.3, 10q26.2q26.3, 16p13.3, 22q11.21, Xp22.33) and 6 duplications (2q32.1, 4p16.1.1, 16p13.11, 17q11.2, Xp22.33). We have identified the genes contained within these regions and performed a genotype phenotype correlation in our patients. Even though some deleted and duplicated regions found in this study were relevant to the phenotype OAVS as *BAPX1* (hemifacial microsomia HM [MIM 164210]); *HMX1* (Oculoauricular syndrome OAS [MIM 612109]). ; *EYA1* (Otofaciocervical syndrome1 OFC1 [MIM 166780]); *YPEL1* (YIPPEE-LIKE 1 YPEL1 [MIM 608082]) and *ERK1* (MITOGEN-ACTIVATED PROTEIN KINASE 1; MAPK1 [MIM 176948]), larger deletions and duplications found in our patients included many genes described in OMIM and genomic regions that could relate to the phenotype. These genes and / or candidate regions were investigated and correlated by reverse dysmorphology. Financial support: FAPESP, Brazil (2013/04623-2).

2778M

An autosomal recessive PGAP3 novel mutation was identified in patients with severe intellectual disability, dysmorphism and hyperphosphatasemia from 2 unrelated families using whole-exome sequencing. *V. Adir, A. Shalata, K. Golinker, E. Shahak, M. Mahroum, S. Tzur, Z.U. Borochowitz.* Molecular Genetics, Bnai Zion Medical Ctr, Haifa, Israel.

Whole-exome sequencing (WES) is very efficient method, that can be used as a diagnostic tool for identifying the molecular basis of genetic syndromes that are challenging to diagnose. In this study we report on hereditary intellectual disability and dysmorphism in two large Arab families with consanguineous marriage in northern Israel. WES revealed that the patients from the two unrelated families carry a novel homozygous mutation in the PGAP3 gene ;Chr17 g.37,829,358T>C; c.845 A>G; p.D282G. According to previous studies, the PGAP3 gene encodes a product which is a post-Glycosylphosphatidylinositol (GPI) attachment to protein factor 3, a protein that is involved in GPI-anchored maturation. The aspartic acid in position 282 is highly conserved from worms to humans, it is likely that this change has critical impact on the PGAP3 function, and ability to interact with other proteins and harm the GPI-anchor maturation process. GPI-anchor is a glycolipid structure that is added to the C-terminus of many proteins and is anchor their attachment to the cell membrane. Thus mutation in proteins that are involved in this process can affect signal transduction, neuronal development and can cause many pathologies including a wide spectrum of intellectual disabilities (Chesebro et al). Recently Howard et al described four different mutations in the PGAP3 gene that causes a subtype of hyperphosphatasia with intellectual disabilities. The four patients we analyzed had normal MRI and have characteristic phenotypic features, they suffer from seizures, sever intellectual disabilities, low head circumference and high body weight. They also have coarse face features which includes narrow foreheads, long palpebral fissures, cleft palates, short noses with a wide nasal tip, full cheeks, tent shape wide mouths and large ear lobes. The phenotype of these patients is similar to the one described previously by Howard et al. Since we found the same mutation in four patients, in two unrelated families, we wanted to develop an easy and affordable test that will allow us to screen the relevant population from the same ethnicity. We developed a PCR RFLP test that would enable easy analysis of the c.845 A>G mutation in the PGAP3 gene. The mutation was not found in 40 control chromosomes from the same ethnicity and geographical region. We suspected that there is a blood relation between the this two families and that they are probably connected via a third family from the same area.

2779T

Significant Secondary Findings of Exome Sequencing in Minor Anomalies with Autism Spectrum disorder. *A. Alsadah.* Benioff Children's Hospital, University of California at San Francisco, San Francisco, CA.

Significant Secondary Findings of Exome Sequencing in Minor Anomalies with Autism Spectrum disorder A. Alsadah, J.T.C. Shieh. Division of Medical Genetics, Department of Pediatrics, Institute of Human Genetics, Benioff Children's Hospital, University of California San Francisco, San Francisco, CA The results of exome sequencing in patients with minor anomalies with autism spectrum is unclear. Secondary or incidental findings on exome testing may also yield unexpected results, however findings of clinical consequence may go beyond current ACMG incidental findings recommendations. We present a instructive example of exome sequencing performed as part of developmental delay/autism genetic work up after normal microarray and metabolic testing at the UCSF Genetics Exome Clinic. A dedicated pre-test counseling visit was performed and incidental findings were discussed. The male patient avoided eye contact and had sloped shoulders and vague knee pains. Whole exome sequencing demonstrated compound heterozygosity for p.(P643L) and p.(G741R) variants in the SLC12A3 gene; both of these mutations have been reported in Gitelman syndrome (Cruz et al., 2001, Simon et al., 1996). Gitelman syndrome is an autosomal recessive disorder characterized by hypokalemia, metabolic alkalosis, hypomagnesemia, hypocalciuria, and hyperreninemia and typically presents in adults. As mutations in SLC12A3 affect the thiazide-sensitive Na/Cl cotransporter, we assessed electrolytes and found low potassium and abnormally low urine calcium confirming an early diagnosis of Gitelman syndrome. Our results demonstrate that exome sequencing can detect significant and treatable secondary findings as in this patient. He was started on potassium supplements to correct his electrolytes and prevent possible complications such as cardiac arrhythmias, which could be fatal. Autism is not the expected presentation for this renal disease, however growth retardation and developmental delays have been reported in few patients with Gitelman syndrome (Skalova et al., 2013). We conclude that large-scale clinical sequencing can reveal important treatable conditions, and future studies will reveal potential effects on clinical outcomes. .

2780S

Association of UBE3B and GRIN2B gene variants with autism spectrum disorders and non-syndromic intellectual disability: a case report. *A.I. Sanchez[1], F. Tobar[1], P.M. Hurtado[1,2].* 1) Pontificia Universida Javeriana Cali, Cali, Colombia; 2) Centro Medico Imbanaco.

Introduction: Autism spectrum disorders (ASD) are neurodevelopmental conditions characterized by language abnormalities, impaired social function and repetitive behaviors. Prevalence among 8-year-old (y/o) children is 1% and intellectual disability (ID) occurs in 70% of them. Together affect 3-5% of population. Genetic causes are identified in 10-20% of ASD and 25-50% of ID. Case presentation: This is an 8 y/o male patient who was born by uncomplicated spontaneous vaginal delivery after an uneventful first full term gestation from non-consanguineous parents. Mother was 21 y/o. Birth weight was 3,4 kg. Patient had normal development until 2 y/o when started to have hampered verbal communication. Significant family history was denied. Physical examination revealed a severe language problems but no other abnormal findings. Hearing tests, electroencephalography and MRI of the brain showed no abnormalities. Whole genome sequence analysis on DNA extracted from patient and parents' blood was performed. Heterozygous variants of unknown clinical significance in UBE3B and GRIN2B genes were detected. Discussion: GRIN2B variant was inherited from mother (autosomal dominant). For UBE3B variant (autosomal recessive), a second allele was not detected in parents. Deletions or duplications cannot be ruled out because both genes are located in chromosome 12. UBE3B codifies for an E3 ubiquitin-protein ligase that plays an important role on neurodevelopment. The variant c.421T>G - pD141A was identified in the exon 6, which is known to be a preserved exon. Other UBE3B variants have been associated with blepharophimosis-ptosis-intellectual disability syndrome (BPIDS), but the patient's phenotype is consistent with ASD and non-syndromic ID (NSID). Therefore we propose a new clinical significance for this variant. Additionally, the variant c.3832A>C - pT1278P was identified in the exon 13 of GRIN2B. This gene codifies for glutamic N-methyl-d-aspartate (NMDA) receptor 2B subunit. Other GRIN2B variants have been related to ASD, ID, psychiatric disorders and problems in retina, skin and testis. Both genes are related to glutamic acid signaling pathway which is important in neurodevelopment, therefore we propose an association between these variants and ASD/NSID. Conclusions: Heterozygous variants in GRIN2B and UBE3B were detected in a patient with autism and NSID. Further studies are necessary to uncover the whole phenotypic spectrum of disruption in these genes.

2781M

Congenital Limb Reduction Defects Associated with Maternal Thrombophilia. *L. Ordal[1], J. Keunen[2], N. Martin[3], N. Shehata[4], G.H. Borschel[5,6], H.M. Clarke[5,6], A. Toi[7], C. Shuman[1,8], D. Chitayat[1,2,3,8].* 1) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 2) Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; 3) Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, Toronto, Ontario, Canada; 4) Department of Medicine, Division of Hematology, Mount Sinai Hospital, Toronto, Ontario, Canada; 5) Division of Plastic and Reconstructive Surgery, Hospital for Sick Children; 6) Department of Surgery, University of Toronto, Toronto, Ontario, Canada; 7) Department of Medical Imaging, Mount Sinai Hospital, Toronto, Ontario, Canada; 8) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, Ontario, Canada.

Congenital limb reduction defects (LRDs) comprise an etiologically heterogeneous group of conditions caused by amniotic bands, early chorionic villus sampling, single-gene disorders, chromosome abnormalities, and teratogen exposure. However, a substantial number of cases occur without an obvious cause. Vascular disruption events may be responsible for some of these cases, especially the terminal transverse reduction defects. Some studies have found an association between maternal thrombophilia and congenital LRDs while other studies have not confirmed this association. We further investigated this association through a review of all prenatally identified LRDs at a major tertiary care center in Toronto, Canada over a 12-year period. Research ethics board approval was obtained at all applicable institutions. Using a Chi-square analysis, our results showed a higher prevalence of inherited thrombophilias (specifically, Factor V Leiden, prothrombin G20210A, and heterozygosity for both mutations) among women with pregnancies affected with an LRD when compared to the general population [$\chi^2(3) = 54.63$, $p < .01$]. Our work was strengthened by the inclusion of affected pregnancies regardless of outcome and using strict criteria to try to eliminate including cases of LRDs with an unrelated etiology. The LRDs in our sample were identified at mean age of 19.7 weeks gestation (SD = 2.8). We found an excess of left-sided defects among terminal transverse but not longitudinal reductions; additionally, all cases of thrombophilia occurred in the terminal transverse group. Our results support the hypothesis that maternal thrombophilia is associated with fetal LRDs. This supports the use of appropriate maternal thrombophilia screening (i.e., screening that is not affected by pregnancy, such as molecular genetic testing or antiphospholipid assays) in pregnancies with an identified LRD. Additionally, our findings emphasize the importance of careful examination of the extremities, including digits, during the routine 18-20 week anatomy ultrasound in women with a known thrombophilia or history of thrombosis.

2782T

Whole Exome Sequencing of Moyamoya Disease. *S. Jang[1], S. Lee[1,2], J. Chae[3], B. Lim[3], J. Kim[1,2].* 1) Biomedical Sciences, Seoul National University, Seoul, South Korea; 2) Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul, South Korea; 3) Department of Pediatrics, Seoul National University College of Medicine, Seoul National University Children's Hospital, Seoul, South Korea.

Moyamoya disease (MMD) is a cerebrovascular condition distinguished by bilateral stenosis and occlusion in the internal carotid arteries, causing development of an abnormal collateral network of vessels for compensatory circulation. Previous studies have reported RNF213 as a major susceptibility gene for MMD. In this study, we suspected there may be another element that could cause or have an effect on MMD since there are patients with MMD who do not carry any variant in RNF213. To discover additional factors that may cause MMD, we performed exome sequencing on 15 individuals (11 MMD patients and 4 of whom are the parents of 2 of these patients). We narrowed down the variants to SNPs in exonic region that were frequent in MMD patients, but not in general population. Our study discovered 6 nonsynonymous variants: EML6, SYNE1, CHRNB3, CDK5RAP2 , ZEB1 and SIPA1L3: all of these variants were shown twice in our patient samples. All 6 genes are candidate genes for causing MMD and may explain why RNF213 negative patients still display MMD symptoms. Further study is needed for verifying the specific pathways these genes are involved in to be assertive of these results.

2783S

Defining a new syndrome of cutis laxa, holoprocencephaly and cerebellar agenesis with overexpression of NRG3. *A. Ramalingam, K. Scott, G. Preston, A. Janssen, T. Kozicz, T.-J. Chen, E. Morava.* Hayward Genetics Center, Tulane University School of Medicine, New Orleans, LA 70112.

Neuregulins are signaling proteins that mediate cell-cell interactions. They play a critical role in the normal growth of organ systems and are essential for the embryologic development of the central nervous system. NRG3, encoding neuregulin3, is located at 10q23.1 and is expressed mainly in brain, testis, skin and muscle. Extracellular cleavage product of NRG3 binds to ERBB4 resulting in ligand-mediated tyrosine phosphorylation. A newborn with alobar holoprosencephaly, monoventricles, cerebellar agenesis and severe cutis laxa was diagnosed with a novel developmental syndrome. Skin biopsy of the patient showed normal elastin and collagen staining but an increased thickness of the epidermis with excessive wrinkling, decreased hair growth and decreased pigmentation. Microarray CGH analysis of the patient's DNA revealed a 34.7kb deletion in intron 5 of the NRG3 gene. The deletion was confirmed by qPCR to be present in the proband but not in the apparently normal mother. Transgenic mice that overexpress NRG3 in skin are hairless and have thick, pale and wrinkled skin as seen in cutis laxa. Patient fibroblasts showed increased cell growth and proliferation suggesting overexpression of cell signaling. The deleted region of the intron 5 of NRG3 contains binding site for a strong transcription repressor. We initiated a reporter assay to evaluate the functional consequences of the intronic NRG3 deletion. Similar to the NRG3 transgenic mice, immunohistochemistry in cell culture confirmed accelerated cell proliferation, compared to control cells, without increased apoptosis. We defined a new syndrome of holoprosencephaly, cerebellar anomalies and cutis laxa. Based on the presence of the strong repressor site in the intron 5 of NRG3, the overexpression phenotype in our patient, and the finding of cutis laxa in the overexpression mice model of NRG3 we hypothesize the pathogenic role of NRG3 in this complex developmental skin and brain anomalies.

2784M

A case of 16p11.2 duplication syndrome and review of the literature. *B.J. Ilagan, G. Ghaffari, M. Rodriguez, C. Hung, SA. Hosseini, OA. Bodamer.* Human Genetics, University of Miami, Miami, FL.

The 16p11.2 duplication syndrome is characterized by a spectrum of clinical findings including autistic behavior, developmental delay, intellectual disability, cranio-facial dysmorphism, microcephaly and epilepsy. Severity and expressivity of clinical symptoms can vary based on duplication size and other yet unknown factors with no evidence of gain or loss of specific gene function. The estimated prevalence of the 16p11.2 duplication syndrome is 3:10,000 individuals and as high as 1% in patients with autism spectrum disorder. The syndrome is a result of genomic structural change likely due to the 16p11.2 locus of about 600 kb (minimal critical duplication region) being situated between flanking segmental duplications having >99 percent sequence identity and regions of non-allelic homologous recombination (NAHR) events. Here we report a five year old female patient with a 3.6 Mb duplication including the 16p11.2 locus obtained by using Illumina's Infinium SNP Array platform. Genes in the duplicated region that may contribute to the patient's phenotype include PRRT2, PKC, DYT10, EKD1, BFIS2, BRIC2, ICCA, FUS, TLS, ALS6 and ETM4. The patient presented with developmental delay involving motor, speech and language, learning difficulties and microcephaly (<3rd percentile). There is no definitive history of seizures and/or autistic spectrum disorders. She was born as the first child of healthy, non-consanguineous parents of mixed ancestry. Diagnostic work-up including EEG, brain MRI/MRS and basic metabolic testing were within normal limits. We report this case to serve as awareness of 16p11.2 duplication syndrome and its complexities as indicated in current literature. Variation in and outside the 16p11.2 locus may or may not be considered part of the syndrome.

2785T

Maternal UPD(16) with IUGR, transient neonatal hypoglycemia and cholestasis. *H. Lesmana, R. Hopkin.* Human Genetics, Cincinnati Children's Hosp Med Center, Cincinnati, OH.

Uniparental disomy (UPD) is defined as the presence of homologous chromosomes from one parent. Maternal UPD of chromosome 16 is the most common reported UPD in the literature other than UPD(15) and has previously been reported to cause intrauterine growth retardation (IUGR) and a variety of congenital malformations. However there is no specific pattern of features documented in these cases suggesting for imprinted genes in this chromosome. Most cases are reportedly associated with confined placental mosaicism of trisomy 16 and some degree of placental insufficiency contributing to IUGR. Neonatal cholestasis has not been previously reported in maternal UPD(16). We report a new case of maternal UPD (16) which was initially suspected due to the presence of long contiguous stretch of homozygosity in chromosome 16 by SNP microarray. Maternal UPD(16) was subsequently confirmed through UPD testing using 12 microsatellite markers spanning chromosomes 16 in the proband, the mother and the father. This test revealed both maternal segmental heterodisomy and isodisomy on chromosome 16. Main clinical features include severe IUGR, transient neonatal hypoglycemia and cholestasis. Metabolic and molecular genetics testings for neonatal cholestasis and other inherited liver disorders failed to reveal any pathologic findings. We discuss reported cases of maternal UPD(16) and consider whether our patient's features may be due to disordered imprinting or unmasking of an autosomal recessive condition.

2786S

Falling serum estradiol levels prior to human chorionic gonadotropin on follicle growth and pregnancy outcomes in in vitro fertilization cycles. *X. Bao, JW. Xu, YP. Sun.* the First Affiliated Hospital of Zhengzhou University, ZhengZhou,China, Henan, China.

Background: Whether the falling of oestradiol on the day of HCG administration could influence in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI) pregnancy outcomes remains unknown. The aim of the study was to analyse the relationship between the falling estradiol levels prior to hCG administration and follicle ruptured before the oocyte retrieval procedure and outcomes on IVF /ICSI treatment. Material and Method: A retrospective study of the database in the First Affiliated Hospital of ZhengZhou University (Zhengzhou, China) to identify 813 IVF /ICSI stimulation cycles whose E2 levels fell prior to the day of hCG administration during IVF/ICSI treatment. Patients were assigned to three groups according to serum E2 drops on the day of HCG administration compare to the day before as follows: group 1 with <10% drop in serum E2 concentration; group 2 with 10-20% drop; group 3 with >30% drop; control group with E2 levels continued to rise until the day of hCG. Results: Follicle ruptured rates (6.1%, 9.1%, 13.5% vs 3.3%, P=0.000) prior to the oocyte retrieval procedure was highly significant among the three groups. The cancellation rates (21.3%,24.4%,28.1% and 17.4%,P=0.018) was significantly higher with the increasing serum E2 drops on the day of HCG administration compare to the control group. Implantation rates (33.0%, 30.9%, 34.6% vs 33.4%, P= 0.826) and pregnancy rates (50.0%, 45.6%, 50.4%, 51.0%, P=0.732) were no statistical significant in all study and control groups respectively. Conclusions: The falling estradiol levels prior to hCG administration might have a potential negative effect on folliclar growth and lead to higher cancellation rates, but the pregnancy rates for embryo transfer cycles are not compromised.

2787S

The LH gene mutation and controlled ovarian hyperstimulation. *M.R. Ranjouri[1], R. Davar,M.D.[2], M.H. Sheikhha,M.D.,Ph.D.[2].* 1) Department of Genetics and Molecular Medicine, Zanjan university of medical science, Zanjan, Iran; 2) Yazd Research and Clinical Center for Infertility, Shahid Sadughi University of Medical Science,Yazd,Iran.

Background: One of the most difficult problems of in vitro fertilization (IVF) treatment is the variability in the response to controlled ovarian hyperstimulation (COH) which ranging from poor to high, leading to IVF failure or complications related to ovarian hyperstimulation syndrome (OHSS). Objective: To evaluate the correlation between *LH*β G1502A polymorphisms in exon 3 of the LH gene and ovarian response to COH. Materials and methods: A total of 220 women treated with a long protocol for ovarian stimulation were studied. Three genotypes of GG, GA and AA were detected by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) analysis. Results: The most frequent genotype was GA (55.5%) whereas 44.5% of patients showed GG genotype and there was no patient with AA genotype. In total, 34 patients were poor responder, 154 were normal responder and 12 were hyper responder. In total 54.5% of normal responders, 61.8% of poor responders and 50% of hyper responders showed GA genotype. Discussion: Our results did not establish a significant relationship between this polymorphism and the ovarian response. Therefore it is still very difficult to use the genotype of patients for prediction of the ovarian response to stimulation.

2788S

Match Study of Sperm Relative Factors on the IVF Outcome. *M. Zhang, J. Xu, Y. Sun.* The first affiliated hospital of Zhengzhou University, Zhengzhou , Henan, China.

Objective: To investigate the effect of sperm relative parameters on the outcome of fertilization, in order to direct the selection of fertilization method. Methods: Collect 2 601 first Controlled ovulation IVF cycles from January 2012 to June 2013, match female factors(age, weight, cause of infertility, base FSH and LH, endometrial thickness on hCG day and the number of acquiring ovum) and get 457 pairs. Analyze the sperm relative factors on different outcomes. Results: Sperm normal morphology rate in pregnant cycles were significantly higher than those of non-pregnancy cycles(P<0.05), but qualified rate were same; sperm density, activity and survival rate were same in different groups(P>0.05). Conclusions: IVF cycles with different outcomes own different sperm normal morphology rate, which has higher reference values then other sperm factors.

2789S

Prenatal diagnosis of Apert Syndrome: fetal brain phenotype on imaging. *Z. Stark[1, 2, 3], G. McGillivray[1, 2], A. Sampson[2, 4], R. Palma-Dias[2, 4, 5], A. Edwards[6, 7], J.M. Said[2, 3, 8], G. Whiteley[9], A.M. Fink[2, 10, 11].* 1) Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Melbourne, Australia; 2) Fetal Medicine Unit, Royal Women's Hospital, Melbourne, Australia; 3) Maternal Fetal Medicine, Sunshine Hospital, Western Health, Melbourne, Australia; 4) Pauline Gandell Imaging Centre, Royal Women's Hospital, Melbourne, Australia; 5) Pregnancy Research Center, Department of Obstetrics and Gynaecology, University of Melbourne, Australia; 6) Fetal Diagnostic Unit, Monash Medical Centre, Melbourne, Australia; 7) The Ritchie Center, Monash Institute of Medical Research, Melbourne, Australia; 8) Northwest Academic Centre, University of Melbourne, Australia; 9) Department of Radiology, Monash Medical Center, Melbourne Australia; 10) Department of Radiology, University of Melbourne, Australia; 11) Medical Imaging Department, Royal Children's Hospital, Melbourne, Australia.

Objectives: The diagnosis of Apert syndrome relies on the identification of the classic triad of complex syndactyly of the hands and feet, craniosynostosis and midface hypoplasia. Diagnosis in the prenatal setting can be difficult as the onset of craniosynostosis is highly variable. We describe the brain imaging findings in 6 fetuses affected by Apert syndrome, and highlight the utility of dedicated neuroimaging as an adjunct to diagnosis. Methods: Retrospective review of ultrasound and MRI brain imaging obtained in 6 fetuses with a diagnosis of Apert syndrome. Results: Five fetuses had attenuation of the septal leaflets, and two had corpus callosum dysgenesis. All six had temporal lobe expansion and overconvolution. The temporal lobe abnormalities preceded the development of cranial deformity in two fetuses. Conclusion: Overconvolution of the temporal lobe cerebral mantle can be detected antenatally and is particularly conspicuous in the fetus when the normal brain is still relatively smooth (approximately 24 to 28 weeks of gestation). Detailed fetal brain imaging by neurosonography or MRI is capable of demonstrating this temporal lobe malformation, and can contribute to diagnostic certainty in cases of suspected Apert syndrome.

2790S

Fibrodysplasia ossificans progressiva: bilateral hallux valgus on ultrasound as clue for the first prenatal diagnosis for this condition -case report. *C. Maftei[1], I. Thiffault[1,3], J. Dubé[2], A-M. Laberge[1], E. Lemyre[1].* 1) Genetics, Ste-Justine Hospital, Montreal, Quebec, Canada, 3175 ch de la Côte-Ste-Catherine, Montréal, QC H3T 1C5, tel: (514) 345-4931, fax : (514) 345 4781; 2) Obstetric & Gynecology, Ste-Justine Hospital, Montreal, Quebec, Canada, 3175 ch de la Côte-Ste-Catherine, Montréal, QC H3T 1C5, tel: (514) 345-4931, fax : (514) 345 4781; 3) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO, USA,64108.

Fibrodysplasia ossificans progressiva (FOP (MIM: 135100)) is a very rare autosomal dominant condition with an incidence of 1:2 000 000. The clinical picture is characterized by congenital malformations of the great toes and progressive disabling heterotopic ossification. The genetic explanation is an activating mutation in the ACVR1, a BMP type I receptor, which promotes osteogenic differentiation of connective tissue progenitor cells. We report here how the presence of fetal bilateral hallux valgus, seen on second-trimester ultrasound, prompted clinicians to consider the diagnosis of FOP in the absence of family history. This classical foot malformation is present in almost all FOP patients reported in the literature. A 35-year-old French-Canadian pregnant woman was referred to our center for investigation and genetic counseling after a second trimester ultrasound identified a right ectopic kidney and a mega cisterna magna. A repeat ultrasound at 23 weeks in our center confirmed a right crossed ectopic kidney fused to the inferior pole of the left kidney, with good cortico-medullary differentiation. Bilateral hallux valgus was also observed. The cisterna magna was normal and all the other structures, including bones, were normal. Fetal growth was normal. Fetal echocardiogram was normal. The karyotype was normal (46,XY), as well as array CGH. A tridimensional ultrasound was performed to better characterize the feet and showed a hallux valgus deformation with plantar deviation of the first phalanx, and the absence of the second phalanx. The fetus was found to have the recurrent c.617G>A; p.R206H mutation in ACVR1 gene, confirming our suspicion of FOP. Post-mortem external examination showed normal growth parameters, some minor dysmorphic features but no other malformations. Bilateral hallux valgus was present and the first metatarsophalangeal articulation was rigid. Fetal X-rays showed important bilateral deviation in hallux valgus, first toes with only one phalanx and plantar deviation. This report highlights that malformed great toes on prenatal ultrasound, especially hallux valgus, should raise the suspicion of FOP even when other atypical anomalies are found. Establishing the diagnosis allows for appropriate genetic counseling and informed decision-making for the family.

† Trainee Award Finalist

2791S

Campomelic dysplasia: Prenatal Ultrasound and Autopsy Findings in Early Pregnancy. *K. Chong[1], N. Martin[1], A. Toi[2], S. Keating[3].* 1) Dept OB/Gyne, The Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, Toronto, ON, Canada; 2) Dept of Medical Imaging, Mount Sinai Hospital, Toronto, ON, Canada; 3) Dept of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada.

Campomelic dysplasia (CD) is a severe skeletal dysplasia characterized by cleft palate, tracheomalacia with respiratory compromise, bowing of the long bones and ambiguous genitalia. CD is an autosomal dominant genetic condition with few published case reports of gonadal mosaicism. Fetal ultrasound findings of CD have mostly commonly been reported in the second trimester with bent femurs, normal upper limbs and hypoplastic/absent scapula, however, some reports of cystic hygroma with lower limb anomalies have on fetal autopsy revealed CD. Translocations, small rearrangements and mutations involving the SOX9 gene on chromosome 17q have been identified as the cause of CD. We report a couple with 2 early pregnancies affected with cystic hygroma +/- other anomalies in the first trimester. The pregnancies were terminated by dilatation and curretage, however, careful pathology examination and X rays combined with molecular testing for SOX9 revealed a diagnosis of CD. Follow-up testing and examination of parents confirmed gonadal mosaicism in a symptomatic parent. This case highlights the importance of fetal pathology, even in early surgical samples, and the importance of clinical examination of parents in rare cases of skeletal dysplasia recurrence.

2792S

Importance of Fetal Fraction Analysis for cfDNA Testing in the General Pregnancy Population. *E. Wang, C. Struble, C. Kingsley, R. Steeke, A. Batey, D. Hollemon, A. Oliphant, T. Musci.* Ariosa Diagnostics, Inc, San Jose, CA.

Introduction: The Harmony™ Prenatal Test uses an assay method, Digital Analysis of Selected Regions (DANSR™), for analysis of chromosomes 13, 18, 21, X and Y as well as other chromosomes to measure fetal fraction. Products from the DANSR assay are then analyzed with the Fetal fraction Optimized Risk of Trisomy Evaluation (FORTE™) algorithm to assess patient-specific risk of trisomy. Alternative tests, such as those using massively parallel shotgun sequencing, use a Z-statistic or Normalized Chromosome Value (NCV) to discriminate between normal and abnormal chromosomal counts without accounting for fetal fraction. **Methods**: A general pregnancy population cohort of 15,841 women between 10.0 to 14.3 weeks gestation were followed to pregnancy outcome. Z-statistics were computed using previously described standard Z-test of proportions and compared to FORTE risk scores generated by the Harmony test. A Z-statistic of ≥ 3 was considered 'Positive' for trisomy and a FORTE risk score of ≥ 1% was 'High Risk' for trisomy. **Results**: The cohort included 38 cases of trisomy 21 (T21) and 10 cases of trisomy 18 (T18). A subset of participants (n = 11,185) had trisomy 13 (T13) analysis.

PPV: Positive Predictive Value

Method	T21	Non-T21	T21 PPV	T18	Non-T18	T18 PPV	T13	Non-T13	T13 PPV
FORTE (%)	38/38 (100)	9 / 15803 (.06)	38/47 (81)	9/10 (90)	1 / 15831 (.01)	9/10 (90)	2/2 (100)	2 / 11183 (.02)	2/4 (50)
Z ≥ 3 (%)	38/38 (100)	34 / 15803 (.22)	38/72 (53)	8/10 (80)	43 / 15831 (.3)	8/51 (16)	2/2 (100)	65 / 11183 (.6)	2/67 (3)

The cumulative test discordant rate for non-trisomies with FORTE was 0.08% compared to 1.07% with the Z-statistic. **Conclusion**: Analysis using FORTE to incorporate fetal fraction in cfDNA testing yields a >10 fold reduction in discordant results compared to the Z-statistic. This will become increasingly relevant as cfDNA testing gets more broadly used in the general pregnancy population.

2793S

Maternal subchromosomal abnormality identified through noninvasive prenatal testing (NIPT). *C. Settler[1], T. Boomer[1], J. Newell[2], P. Santiago-Munoz[2], N. Teed[1], J. Saldivar[1], N. Dharajiya[1].* 1) Sequenom Laboratories, San Diego, CA; 2) Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX.

Introduction: Noninvasive prenatal testing (NIPT) uses circulating cell free DNA (ccfDNA) for the evaluation of fetal chromosomal abnormalities. Studies indicate that about 10-15% of the ccfDNA in maternal plasma is derived from the pregnancy, and thought to be placental in origin. The remaining ccfDNA is contributed by the mother. Thus, evaluation of ccfDNA to identify chromosomal abnormalities has the potential to identify both fetal and maternal anomalies. Here we report a case of a maternal deletion identified through a combination of NIPT and clinical history. Methods: Maternal blood samples submitted to Sequenom Laboratories for MaterniT21™ PLUS testing were subjected to DNA extraction, library preparation, and whole genome massively parallel sequencing as previously described. Sequencing data were analyzed using a novel algorithm to detect trisomies and other subchromosomal events. Results: A 20 year old G3P1010 had NIPT secondary to positive serum screening. NIPT results were negative, but ultrasound identified multiple anomalies. Karyotyping performed on amniocentesis sample subsequently revealed a ring chromosome 18 [46,XY,r(18)(p11.3q23)]. The patient elected pregnancy termination. Maternal karyotype was normal (46,XX). The patient conceived again and NIPT was again negative; no anomalies identified by 20 weeks' gestation. The history of the ring 18 chromosome was provided to the laboratory, and based on this information, detailed bin data was reviewed for chromosome 18, revealing an apparent 0.75 Mb deletion at 18q22.1>q22.1, likely maternal in origin. Maternal microarray analysis confirmed a 773 Kb interstitial deletion with the following result: arr 18q22.1 (63, 884,429-64,657,305)x1. The patient declined invasive testing for the current pregnancy (ongoing). Familial studies are in progress to characterize the segregation of the chromosomal abnormality. Conclusion: Clinicians should be cognizant that maternal subchromosomal abnormalities may be identified by NIPT using massively parallel sequencing. Accurate clinical information provided to the laboratory is critical for interpretation of test results. Additionally, standard karyotyping and in many cases FISH does not have sufficient resolution to confirm subchromosomal events detected by NIPT. In these cases, microarray studies are superior to standard karyotype and FISH testing for confirmation of the suspected abnormality.

2794S

Thrombophilic Mutations for Recurrent Miscarriage in Iranian women with or without Thrombophilia. *H. Mirtavoos-Mahyari[1,3], B. Poopak[2].* 1) National Institute of Tuberculosis and Lung Disease, Tehran, Tehran, Iran National Institute of Tuberculosis and Lung Disease, Tehran, Iran,MD; 2) Department of Hematology and Laboratory Blood Bank, Tehran Medical Branch, Islamic Azad University, Tehran, Iran, PhD; 3) Department of Medical Genetic, Tehran University,Tehran,Iran, PhD Student.

Background: Inherited thrombophilia may increase susceptibility of fetal loss. This study was performed to investigate the frequency of some factor's mutations in ethnic Iranian patients . Methods: This epidemiologic case study performed with 609 patients whom were referred to Payvand Medical and Specialty Laboratory in Tehran, Iran, from 2011 through 2012. Peripheral blood samples of the patients with abnormal thrombosis were analyzed for mutations in Factor V Leiden, FII , MTHFR and PAI-1 . Mutation analysis was accomplished by PCR-reverse dot blot. Results:The mean age of patients was 29.58 years. The results are shown in Table 1. Table1.Distribution of inherited thrombophilia

	Genotype		
	Homozygous	Heterozygous	Wild type
Factor V Leiden(n=176)	0	17	159
MTHFR(n=195)	21	90	84
FII (n=102)	0	6	96
PAI-1(n=136)	23	78	35

Discussion: MTHFR and PAI-1 polymorphisms were identified in more than fifty percent and factor V lieden mutation is higher than some other countries . Therefore, testing and providing genetic consultation may be needed in Iran.

2795S

Validation of a taxonomy of genetic conditions for pre-conception genetic carrier testing. *T.L. Kauffman[1], N. Neil[3], C. McMullen[1], M.C. Leo[1], J. Reiss[1,2], B. Wilfond[4], J. Davis[1], M. Gilmore[2], P. Himes[2], F. Lynch[1], K.A.B. Goddard[1].* 1) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 2) Kaiser Permanente Northwest, Portland, OR; 3) Decision Research, Eugene, OR; 4) Truman Katz Center for Pediatric Bioethics, Seattle Children's Hospital, Seattle, WA.

Background: We are studying the use of genome sequencing in the context of pre-conception genetic carrier testing for women and their partners. The study team created a taxonomy of four categories of results that participants could choose to receive categorizing the types of genetic conditions found. With patient input, the categories were revised to: shortened lifespan, serious, mild, unpredictable outcomes, and adult onset. We then used risk perception methodology to obtain support for the taxonomy. Methods: We emailed a link to an online survey to fifteen hundred Kaiser Permanente Northwest members who had previously received preconception genetic carrier testing. The survey included descriptions of genetic conditions and their associated attributes (physical impairments, cognitive impairments, effectiveness of treatment, age at onset, and variability of expression). The study team selected 20 genetic conditions to include that varied according to these attributes. We created five versions of the survey, each with four conditions. Respondents rated their reactions to each condition on a seven point Likert scale, including impact of the condition on a child and the family, predictability and controllability of the condition, and how it would change their plans for pregnancy. Enrollment was limited to the first approximately 40 respondents per version of the survey. Results: We assigned the 20 conditions into the taxonomy categories. We then determined if survey participants' (N=204) reactions to the conditions differed depending on category. Based on the results of factor and cluster analyses, the conditions were grouped according to ratings of hopeful feelings, controllability, predictability, visibility, severity, and negative feelings. Overall, survey responses supported our five category taxonomy of genetic conditions. Conditions in the serious and mild categories were the hardest for patients to distinguish. Conclusions: Survey respondents made meaningful distinctions between the genetic conditions presented to them. While the categories were derived from experts and revised with focus group input, the results provide empirical support for the category distinctions. Future research utilizing the same methodology with a wider variety of conditions could be beneficial.

2796S

Tay-Sachs disease enzyme carrier screening does not perform well in a pan-ethnic population when compared to DNA analysis. *V. Greger, C. Perreault-Micale, A. Frieden, A. Kramer, N. Faulkner, D. Neitzel, S. Hallam.* Good Start Genetics, Cambridge, MA.

Carrier screening for Tay-Sachs disease (TSD [MIM 272800]), an autosomal recessive disorder, is recommended in the Ashkenazi Jewish (AJ) population due to its high frequency. DNA analysis of the *HEXA* gene using a limited panel of mutations focused on the AJ population and HexA enzyme analysis are typically used for TSD carrier screening individually, or in combination. The goal of this study was to compare the effectiveness of DNA testing for TSD using an expanded panel of mutations along with enzyme analysis, in a large pan-ethnic population. Both HexA enzymatic activity and *HEXA* DNA analyses were performed in 7148 individuals referred to us by IVF clinics in the United States for TSD carrier screening. HexA enzyme activities were measured by an outside independent laboratory. Next Generation DNA sequencing (NGS) of the coding region and exon-intron borders was used to assess for the presence of more than 70 variants with our extensive *HEXA* test panel. Our results showed that 181 individuals tested positive by enzyme analysis. A pathogenic variant was identified in 55 (30%), and a pseudodeficiency allele was present in 23 (13%). DNA analysis revealed no pathogenic variants in 103 of the 181 (57%). In 6397 enzyme negative patients, 3 had a pathogenic B1 allele, as determined by DNA analysis. Of the 570 (8%) patients within the enzyme inconclusive range, 6 had pathogenic variants. The observed carrier frequency in non-AJ populations is substantially higher with screening by enzyme analysis (corrected for pseudodeficiency alleles) than DNA analysis: 1/45 vs. 1/111 overall for enzyme vs. DNA analysis respectively, 1/9 vs. 1/252 for African Americans and 1/86 vs. 1/196 in Caucasians. Estimates based on disease incidence place the carrier frequency for TSD in the general population in the range of 1/250 - 1/300, much closer to carrier frequencies derived by DNA analysis. We conclude that while HexA enzyme carrier screening seems to reflect TSD status appropriately in most AJ individuals, it is problematic for carrier screening in pan-ethnic populations due to its high rate of false positive and inconclusive results. Our data indicate that ACOG's recommendation for the use of TSD biochemical analysis may need to be revised for non-AJ populations and/or the normal, inconclusive, and carrier ranges for enzyme analysis should be re-assessed for non-AJ individuals.

2797S

Analysis of Y-chromosomal microdeletions in an azoospermic patient candidate for an assisted reproductive technique. *M. Rongioletti, F. Papa, C. Vaccarella, MB. Majolini, A. Luciano, M. Belli, P. Scotaccia, B. D'Andrea, V. Mazzucchi, G. Liumbruno.* Clinical Pathology Department, San Giovanni Calibita Hospital, Rome, Italy.

Introduction Y-chromosomal microdeletions are the second most frequent genetic cause of male infertility. As stated in the guidelines of the European Academy of Andrology (EAA) for molecular diagnosis of Y-chromosomal microdeletions, this kind of analysis should be performed in patients with azoospermia or severe oligospermia with sperm concentrations < 2×106/ml. In these patients is also convenient analyze other clinical parameters such as hormone levels, testicular volume, varicocele, infections etc, although these parameters do not have any predictive value. Here we report a lab case of an azoospermic patient candidate to a testicular sperm extraction (TESE) for performing an assisted reproductive technology. Methods According to EAA guidelines, we first performed the multiplex PCR amplification of genomic DNA with the Ampli-Y Chromosome and Ampli-Y Chromosome Extension kits (Diachem - Italy). As a confirmatory test we also utilized the Devyser AZF kit (Devyser Ab - Sweden). The execution of kariotype was performed in an external laboratory. Conclusions The Y-chromosomal microdeletions analysis performed showed a lack of AZF b and c regions of Y-chromosome, confirmed by both kit utilized for Y microdeletions diagnosis. The patient had also high levels of FSH and LH, low level of inhibin b and normal levels of prolactin and testosterone hormones. Moreover after the execution of kariotype test, the patient was found to be carrier of a severe form of chromosomal mosaicism (45X[99]; 46XY[1]) that confirmed Y microdeletion diagnosis. In the overall these results suggested clinicians to do not perform TESE since, as reported in literature, the chance for testicular sperm retrieval in these cases is virtually zero. Our case confirm the need of a good laboratory practice to support clinical decisions.

2798S

First Experiences with Non invasive Genetic testing in Switzerland. *J. Esslinger[1], S. Hotz[1], L. Risch[2], U. Wiedemann[2].* 1) Institute for family research and counseling, University of Fribourg, Fribourg, Switzerland; 2) Labormedizinisches Zentrum Dr. Risch, Liebefeld BE, Switzerland.

Introduction In July 2012 the NIPD Praena-Test of Lifecodexx, was implemented in Switzerland. NIPD-testing received high media coverage. Initially, a lack of clarity regarding the target patient-community, the length of reporting time and the indication for this test existed. Referral was made via local gynaecologist, hospital centre or private person. We evaluated data to clarify whether a significant "patient pattern" exists for this new test. **Material und methods** A total of N=378 cases were included in this analysis, representing all NIPT-tests requested in the laboratory Dr. Risch, as a typical big private laboratory in Switzerland, from the beginning in August 2012 until the reference date December 31st 2013. **Results** As expected due to location, most patients were Swiss (92,1%), wherefrom 72,1% from the German-speaking part of Switzerland, 27,6% from the French-speaking part, 0,3% from the Italian-speaking part. Most foreign patients were from Liechtenstein, France and Italy or other European countries. Patients were between 18 and 46 years old, with a mean of 36 years (*SD*=4,4), and a most prevailing age of 39 years. The mean pregnancy week (PW) was 14 (*SD*=2,4); with a min. PW of 9 and max. PW of 28. Declared main indications for testing: maternal age >35 years (46%), conspicuous first trimester screening (1TT) (14%), combination 1TT/age (13%) and wish (13%). Further combinations such as age/wish, age/previous trisomy as well as familiar burden have each been indicated in 2% of patients. Other indications were the case in 1% or less of women. Two thirds (67%) of patients were referred from local gynaecologists, 30% from hospital centres, 3% were self-registered patients. We found no significant difference in indication between gynaecologists and hospital centres. The mean reporting time was 16 days (*SD*=4,8), with a min. of 6 and a max. of 53 days (incl. repetition). In 4% of the cases, the first blood sampling did not provide a valid result. One case showed a trisomy 13, one case a trisomy 18 and 4 cases a trisomy 21. The typical NIPT-patient after 18 months of test-availability was Swiss, 39 years old, with a gestational age of 14 weeks and maternal age as indication. After initially long reporting time, a statistically significant decrease in reporting time was achieved throughout the examination period during months 10-18 (*M=15*, *SD*=3,4) vs. months 1-9 (*M=17*, *SD=6,4*; p=0.001). The majority of patients (66%) receive their result with in 15 days.

2799S

Carrier screening for recessive disorders through exome sequencing. *P. Makrythanasis[1], A. Massouras[1], S.E. Antonarakis[1,2,3].* 1) University of Geneva, Geneva, Switzerland; 2) University Hospitals of Geneva, Geneva, Switzerland; 3) iGE3 Institute of Genetics and Genomics of Geneva, Geneva, Switzerland.

Detection of carrier status for certain recessive Mendelian disorders is a well-accepted health-care practice in several countries since the 70s and aims at the prevention of frequent severe monogenic disorders. Exome sequencing in combination with the increasing knowledge of human pathogenic variation provide the possibility to perform a carrier screening for all the known recessive Mendelian disorders. Such a screening would allow for a much more informative genetic counseling and may alter the total prevalence of the known recessive disorders. In order to test this hypothesis we have used exome sequencing data from 104 individuals of European origin and have identified the total number of likely pathogenic variants in the >1600 recessive disorders for which the responsible gene is known. The mean value was 18.2 variants per individual. Consequently we have randomly paired these exomes in order to create 5356 fictive couples. 33.14% of these couples have at least one gene for which both members are heterozygous for a likely pathogenic variant. These preliminary results exhibit an upper estimate of at risk couples but more precise knowledge and definition of the pathogenic potential of each variant will render the carrier detection more accurate and make it a potent test for family planning.

2800S

IL-10 Promoter polymorphism (592C/A) in women with recurrent miscarriages in Punjabi population (India). *A. Kaur, N. Sudhir, B. Badaruddoza.* Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

Cytokines, expressed by several cell types and tissues, are diverse family of soluble proteins which play a fundamental role in differentiating the nature of immunologic response. Interleukin -10 (IL-10) a type of cytokine has a dual immunological role as it act as either immunostimulatory or immunosuppressive. IL-10 is fundamental in promoting normal pregnancy outcomes as it is involved in placental angiogenesis and inhibits the secretion of inflammatory cytokines. The promoter region polymorphisms in IL-10 gene have reported to interfere in its regulation. IL-10 down regulation may affect the placental angiogenesis affecting fetal growth and might end in miscarriage. In the present study IL-10 592C/A promoter polymorphism was investigated in women with recurrent miscarriages and in women with atleast one live birth and no history of miscarriages. For this, a total of 182 samples (92 cases and 90 controls) were genotyped using PCR-RFLP method. In the present study, 56.5% of women with recurrent miscarriages has homozygous wildtype genotype (CC), while 4.76% and 2.04% of women were having heterozygous and homozygous mutant genotype, respectively. In case of control group 52.2% showed CC gentotype, 40% had CA genotype and homozygous mutant (AA) was found in 7.8% of women. Three genetic models such as dominant, codominant (additive) and recessive were used to analyze the data, none of the models have been found significant except recessive model which showed positive association (OR: 1.45;0.53-3.98,P= 0.373). no significant difference were also observed in either genotype (P= 0.542) or allele (P=0.907). The present case control study did not show any significant association with recurrent miscarriages. Therefore, no overall significant association was found between IL-10 592C/A and recurrent miscarriages in present population.

2801S

Tay-Sachs carrier screening by enzyme and molecular analyses in the New York City Black population. *G.A. Lazarin[1], E. Spiegel[2], K. Berentsen[1], K. Brennan[2], N. Mehta[1], I.S. Haque[1], R. Wapner[2].* 1) Counsyl, South San Francisco, CA; 2) Department of Obstetrics and Gynecology, Columbia University Medical Center, New York, NY.

Introduction: Carrier screening for Tay-Sachs disease (TSD [MIM 272800]) analyzes the *HEXA* gene or HexA enzymatic activity. The former can be more costly or may not detect all carriers. Enzymatic assay approximates phenotype but is subject to misinterpretation (e.g., indeterminate results). Enzymatic assay is often suggested as the optimal detection method in non-Ashkenazi Jewish people. However, this protocol has only recently been considered for routine, pan-ethnic TSD carrier screening with the advent of expanded carrier screening (ECS). Columbia University implemented universal ECS in 2013. Results of DNA and enzymatic analyses for TSD carrier status in the Black population are reported. **Methods:** 123 individuals reporting Black ancestry underwent ECS, including *HEXA* mutation screening (9 mutations) and lymphocyte HexA activity levels (WBC HexA %). Mutation screening was performed at Counsyl (South San Francisco, CA) and most enzyme assays were performed at Mount Sinai Genetic Testing Laboratory (New York, NY); one enzyme assay was performed at Mayo Medical Laboratory (Rochester, MN). Individuals with carrier or indeterminate enzyme levels and negative targeted mutation analysis were subsequently offered next-generation sequencing (NGS) of exons 1-14, also at Counsyl. **Results:** Of 123 individuals screened by enzyme, n=16 (13.0%) were deemed carriers and n=15 (12.2%) had inconclusive results. Mean WBC HexA % were: 58% in all individuals, 45% in positive results, 52% in inconclusive results and 61% in negative results. All of these 31 individuals were negative for a 9-mutation DNA analysis. At time of writing, NGS was completed on 4 individuals with positive results and 4 with inconclusive results. One likely deleterious variant, c.1510C>T, was found in 1 individual. One individual had no variation from the reference sequence. In the remaining 6, 2-5 known benign variants (total of 6 unique variants detected 34 times) were identified. **Conclusions:** Using enzyme analysis in this Black population, we found an unexpectedly high positive or inconclusive rate. Such results necessitate molecular testing, and we have initially found most do not have pathogenic mutations. We continue to collect data characterizing non-negative enzyme results, which may reveal unrealized deleterious variants or pseudodeficiency alleles. The data may suggest recalibration of enzyme reference ranges or an alternative routine screening protocol for certain populations.

2802S

Prospective exome sequencing of consanguineous couples: First steps. *H. Meijers-Heijboer, M. Teeuw, MD, Q. Waisfisz, PhD, P.J.G. Zwijnenburg, MD, PhD, E.A. Sistermans, PhD, J. Weiss, PhD, L. Henneman, PhD, L.P. ten Kate, MD, PhD, M.C. Cornel, MD, PhD.* VU University Medical Center, Amsterdam, Netherlands.

With 10% of people worldwide in a consanguineous relationship or having consanguineous parents, consanguinity is one of the most frequent risk factors for congenital disorders. In theory, prospective exome sequencing of consanguineous couples could identify couples who are carriers for autosomal recessive diseases, and empower such couples to make informed reproductive decisions. The expected yield of this approach can be investigated in silico, leaving the question unanswered whether expectations would be realized in practice. Therefore, we sent blood samples of 4 pairs of parents, having one or more children affected by an autosomal recessive disorder, without revealing any diagnostic information, to our laboratory (which had not been involved in the testing of the children earlier on). The study was restricted to find, in both parents of each couple, identical, previously described, or evidently pathogenic mutations in over 400 genes known to result in severe autosomal recessive disorders. After de-blinding, results were compared to pre-existing information. Out of the 6 autosomal recessive disorders known to the 4 couples studied, two were correctly identified. As expected, given the pipeline used, large deletions were not identified, explaining one unidentified disorder. The same applied to mutations in genes not present in the gene list, mutations outside the exons and consensus splice sites, and mutations that were not evidently pathogenic and not reported before, each explaining an unidentified disorder. In one of the couples carriership of both partners was found for a disorder which was not known to the family yet. As our approach was not able to identify intronic mutations or large deletions, provisions for frequent disorders with these kind of mutations (e.g. thalassemia's and spinal muscular atrophy) should be realized. The restriction to detecting only couples with identical mutations diminishes the risk of finding unsolicited findings and shortens the time needed for analysis, but also results in missing pairs with different mutations in the same gene. On the other hand, for disorders with all mutations identifiable, finding couples with identical mutations (either by descent or state) reduces the risk of the consanguineous couple to a level below the risk for (untested) non-consanguineous couples. Again special provisions might be needed for frequent disorders with one frequent mutation, such as cystic fibrosis.

2803S

A new computational approach for reproductive genetic risk assessment with over 10-fold greater sensitivity than carrier screening. *A. Silver[1], C. Borrot[1], B. Spurrier[1], M. Silver[1], J. Greisman[1,3], N. Delaney[1,2], R. Lim[1], L. Silver[1,3].* 1) GenePeeks, Cambridge, MA; 2) Broad Institute, Cambridge, MA; 3) Princeton University, Princeton, NJ.

Carrier testing (CT) is unique among diagnostic tests in that disease can only be observed in persons other than those being tested. As a result, clinical validation of many rare disease-causing mutations may be difficult if not impossible. Nevertheless, CT can be applied only to validated mutations in the detection of recessive disease risk prior to conception. CT is a particularly poor indicator of reproductive risk for anonymous donors to sperm and egg banks. We have developed a computational system that simulates the creation of haploid gametes from a person's genomic or exomic sequence information. The Monte Carlo-generated "virtual gametes" from two prospective genetic parents are combined to form a pool of diploid "virtual progeny" (VP) genomes. Each VP genome is analyzed individually for the likelihood that it will induce a disease phenotype based on clinical annotations and the combined biochemical properties of the products of the two copies of each gene. Our isolated variant scoring method shows a sensitivity of 95% on defined clinical variants. We applied our analysis to 440 well-characterized recessive disease-associated genes in virtual matings of 2,500 exomes from phase 3 1000 genomes data. The detected risk of recessive disease is over 10-fold greater than predicted by an ideal all-encompassing carrier testing protocol.

2804S

Genetic Polymorphism Of Hsp-70 & Tp-53 Gene In Indian Preeclamptic Women With Placental Vasculopathies. *S. KUMAR, A. LOMASH, A. SINGH, S. KHALIL, SK. POLIPALLI, S. GUPTA, S. KAPOOR.* MAULANA AZAD MEDICAL COLLEGE, NEW DELHI, NEW DELHI, India.

Objective: Complications in the later phases of pregnancy are often summaried as placental vasculopathies, which comprises placental infarction, villous infarct, villous fibrosis, fibroid necrosis of decidual vessels, calcification, cytotrophic proliferation, increases the maternal & fetal morbidity and mortality upto 40%. Polymorphic variation in HSP-70 & TP-53 genes results in convergence of multiple etilogical factors, leading to endothelial cell dysfunction with clinical manifestations of placental vasculopathies with preeclampsia. Aim of the study was to detect polymorphic variation within the HSP70 & TP53 genes with respect to placental vasculopathies in preeclamptic women. Methods: Pilot study consisting 30 preeclamptic (proteinuria & hypertensive) women with placental vasculopathies & 30 healthy pregnant controls were recruited post diagnosis. Polymorphic varation in HSP 70gene (H S P A 1 A G 1 9 0 C { H S P 7 3 } , H S P A 1 B A 1 2 6 7 C { H S P 7 2 } , HSPA1LT2437C{HSP70-hom}) & TP53(p53 Codon 72 polymorphism) were ascertained by using PCR-RFLP Method. Results: Control & Patient group's, gestational age 38.5657+/-1.03138 & 36.4743+/-3.206(p=0.003), maternal age 24.64+/-2.447 & 25.04+/-3.259 yrs, Placental weight 415+ 45 gm & 280+86gms(p=0.004).Immunohistopathological features in Controls/Case were Placental Infarct 0.00%/16.66%, Fibroid necrosis of DV's 0.00% / 20%, Cytotrophoblast proriferation 10.0%/53.33%, Villous fibrosis 20.0%/80.0%, Hypovascularity of villi16.66%/86.66%.Genotypic distribution in Control vs Case for HSP70 gene were , HSPA1A190 (GG,GC,CC) 0.00%,100%,0.00% vs 0.00%,90.0,10% ; HSPA1L2437 (TT,TC,CC) 20%,80%,0.0% vs 0.0%,80%,20% ; HSPA1B1267 (AA,AG,GG) 60%,40%,00 vs 0.00,100%,0.00% & for TP53 gene (p53 Codon 72 Proline/Arginine) were Pro/Pro 36.66% / 23.33% & Pro/Arg 63.33%/76.66%. Conclusion: This is a pilot study to identify that HSPA1B (A1267G) {HSP72},polymorphism is a predisposing genetic risk factor (p value= 0.0022) in preeclamptic women with placental vasculopathies. It shows positive association with gestational age, placental weight , Hypovascularity of villi and along with Fibrous villous in patients. However. HSPA1LT2437C, TP53(p53 Codon 72 polymorphism) (HSPA1AG190C{HSP73}, does not show any association independently or in combination with the disease. The application of the above study to the entire population would greatly enhances its importance.

2805S

Exome chip evaluation of genetic variants for association with uterine fibroids. *M.J. Bray[1], T.L. Edwards[1,2,3,4], K.E. Hartmann[2,3,5], D.R. Velez Edwards[1,3,5,6].* 1) Center for Human Genetics Research, Vanderbilt University , Nashville, TN; 2) Department of Medicine, Vanderbilt University, Nashville, TN; 3) Vanderbilt Epidemiology Center, Vanderbilt University, Nashville, TN; 4) Institute for Medicine and Public Health, Vanderbilt University, Nashville TN; 5) Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN; 6) Division of Epidemiology Department of Medicine, Vanderbilt University, Nashville, TN.

Uterine fibroids (UFs) affect up to 77% of women by menopause, account for $9.4 billion in yearly healthcare costs, and disproportionately impact minorities. Although UFs are heritable, genetic causes are poorly understood. The first genome-wide association study (GWAS) of UFs was performed in 2011 in a Japanese population. However, to date few large-scale genetic studies have been performed in US populations. The objective of this study is to conduct a whole exome association study of UF risk in European Americans (EA) and African Americans (AA). This is a case-control genetic association study of UFs defined using pelvic imaging data obtained from an electronic medical record biorepository (BioVU DNA Repository). Logistic regression adjusted for ancestry and age was used in 728 EA and AA DNA samples (EA: 246 cases, 243 controls; AA: 121 cases, 118 controls) to evaluate SNPs for association, stratified by race. Meta-analyses of results with fixed effect models were performed to obtain combined evidence for associations across racial groups. Our strongest association within EAs was within t-complex 11, testis-specific like 1 (TCP11L1, rs1064005, OR = 1.79, 95% CI = 1.36-2.32, p = 3.94×10-5) and within multiple SNPs within the HLA region (smallest p 9.28×10-5). It is of note that the exome chip has enhanced coverage of the HLA region compared to GWAS arrays. Among AAs strong associations were observed within glipican 4 (GPC4, OR = 0.46, 95% CI = 0.31-0.68, p = 8.73×10-5). Meta-analysis across EA and AA further strengthened the associations observed at TCP11L1 (smallest meta-p = 4.09x-7, with effect sizes consistent across races). Prior gene expression studies of TCP11L have shown increased expression in tumor tissues compared to normal tissues across several different cancers, and variants within HLA genes have been associated with UF risk in prior studies. These pilot data suggest common variants in exonic regions increase risk for UF in both EA and AA populations. However, further validation of our study findings is needed to confirm our results. We are currently genotyping >4,000 samples for GWAS and exome arrays to further evaluate the relationship between gene variants and UF risk.

2806S

Familial Infertility with Sex-limited Autosomal Recessive Inheritance. *H. Huang[1,2], H. Deng[1], H. Xiao[2].* 1) Department of Biostatistics and Bioinformatics, Tulane University, NEW ORLEANS, LA; 2) the Institute of Reproduction & Stem Cell Engineering, Central South University , Changsha, P. R. China.

BACKGROUND: In mammals, the zona pellucida is a glycoprotein matrix surrounds oocytes, which is vital to the production of oocytes in early development, fertilization, and protection of the early prior to implantation. Previous studies of zona pellucida defects in humans have not identified mutations the four genes *(ZP1, ZP2, ZP3,* and *ZP4*) encoding the ZP glycoproteins, which are secreted by growing oocytes. **METHODS:** The clinical and genealogical characterization was collected from the patients who failed in IVF (*in vitro* fertilization). Oocytes were removed away from granular cells and were observed with a visible light microscope. We firstly chose four candidate genes (*ZP1-ZP4*) to screen for potential mutations. We analyzed ZP1 and normal ZP3 in oocytes *in vitro*, from the patients and control subjects using immunofluorescently tagged antibodies. We gave one of the patients ART with oocyte donation, and followed all of the patients' birth situation. **RESULT:** Four sisters of the family were diagnosed with primary infertility. The eggs of two of the four sisters had a similar abnormal phenotype were not surrounded by a zona pellucida. We detected a homozygous frameshift mutation in *ZP1* in six members of the reported family, resulting in the truncation of ZP1 (p. I390fs404X). We propose that this prevents the formation of the zona matrix, and thus sterility. The immunofluorescen in oocytes showed no evidence of extracellular ZP1 and ZP3 surrounding the mutated eggs. ZP3 and mutant ZP1 showed diffuse staining in the patient's eggs. Merged images showed co-localization of signals throughout the cytoplasm in eggs from the patient. The treatment was successful and the patients delivered two babies. Her brother who carried a homozygous mutation had male fertile feedback, because his partner was pregnant after cohabitating for three months. **CONCLUSIONS:** We describe a novel form of infertility with a sex-limited autosomal recessive mode of inheritance, characterized by abnormal eggs that lack a zona pellucida. We identified a homozygous frameshift mutation in *ZP1* in six members of the family. Our data suggest that the aberrant ZP1 results in the sequestration of ZP3 in the cytoplasm, thereby preventing the formation of the zona pellucida around the oocyte, leading to infertility. The egg-donated treatment was effective. **KEY WORDS:** Infertility, ZP1, zona pellucida, oocyte, IVF, ART,sex-limited inheritance, egg donor.

2807S

Expression of Aurora Kinase C splice variants in human oocytes and cumulus cells. *J.E. Fellmeth[1], K. Schindler[1], N.R. Treff[2].* 1) Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, New Jersey, United States of America; 2) Division of Reproductive Endocrinology, Department of Obstetrics, Gynecology and Reproductive Science, Robert Wood Johnson Medical School, Rutgers University, New Brunswick; Reproductive Medicine Associates of New Jersey, Morristown, New Jersey.

Objective: Accurate chromosome segregation during meiosis I (MI) is essential for generating eggs with the proper chromosome complement. This segregation is highly error-prone in humans leading to miscarriage or offspring with developmental disorders. The Aurora family of protein kinases is a well-established regulator of chromosome segregation during mitosis and meiosis. Aurora kinases A and B (AURKA and AURKB) are expressed in nearly every cell type, whereas Aurora kinase C (AURKC) is only highly expressed in gametes. AURKC was identified in sperm and missense mutations in AURKC are linked to formation of polyploid sperm and infertility. Studies in mouse oocytes indicate that AURKC is also required to regulate MI chromosome segregation in females. To begin to determine if AURKC function is conserved between mouse and human, we used quantitative real-time PCR to detect the presence of AURKC message in single human oocytes. Design: Observational Materials and Methods: To determine which variants are present in human oocytes, we designed Taqman probes that are specific for detecting each variant, 1, 2, and 3. To compare the expression levels between meiotic and mitotic cells, we compared the expression of AURKC in oocytes to that of its expression in oocyte-matched cumulus cell samples. Sperm samples known to express all 3 variants were used as positive controls. Results: We found that AURKC expression levels varied from oocyte to oocyte, and that expression levels are greater in the oocyte than sperm, where variants were originally found. Oocytes express all three variants, however, variant 1 is expressed approximately 10 fold greater than variants 2 and 3. Surprisingly, cumulus cells also express AURKC, although this expression was restricted to variant 1. Similar to the single oocytes, AURKC expression levels varied amongst the cumulus cell samples. Conclusion: In sum, like sperm, human oocytes and cumulus cells express AURKC suggesting that missense mutations may also affect female MI and thereby compromise fertility. Future work may involve characterizing whether varying expression levels amongst oocytes correlate with aneuploidy development. Support: ASRM Research Grant.

2808S

Understanding the genetics of spermatogenic failure by resequencing the sex chromosomes of infertile men. *R. George[1], J. Hughes[1], L. Brown[1], L. Lin[2], D. Koboldt[2], K. Meltz-Steinberg[2], R. Fulton[2], R. Wilson[2], R. Oates[3], S. Silber[4], S. Repping[5], D. Page[1,6,7].* 1) Whitehead Institute, Cambridge, MA; 2) The Genome Institute at Washington University, St. Louis, MO; 3) Urology Department, Boston University, Boston, MA; 4) Infertility Center of St. Louis, St. Louis, MO; 5) Academic Medical Center, Amsterdam, The Netherlands; 6) Biology Department, MIT, Cambridge, MA; 7) HHMI.

The sex chromosomes harbor a large number of genes involved in spermatogenesis and are hemizygous in males. For these reasons, mutations that cause spermatogenic failure—the production of very few or no sperm—are disproportionately likely to be found on the X and Y chromosomes. While several genetic causes of spermatogenic failure, such as microdeletions of the Y chromosome and large-scale cytogenetic abnormalities (e.g. Klinefelter's syndrome; XXY), have been identified, they only account for 20-30% of cases and the majority of genetic causes remain unknown. To identify new mutations involved in spermatogenic failure, we have captured and sequenced the coding and conserved non-coding regions of the X and Y chromosomes in 300 men with nonobstructive azoospermia and 300 controls. Our targeted regions include 838 and 53 protein coding genes from the X and Y chromosomes respectively, along with 22 Mb of non-coding sequence that contain ncRNAs and putative regulatory regions. We will describe the initial analysis of these sequences, and the identification of putatively causal coding mutations and copy number variants.

2809S

A rare familial non-Robertsonian translocation involving chromosomes 15 and 21 and failure of reproduction: Is there a correlation? *R. Frikha, F. Turki, S. Daoud, O. Kaabi, R. Louati, O. Trabelsi, T. Rebai, N. Bouayed Abdelmoula.* Histology Laboratory, Univ Medicine, Sfax, Tunisia.

Robertsonian translocations are among the most common balanced structural rearrangements in humans involving two (often non homologous and rarely homologous) acrocentric chromosomes (13, 14, 15, 21 and 22). Nevertheless, non-Robertsonian translocation involving these chromosomes is a rare event and only few cases were reported. Here, we report a familial non-Robertsonian translocation involving chromosomes 15 and 21. The index case is a 38-years-old man for whom chromosomal investigation was carried out to explore a male infertility of 3 years related to azoospermic profile at the semen level, a first unsuccessful attempt of sperm retrieval using TESE and a second diagnostic biopsy of testis showing a sertoli cell only syndrome. Cytogenetic analysis carried out using RHG banding, disclose the presence of a balanced translocation between the long arm of chromosome 15 and the long arm of chromosome 21: t(15;21)(q21;q21). At the genetic counseling, familial history showed that the patient had consanguineous parents , an infertile paternal uncle and two hypofertile brothers with only one son for each of them (as well as another brother and 2 sisters who were fertile). One of hypofertile brother's patient had a history of recurrent pregnancy losses (RPL) (n=3) and a teratozoospermia with a normal count of spermatozoa (98×106/ml). His karyotype revealed at the occasion of familial genetic counseling, was abnormal with the same translocation t(15;21)(q21;q21). Non-Robertsonian translocations, involving chromosomes 15 and 21 are uncommon and were described only three times at literature. However, these translocations involved other break-points and were reported for parents of malformed children harboring partial trisomy 15 [t(15;21)(q13;q22.1)mat], partial trisomy 21 [t(15;21)(q26;q22.1)pat], and Prader willi syndrome [t(15;21)(q15;q22.1)pat]. This new translocation is different by its breakpoints and its association with male infertility and RPL. However, its consequences at the semen level were different for at least the two cases in which the karyotype was done. Molecular delineation using comparative genomic hybridization will be important for phenotype-genotype correlation study which should be considered for genetic counselling and ART.

2810S

Screening Uniparental disomy in recurrent miscarriage couples. *YP. Sun, JW. Xu.* Reproductive Medicine Centre, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China.

Background: Approximately 15-20% of all clinically recognised pregnancies miscarriage, most commonly between 8-12 gestational weeks. The majority genetic factors of early pregnancy losses is attributed to aneuploid abnormalities. However, the aetiology of approximately 40% of early abortions remains unknown. To investigate additional factors causing spontaneous abortions, we screened uniparental disomies (UPD) in cytogenetically normal diploid spontaneous abortion couples. **Materials and methods:**We screened 165 couples who came for assisted reproduction at Reproductive medical center of The First Affiliated Hospital of Zhengzhou University (Henan, China). And all the patients have undergone more than 2 times miscarriage. The karyotype of these couples was 46,XX and 46,XY detected by G-banding technique. Uniparental disomy was detected by illumine HumanCytoSNP-12 DNA Analysis BeadChip and the protocol has been described previously. **Results:**Our results shows that the UPDs mostly occurred on chromosome 6 (7 couples has at least one UPD located chromosome 6)and 11(4 couples has at least one UPD located chromosome 11). The UPDs were between 1M and 15M length, and the most interesting molecular kyrotype was 46,XX,upd(1)(p21.2→p31.1),upd(2)(q33→q34), upd(6)(q11.1→q13),upd(9)(p23→p24.2),upd(9)(q21.13→q21.32), upd(11)(p11.12→q14.1),upd(16)(p12.3→p13.2),upd(19)(q13.2→q13.33), upd(22)(q12.1→q12.3),upd(22)(q13.2), has UPDs locates on chromosome 1, 2,6,9,11,16,19 and 22.We also showed that a patient with about 4M UPD located on chromosome 1 and a 0.01M deletion located on chromosome 8 (46,XY,upd(1)(p32.3→p33)(49215178→53137328)×2,del(8)(q24.12) (121848163→121861465)×1). However, the association between these genes locate in these region and recurrent miscarriage is still unknown. Further studies need to investigate the susceptibility locus and DNA methylation status of these genes. **Conclusion:** Our study suggests that UPD is a potential genetic aetiology of recurrent miscarriage.

2811S

The correlation between Y chromosome partial micro-deletions and recurrent pregnancy loss. *Z. Sarrafi[1], R. Mirfakhraie[1], M. Salimi[1], G. Modabber[2].* 1) Shahid Beheshti University of Medical Sciences, Tehran, Iran; 2) Tadbir Fan Azma Co. , Tehran, Iran.

Introduction: Abortion is defined as preterm labor with no chance of live birth. It is considered as the end of pregnancy before the 20th week of gestation, or a fetus birth weighing less than 500g. Recurrent pregnancy loss, which involves 5% of couples, means 2 or more consecutive pregnancy losses. The studies made so far on the matter have been mainly focused on female causes however, among men, evaluations have only been made on karyotype and chromosomal abnormalities. There are 3 azoo-spermia factor regions on Yq11, called AZFa, AZFb and AZFc, of which AZFc is more than others prone to deletions. In AZFc, there are a few kinds of partial deletions including gr/gr, b1/b3 and b2/b3 of which the most common sub-deletions among infertile persons are gr/gr and b2/b3. It was only in recent years that some studies were conducted to realize the relation between Y chromosome micro-deletions and recurrent pregnancy losses, in which different results were achieved. Among them the relation between partial micro-deletions of the Y chromosome and recurrent pregnancy losses, has been less investigated. Materials and methods: In this study 87 men of couples with idiopathic RPL and 50 healthy fertile men, who had at least one healthy child were examined. After DNA extraction from peripheral blood by salting out and phenol-chloroform methods, multiplex PCR was used to investigate the existence of gr/gr and b2/b3 micro-deletions. Results: No significant difference in the frequency of gr/gr micro-deletions were observed between cases (5 patients, 5/7%) and the controls (1 person, 2%) (p value = 0.282). There was not any significant difference in the frequency of b2/b3 micro-deletions between the cases (2 persons, 2/3%) and the controls too (1 person 2%) (p Value=0/699). Conclusion: gr/gr and b2/b3 deletions have no correlation with recurrent pregnancy losses among Iranian couples. Key words: Recurrent pregnancy loss, Y chromosome, gr/gr deletion, b2/b3 deletion.

2812S

Predictive value of sperm count and motility in the assessment of sperm morphology in infertile men. *S. Daoud, A. Sellami, N. Chakroun-Feki, R. Frikha, N. Bouayed-Abdelmoula, L. Ammar-Keskes, T. Rebai.* Histology-Embryology-Biology of Reproduction Laboratory, Sfax Faculty of Medicine, Sfax, Tunisia.

Introduction: Sperm morphology has an important impact on the success of fertilization. However, it is known that morphology assessment is subjective and highly variable between laboratories and technicians. We thus aimed to determine whether the value of sperm count or motility in routine semen analysis would be predictive of normal morphology. **Patients and Methods :** Data from semen analyses performed in 763 patients undergoing routine fertility evaluation in our laboratory were collected. 689 (90,3%) of them had abnormal morphology. Of the 74 specimens having normal morphology, 61 (8%) had normal count and motility. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were obtained from 2 x 2 contingency tables of sperm count and motility against sperm morphology, and Fisher's exact test was used for the statistic analysis. **Results:** A significantly greater proportion of normal morphology samples exhibited normal count (96% vs 75,6%; p<0,001) and normal motility (86,3% vs 59,3%; p<0,001). A normal count predicted concurrent normal morphology with a NPV of 96% and a sensitivity of 98,2%. Normal motility predicted normal morphology with a NPV of 86,3% and a sensitivity of 96,4%. A significantly greater proportion of normal morphology samples exhibited either normal count or motility (83,6% vs 53%; p<0,001) and both normal count and motility (83,6% vs 53%; p<0,001). The presence of either normal count or normal motility predicted concurrent normal morphology with a NPV of 83,5% and a sensitivity of 96,3%. The presence of both normal count and normal motility predicted concurrent normal morphology with a NPV of 98,6% and a sensitivity of 99%. **Conclusion:** We have found that normal morphology is more likely in the presence of both normal count and motility compared with the presence of normal count, normal motility, or either normal count or motility. This category includes only 8% of specimens in our study population. So, a complete assessment of sperm morphology remains necessary in order to have all the important keys of male infertility diagnosis.

2813S

Nelf affects GnRH migration and secretion in mouse puberty and fertility. *EK. Ko, SD. Quaynor, RC. Cameron, LP. Chorich, LC. Layman.* Institute of Molecular Medicine & Genetics, Georgia Regents University, Augusta, GA.

The hypothalamic-pituitary-gonadal (HPG) axis controlled by gonadotropin releasing hormone (GnRH) plays a crucial role in normal puberty and fertility. Proper specification, migration, and regulation of GnRH neurons are prerequisites for normal GnRH function. When there is impairment of GnRH action, hypogonadotropic hypogonadism results. These patients present with delayed puberty, low sex steroids, and low or inappropriately normal levels of gonadotropins. Sense of smell may be normal, as in normosmic hypogonadotropic hypogonadism (nHH), or impaired as with Kallmann syndrome (KS). Mutations in the NELF gene have been identified in human nHH/KS patients, but the mechanism of how NELF mutations impair puberty is unknown. To address this question, we have generated Nelf knockout (Nelf -/-) mice, and have shown that females manifest delayed puberty, while both males and females have reduced litter sizes. GnRH neurons occupy a smaller region of the brain in adult female Nelf -/- mice compared with Nelf +/+ mice suggesting altered migration. We hypothesized that pubertal aged mice (30 days) will have impaired GnRH neuron migration and manifest hypogonadotropic hypogonadism. Serum follicle stimulating hormone (FSH) and luteinizing hormone (LH) were drawn on all animals as were serum testosterone (males) and estradiol (females). Brains from 6 Nelf -/- and 6 Nelf +/+ mice of each sex (24 total at 30 days of age) were frozen and 10 um-thick coronal sections collected anterior to the optic chiasm through the posterior cerebrum were processed for immunofluorescence microscopy using antibodies directed against GnRH. Immunolabeled sections were examined at 20x using LSC microscopy and GnRH neuron number determined in sections at 40 um intervals by cell counting. Initial serum assays at 30 days demonstrate no differences in testosterone (males) or estradiol (females) between Nelf -/- and Nelf +/+ mice. Our preliminary findings suggest that Nelf -/- mice have normal sex steroid levels at puberty. Examination of pubertal mice for GnRH neuron number and distribution is ongoing, and could indicate changes that may explain the altered patterns of GnRH neurons observed in adults.

2814S

High throughput sequencing of short sequence tags(STS) uncovers novel Y chromosome deletions associated with non-obstructive azoospermia. *X. Liu[1,2], Z. Li[4], Z. Su[1], Y. Gui[3].* 1) BGI, Shenzhen, Guangdong, China; 2) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 3) Peking University Shenzhen Hospital, Shenzhen, Guangdong, China; 4) 2nd people's hospital, Shenzhen, Guangdong, China.

Male infertility, which affects approximately one in ten men among population, is closely associated with Y chromosome deletions, especially in Azoospermia Factor(AZF) region. Traditional PCR based deletion detection method traces a couple of STS markers located in palindromic regions, to check the existence of long range deletions in AZFa, AZFb and AZFc region. This low resolution technique limits our exploration on extensive and novel Y chromosome deletions associated with male infertility. Here we present a novel methodology by sequencing thousands of short sequence tags(STS) distributed across the entire male specific Y chromosome(MSY) in next generation sequencing platform to identify all Y deletions with significantly increased sensitivity and resolution. By exquisite algorithm and validation the accuracy of our method was proved. We applied this approach in 766 Chinese men with non-obstructive azoospermia(NOA) and 683 ethic matched healthy individuals, and identified 481 and 98 STS deleted in NOA and control group respectively, which enormously expanded the previous picture. Overall, we found NOA patient tended to carry significantly more yet rarer deletions. We also discovered a couple of novel deletions in Y chromosome significantly impact functions of spermatogenic genes thus very likely directly resulted in the incidences. Haplogroup O2* seemed to be associated with NOA in Chinese population. In summary, our work reflected a new high-resolution portrait of deletion in Y chromosome, both in NOA patients and normal population.

2815S

Expression of hsa-miR-34b, hsa-miR-181c, hsa-miR-449b, hsa-miR-517c and hsa-miR-605 in FFPE testicular tissues of infertile men with different impairments of spermatogenesis. *D. Plaseska-Karanfilska[1], K. Popovska-Jankovic[1], P. Noveski[1], V. Filipovski[2], K. Kubelka[2].* 1) RCGEB "Georgi D. Efremov", Macedonian Academy of Sciences and Arts, Skopje, Macedonia; 2) Clinical Hospital “Acibadem Sistina”, Skopje, Macedonia.

Using microarray analysis we have previously detected 32 differentially expressed miRNAs in FFPE testicular tissues of infertile men with hypospermatogenesis (Abstract presented at the 5th Florence-Utah Symposium on the Genetics of male infertility, September 2013, Florence, Italy). Among these, hsa-miR-34b, hsa-miR-449b and hsa-miR-517c were one of the most significantly down regulated miRNAs, while hsa-miR-181c and hsa-miR-605 were up-regulated only in the patients with hypospermatogenesis and AZFc deletion. In this study the expression of these five miRNAs was studied using qRT-PCR in a total of 74 infertile men with different impairments of spermatogenesis. Based on the histopathological examination and molecular analysis the studied men were divided in the following groups: normal spermatogenesis (n=18), hypospermatogenesis (n=27), hypospermatogenesis and AZFc deletions (n=3), Sertoli cell only syndrome (SCOS) (n=8), maturation arrest (MA) (n=2), testicular atrophy/ fibrosis/ hyalinisation (n=8) and Klinefelter’s syndrome (n=8). Total RNA was extracted from FFPE testicular tissues using commercial DNA/RNA FFPE tissue kit (Qiagen, Hilden, Germany). RNA quantity and purity were determined using the NanoDrop Spectrophotometer and Agilent Bioanalyzer 2100. MiRNA quantitation was performed by stem-loop RT-PCR followed by Taq-Man PCR analysis using TaqMan MicroRNA Reverse Transcription Kit, TaqMan Universal PCR Master Mix and 7 TaqMan MicroRNA Assays (miR-34b, miR-449b, miR-517c, miR-181c, miR-605, as well as RNU44 and RNU6b as control genes) (Life Technologies, Carlsbad, CA, USA). The expression of each miRNA relative to RNU6b was determined using the $\Delta\Delta$Ct method. Our results showed that hsa-miR-449b and hsa-miR-34b were down-regulated in all studied groups of infertile men; hsa-miR-517c was also down-regulated in all groups, except in the MA group. Hsa-miR-605 and hsa-miR-181c were up-regulated in all, except the hypospermatogenesis group, although the later with a fold change below 2. In conclusion, the qRT-PCR results of the five studied miRNAs in the patients with hypospermatogenesis were consistent with the microarray data. They, furthermore, suggest that the expression pattern of these miRNAs in patients with hypospermatogenesis who also have AZFc deletions is more similar to that of the men with more severe impairments of spermatogenesis, such as SCOS than to the hypospermatogenic patients without AZFc deletion.

2816S

Novel mutations in spermatogenesis genes in azoospermic and severely oligospermic men. *K.A. Fakhro[1], A. Robay[1], J. Rodriguez-Flores[2], A. Al-Shakiki[1], H. Miqdad[3], C. Abi Khalil[1], M. Arafa[3], H. El-Bardisi[3], S. Said[3], J. Mezey[2], R.G. Crystal[1,2].* 1) Genetic Medicine, Weill Cornell Medical College - Qatar, Doha, Qatar; 2) Genetic Medicine, Weill Cornell Medical College - New York, NY; 3) Urology Department, Hamad Medical Corporation, Doha, Qatar.

Primary male infertility - the inability of a healthy male to achieve pregnancy in a fertile female - affects up to 7% of couples worldwide. In some cases, infertility is a result of physical or hormonal causes, often treatable by surgical or pharmacological intervention. However, in cases with no clear environmental cause, a primary genetic defect may be suspected, prompting routine genetic evaluation. In these cases, the infertile men may present with any combination of impaired sperm motility (asthenospermia), abnormal morphology (teratospermia) and/or reduced count (oligospermia), and in the most severe instances, a complete absence of sperm (azoospermia). At the Urology Department at Hamad Medical Corporation in Qatar, over 80% of these patients are classified as having idiopathic infertility, where cytogenetic testing is negative in addition to environmental causes being ruled out. Of these patients, about 10% have a strong family history of infertility (at least one concordant brother), further implicating a genetic basis. We have collected DNA from 18 such families, each comprising a core set of at least two infertile brothers and one fertile brother, plus other available family members. We use exome sequencing to identify rare, severe protein-altering variants segregating with disease in these families and have found novel candidate mutations in all 12 families sequenced to date. In 2 families, we identified novel deleterious mutations in CFTR and DNAI2, genes in which mutations have previously been reported to cause male infertility. In the remaining 10 families, we identified genes not previously implicated in human infertility, including TEX14, TULP3, MED14, SPO11, PDZD7, SPATA21, NOTCH4, ODF3, EEA1 and SYNPO2. All mutations were novel or very rare in all public databases, including >800 ethnically similar controls sequenced on the same platform. Notably, for the 5 latter genes, we observed the same mutations in 14 additional individuals from a cohort of 128 men with sporadic infertility. Preliminary examination of these genes supports involvement in spermatogenesis, with 8 genes found to either have testicular expression in the Human Protein Atlas and/or to cause infertility when knocked out in mouse. Altogether, we present an approach of identifying genes in a severe sporadic disease by focusing on its familial presentation in Qatar, and believe these findings could enhance understanding of pathophysiology and improve treatment options.

2817S

HLA-G gene polymorphisms in Mexican women with recurrent abortions. *A. PORRAS[1,2], A. LAZCANO[3], T. DA SILVA[1,2], C. JUAREZ[1,4], J. JUAREZ[1,2], F. PEREA[1,2], J. GARCIA[2].* 1) DIVISION DE GENETICA, UNIVERSIDAD DE GUADALAJARA, GUADALAJARA, JALISCO, Mexico; 2) DIVISION DE GENETICA, CENTRO DE INVESTIGACION BIOMEDICA DE OCCIDENTE, INSTITUTO MEXICANO DEL SEGURO SOCIAL, GUADALAJARA, JALISCO, MEXICO; 3) SERVICIO DE GINECOOBSTETRICIA, HOSPITAL GENERAL 180, INSTITUTO MEXICANO DEL SEGURO SOCIAL, TLAJOMULCO DE ZUÑIGA, JALISCO, MEXICO; 4) DIVISION DE MEDICINA MOLECULAR, CENTRO DE INVESTIGACION BIOMEDICA DE OCCIDENTE, INSTITUTO MEXICANO DEL SEGURO SOCIAL, GUADALAJARA, JALISCO, MEXICO.

The human leukocyte antigen (HLA)-G is an important immunomodulatory molecule for the maintenance of maternal-fetal relationships because it contributes to the acceptance of semi-allogeneic fetuses. Some single nucleotide polymorphisms (SNPs) in the noncoding regions of the HLA-G gene may influence cellular levels of HLA-G, contributing to pregnancy complications such as preeclampsia or recurrent spontaneous abortions (RSA). In this study, we analyzed the -725C>G (rs1233334), -201G>A (rs1233333) and 14-bp deletion/insertion (14-bp del/ins) (rs66554220) polymorphisms in the HLA -G gene by polymerase chain reaction amplification sequence-specific oligonucleotide probing (PCR-SSOP) and polymerase chain reaction amplification (PCR), respectively in 58 RSA women ($\geq$ 2 miscarriages) without identifiable risk factors and 56 unrelated fertile women ($\geq$ 2 live births). The ages of individuals in both groups ranged from 18 to 42 years. We found no significant differences in the genotype distributions that were analyzed between the RSA women and those who were fertile. In addition, the combinations of SNPs were not in linkage disequilibrium (r2<0.1140). Therefore, this study suggests that these polymorphisms in the HLA-G gene are in linkage equilibrium and do not influence the risk of RSA in Mexican women. Key words: Human leukocyte antigen, recurrent spontaneous abortion.

2818S

Performance and Limitations of Sequenced-Based Cell-Free DNA Aneuploidy Screening: Experience of a Tertiary Referral Center and Importance of Confirmatory Follow-Up Studies. *Y. Liu, W. Neufeld-Kaiser.* Dept Pathology, Univ Washington, Seattle, WA.

Non-invasive prenatal screening (NIPS) using cell free DNA in maternal serum has had a major influence on diagnosis of fetal aneuploidies in the clinical setting. Much higher sensitivity and specificity than maternal serum screening was reported in recently published validation studies. Follow-up diagnostic cytogenetic testing and/or ultrasound evaluation can provide the positive predictive value (PPV) and negative predictive value (NPV) of NIPS that are the more relevant metrics in the clinical setting. We performed a retrospective chart review for 56 abnormal cfDNA screen results from March 2012 through December 2013 in high risk pregnancies. Of 648 consecutive patients referred for non-invasive prenatal screening and genetic evaluation for chromosomal aneuploidy, 5 patients (0.8%) were referred either because of inconclusive results or repeated test failures. There were 586 true negative NIPS results and one false negative NIPS finding, resulting in 99.8% negative predictive value. The false negative NIPS case was trisomy 22 mosaicism diagnosed by amniocentesis and microarray analysis. For a variety of aneuploidies, there were 45 true positive NIPS results and 12 false positive NIPS results, resulting in 78.6% Positive Predict Value. Therefore, NIPS had a sensitivity of 97.8% and a specificity of 98% in this clinical setting. Among the 12 false positive results by NIPS, there were 3/33 trisomy 21, 4/12 trisomy 18, 1/2 trisomy 13, 2/5 monosomy X, and 2/3 XXX. Although NIPS has been shown to be the best non-invasive prenatal screen test to date, it is not a diagnostic test as shown by the false positive rate of 21.4%, an indicator of its diagnostic performance. A number of case reports have been published showing that discordant results can be due to a vanished twin, confined placental mosaicism, low-level maternal mosaicism, complex fetal chromosomal imbalances, and statistical false abnormal. Selected cases with some of these interesting findings will be discussed. Our experience shows that clinicians must recognize the limitations of cfDNA as a screening test. Diagnostic testing via fetal chromosome analysis should always be offered following abnormal NIPS results. In addition, the correlation between the maternal age and gestational age and NIPS test results will be discussed.

2819S

Referral of patients for pre-implantation genetic diagnosis: a survey of clinicians. *K. Barlow-Stewart[1], A. Morrow[1], S. Seeho[2], B. Meiser[3], J. Fleming[1], J. Karatas[4,5].* 1) Northern Clinical School, University of Sydney, Sydney, NSW, Australia; 2) Perinatal Research Group, Kolling Institute of Medical Research, Sydney, NSW Australia; 3) Psychosocial Research Group, Prince of Wales Clinical School, University of New South Wales, Sydney, NSW Australia; 4) Centre for Genetics Education, Royal North Shore Hospital, Sydney, NSW Australia,; 5) IVF Australia, Sydney, NSW, Australia.

BACKGROUND: Pre-implantation genetic diagnosis (PGD) is an assisted reproductive technique, in which embryos are tested for specific genetic abnormalities to enable the selection of those unaffected for implantation and pregnancy. This provides an alternative to prenatal diagnosis and potential pregnancy termination for couples at risk of transmitting a genetic condition to their children. There is recent Australian evidence to suggest that women who are not informed by their obstetricians about the availability of PGD feel disempowered and distressed about not having been provided this option. This study aimed to explore obstetrician knowledge regarding PGD and to identify potential barriers to referral for PGD. METHODS: An online questionnaire was e-mailed to fellows and accredited trainees of the Royal Australian and New Zealand College of Obstetricians and Gynaecologists. The questionnaire assessed respondents' knowledge of PGD and barriers to referral. RESULTS: Of 372 practicing obstetricians who responded, 315 were fellows and 57 trainees. Obstetricians' perceptions of their patients' financial status and ability to access PGD services were identified as the main barriers to referral. There was variability in obstetrician discussion of referral for PGD and/or to genetic services according to the presenting scenario. Variability was also observed regarding perceived appropriateness of PGD for different indications. Obstetricians who had received training or professional development related to PGD were more likely to discuss the option of PGD than those who had not ($\chi 2 = 6.44$; $p < 0.01$). CONCLUSIONS: Results from this study highlight the need for training opportunities, educational resources and recommendations to guide health professionals and ensure that couples eligible for PGD are informed and appropriate referral pathways are in place.

2820S

Four years later: the state of PGD in Quebec. Professional perspectives on medical indications and regulations. *F. Duplain-Laferrière[1], R. Drouin[2], C. Bouffard[1].* 1) Division of Genetics, Department of Pediatrics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Quebec, Canada; 2) CHR du Grand-Portage, Department of Obstetrics-Gynecology, Quebec, Canada.

For four years now, the Quebec healthcare system has covered costs related to preimplantation genetic diagnosis (PGD), for couples at risk of having a child affected by a serious genetic disease. This minimal and unclear regulation, both concerning medical indications and practices, is leading to significant decision-making, organizational and ethical challenges for health professionals. Objectives: To better understand the situation, we have consulted Quebec geneticists, obstetricians-gynecologists and genetic counselors to learn about their perceptions on: 1) the medicals and social reasons that justify PGD and 2) the principal players that should be involved in the regulation of PGD. Methodology: Qualitative Research Design — Online questionnaire containing 34 questions (13 open, 11 semi-open, 10 closed-ended) has been completed by 15/30 obstetricians, 15/30 geneticists and 17/30 genetic counselors engaged in activities related to prenatal diagnosis in the province of Quebec. Data analysis: General inductive and thematic analysis. Results: Overall, the participants agree that the use of PGD for multifactorial illnesses or for social reasons should be prohibited. For them, PGD should be restricted to fatal or incapacitating illnesses. However, professionals lack consensus on certain diseases or syndromes. Regarding regulation, the majority of participants want PGD to be regulated while allowing them the flexibility to adjust their clinical management or medical counselling to particular cases. While they consider that the patients should be involved in the decision-making process, they do not believe that patients should decide everything. Conclusion: This project has made it possible to develop novel knowledge through the comparison of the perceptions of geneticists, obstetricians and genetic counselors in Quebec. In light of these results, it is necessary to find new approaches to understand the concepts of serious genetic diseases and the sharing of responsibility, to build guidelines that are more adapted to social and professional contexts.

2821S

4-Hour Concurrent Preimplantation Genetic Diagnosis of 24-Chromosomes Aneuplody, Single Gene Disorders, And Micro-Deletion and Duplications. *C. Jalas[1], X. Tao[2], A. Fedick[2], C. Bohrer[2], S. Kloskowski[1], D. Gabriele[2], B. Levy[3], R.T. Scott[2], N.R. Treff[2].* 1) Foundation for the Assessment and Enhancement of Embryonic Competence , Basking Ridge, NJ; 2) Reproductive Medicine Associates of New Jersey, Basking Ridge, NJ; 3) Columbia University, NY.

Combining accurate preimplantation comprehensive chromosome screening (CCS) with single gene disorder (SGD) or insertion/deletion (indel) diagnosis from the same biopsy has presented many challenges. Current methods can either fail to detect all origins of aneuploidy, require excessive workup times, involve long PGD turn-around times, or are expensive to perform. The present study develops a method which overcomes each of these limitations using quantitative real-time (q)PCR. Workups involved identifying informative SNPs in the parents using SNP arrays and phasing the markers using qPCR on family members. Blastocyst biopsies underwent targeted multiplex PCR of several loci followed by qPCR for SGD, indel, contamination, and CCS analyses. For insertions, a linkage only approach was used. For deletions, multiple informative markers within the deletion were used along with linked markers outside the deletion. SGD results were compared for consistency and reliability to conventional diagnoses which were obtained from a reference lab. Additional linked informative markers were evaluated to resolve discordant cases. The typical workup involved approximately 4 weeks to complete. 17 SGD cases, including autosomal and X-linked, recessive and dominant, and compound heterozygosity disorders, provided 150 embryos for analysis. 150 (100%) gave a 4-hour diagnosis by SNP qPCR, while 139 (92.7%) gave a result by conventional STR analysis. Concordance in diagnosis was 97%, with 4 discrepancies. Subsequent analysis of additional markers confirmed the SNP qPCR diagnosis in all cases. The use of SNPs rather than STRs generally resulted in having markers nearer the mutation, enhancing the ability to avoid recombination-based misinterpretation or a failure to make a diagnosis in many cases. 14 indel cases were processed including a total of 157 embryos, all of which gave an indel and CCS diagnosis. 7 patients have had embryo transfers from which 5 have delivered, 1 has an ongoing pregnancy, and 1 did not become pregnant. In 2 cases, DNA from the newborns was available for analysis and confirmed the PGD and CCS diagnoses. This new approach to combined CCS and SGD or indel PGD involves a short and inexpensive workup and the ability to reliably and rapidly produce accurate PGD results in parallel with CCS from the same biopsy.

2822S

The frequency, type and classification of chromosome errors differs at the pre-implantation stage from that observed during pregnancy. *M. Schweitz, S. McReynolds, W.B. Schoolcraft, M.G. Katz-Jaffe.* FLC, Lone Tree, CO.

Women at 40 years of age have a 50% reduction in fecundity compared to women a decade younger, primarily due to progressive oocyte depletion and an increase in meiotic errors resulting in aneuploidy. It is well documented that oocyte aneuploidy is a major contributor to spontaneous miscarriage. The aim of this study was to evaluate the frequency, type and classification of chromosome errors in human pre-implantation embryos relative to maternal age. Infertility patients consented, under IRB approval, to an in vitro fertilization cycle with comprehensive chromosome screening (CCS). Embryos were cultured to the blastocyst stage for a trophectoderm (TE) biopsy (n=15,750 blastocysts). Biopsied TE cells were examined for chromosome numeration using quantitative PCR (RMA-NJ). Statistical analysis involved Chi square test with significance at $P<0.05$. Blastocyst CCS results revealed all 22 autosomes and both sex chromosomes were involved in chromosome gains and losses. Aneuploidy was identified in 50.2% (7,905) of these human blastocysts, with close to even numbers of trisomies (51.1%) and monosomies (48.9%) that are not typically observed in pregnancy. Predictably, for younger infertility patients (<38 years) a lower incidence of aneuploidy was observed (35.9%), which was significantly less than for woman of advanced maternal age (AMA; ≥38 = 63.1%; $P<0.0001$). Interestingly, specific chromosome errors were associated with AMA, chromosomes 14, 15, 16, 17, 18, 20, 21 & 22 were more likely (average 2.3 fold) to be aneuploid in human blastocysts from AMA infertility patients ($P<0.05$). In addition, the classification of these chromosome errors was significantly different relative to maternal age. Blastocysts from young infertility patients showed a significant increase in errors involving the large metacentric and submetacentric chromosomes (1-5), which are not typically observed in clinical pregnancy losses (<38=16.5% vs. ≥38=13.5%; $P<0.01$). The frequency, type and classification of chromosome errors in pre-implantation human blastocysts were significantly different in relation to maternal age. In addition, the incidence of aneuploidy in pre-implantation human blastocysts has a different profile to aneuploidy observed in first trimester losses or live births. These results provide valuable information for clinical patient management prior to infertility treatment.

2823S

A Simple and Streamlined Next-Generation Sequencing-based approach to Preimplantation Genetic Screening. *M. Umbarger, J. Gole, A. Gore, G. Porreca.* Good Start Genetics, Cambridge, MA.

Preimplantation genetic screening (PGS) is used to assess the chromosome copy number of embryos. Although increasing evidence indicates that euploid embryo transfer increases and decreases implantation and miscarriage rates, respectively, PGS adoption has been limited at least in part due to the high cost associated with traditional PGS approaches. However, increased use of trophectoderm biopsy followed by vitrification and subsequent frozen embryo transfer, coupled with streamlined workflows employing next-generation DNA sequencing (NGS), are poised to enable broader PGS adoption.

We have developed and implemented an automated PCR-based method that amplifies regions from each chromosome and simultaneously attaches the sequencing adapters and sample-specific barcodes necessary for multiplexed NGS. 12 pg DNA purified from cell lines (~2 diploid cells) or lysate derived from 2-cell isolated cultured lymphocytes served as template for the PCR reactions. The products were sequenced to generate count data for each chromosome, and this data was subsequently used to infer chromosome copy number.

A total of 37 true positive aneuploid chromosome calls were made across the DNA from 21 aneuploid cell lines. The method generated 789 correct diploid chromosome calls, 2 incorrect aneuploid (false positive) chromosome calls, and zero incorrect diploid (false negative) chromosome calls. Both incorrect aneuploid calls were in samples containing other aneuploid chromosomes, thus yielding perfect sample-level specificity and perfect chromosome-level sensitivity. Aneuploidies detected included: trisomy 2, 8, 9, 13, 18, 20, 21, 22, 16+21, 2+21, monosomy X, tetrasomy X, XXY, and disomy Y. The technique also detected trisomy 21 and XXY when lymphocytes were used as a template. Collectively, our results indicate that we have developed a simple and accurate NGS-based PGS approach.

2824S

Preimplantation genetic risk reduction (PGR) - a new concept in the era of microarray CGH and exome sequencing. *G. Altarescu[1], R. Beeri[1], T. Eldar-Geva[2], G. Lazer-Derbeko[1], E. Levy Lahad[1], P. Renbaum[1].* 1) Zohar PGD Lab, Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel; 2) In vitro Fertilization Unit, Shaare Zedek Medical Center, Jerusalem, Israel.

While the ESHRE guidelines of best practice standardizes the genetic and IVF workup prior to and during the PGD cycle, new technologies such as Chromosomal Microarray (CMA) and exome or genome sequencing reveal variants of unknown significance (VOUS). These variants raise challenges in the decision making process in respect to the indications calling for preimplantation genetic diagnosis. Results: We present two couples that came to our PGD unit at the Shaare Zedek Medical Center requesting preimplantation genetic diagnosis for VOUS. In the first couple, a healthy male, carrier of the variant V51D in the MLH1 gene requested PGD as risk reduction for lynch syndrome. His family history included his father who died of colon cancer (no DNA sample), paternal grandmother died of gastric cancer (no DNA sample), a paternal uncle with colon polyps carries the familial variant and another paternal uncle who also underwent polypectomy, but is not a carrier of this variant. While the MLH1 V51D has not been reported as pathogenic it is predicted by Polyphen to be probably damaging (score 0.96), by SIFT to be damaging (score 0)and by SNPs@Go to be a polymorphism. Despite the uncertainty of the pathogeneity of this variant as the cause of multiple cancers in this family, the couple, who do not require IVF for infertility, requested PGD to possibly reduce the risk of cancer. The second couple came to our PGD unit with one child diagnosed with PDD. Results from CMA revealed a 570,000 bp microduplication on chr:X p.22.3, and a 637,000 bp microdeletion on chr:17q21.31) in the affected child. The X linked microduplication was maternaly inherited and also present in the couple's healthy daughter, while the microdeletion was "de novo". Both microduplications and microdeletions in these regions have been reported to be associated with intellectual disability and congenital malformations. The couple has requested PGD for both findings. Conclusions: Chromosomal microarrays and exome findings of uncertain significance pose dilemmas for both couples and geneticists in regard to counseling and performing PGD. As CMA and exomes are becoming standard tests, couples are requesting PGD for VOUS to reduce their risk of possible genetic disease. These technologies demand urgent discussion and guidelines for preimplantation risk reduction.

2825S

Fetal intracerebral hemorrhage and cataract: think *COL4A1*. *E. COLIN[1], A. DENOMME-PICHON[1], A. GUICHET[1], M. GORCE[1], S. CHEVALIER[2], M. MINE[3], E. TOURNIER-LASSERVE[3], L. SENTILHES[4], D. BONNEAU[1].* 1) Department of Biochemistry and Genetics, CHU Angers, Angers, France; 2) Department of Ophthalmology, CHU Tours, Tours, France; 3) Reference Center for Rare Vascular Diseases of the Eye and the Brain, Lariboisière Hospital, Paris, France; 4) Department of Obstetrics and Gynecology, CHU Angers, Angers, France.

The *COL4A1* gene (MIM 120130) encodes the alpha1 chain of type IV collagen, a crucial component of nearly all basement membranes. Mutations in *COL4A1* were first associated with cerebral microangiopathy and familial porencephaly and have later been implicated in a clinicopathologic broad-spectrum affecting the brain, eyes, kidneys and muscles. Recently, *COL4A1* mutations have also been identified prenatally in fetuses with intracranial hemorrhage (ICH). We report two additional prenatal cases of *COL4A1* mutations in fetuses with ICH and cataract.

Case 1: Fetal ultrasound examination (US) at 23 weeks' gestation (WG) showed left cataract, left ventriculomegaly and hyperechogenic lesion of basal ganglia. Fetal magnetic resonance imaging (MRI) at 32 WG confirmed the subependymal hemorrhage affecting the left hemisphere.

Case 2: Fetal US examination at 31 WG showed hyperechogenic lesion in left hemisphere with thalamic echogenicity and bilateral cataract. Fetal MRI at 32 WG showed a left-sided periventricular parenchymal hemorrhage and mild ventriculomegaly.

In both these cases, the involvement of *COL4A1* was evoked because congenital cataracts had been previously reported in association with ICH in pediatric cases.

The sequencing of *COL4A1*, performed on fetal DNA after termination of pregnancy, evidenced two heterozygous novel missense mutations c.G2317A (p.G773R) and c.G3005A (p.G1002N) in fetuses 1 and 2 respectively.

The two cases reported here show that the *COL4A1* mutation should be envisaged in fetuses with prenatal ICH especially in the presence of lens abnormalities at US examination. Molecular confirmation of a *COL4A1* mutation may have important implications for the outcome of the pregnancy and for genetic counseling.

2826S

The Israeli experience of the first 300 Panorama™ tests that use 19,488 single nucleotide polymorphisms (SNPs) followed by high-throughput sequencing for common trisomies risk assessment. *H.N. Baris Feldman[1], Z. Weiner[2], I. Solt[2], M. Shohat[3], D.M. Behar[1].* 1) The Genetics Institute, Rambam Health Care Campus, Haifa, Israel; 2) OBGYN department, Rambam Health Campus, Haifa, Israel; 3) The Recanati Genetic Institute, Rabin Medical Center, Petach Tikva, Israel.

Background Cell free DNA (cfDNA) has emerged over the last year as an alternative for amniocenthesis for diagnosis of the common aneuploidies looking at trisomy 21, 13, 18, sex chromosomes and triploidy. Methods We present our experience of the first 300 Panorama™ tests sent from Israel. This method is based on massively multiplexed PCR amplification of cfDNA isolated from maternal plasma, targeting 19,488 SNPs, followed by high-throughput sequencing. The fetal fraction is determined. The SNP pattern of maternal DNA (from buffy coat) is compared to the SNP pattern of free DNA from maternal plasma, which contains maternal and fetal DNA. Paternal genomic samples, when available, were included in the analysis; in the absence of a paternal sample, the algorithm considers population allele frequencies. Combining the maximum likelihood ratio with a priori risk generates a risk score. Results The results of the first 300 sequential tests performed in Israel were analyzed. Fifteen samples necessitated redraw, two samples failed analysis. Four samples yielded high risk scores: two cases for trisomy 21, one for Kleifelter syndrome (KS) (47,XXY) and one for trisomy 18. Confirmation of both trisomy 21 and one KS were done by CVS or amniocenthesis. The mother of suspected trisomy 18 was not interested in invasive testing in view of normal ultrasound scans and delivered a healthy baby. Karyotype was not done so mosaic state was not excluded. There are no known false negative results. Discussion Panorama™ test is a reliable tool for identification of pregnancies at high risk for fetuses with the common aneuploidies with a high success rate. We recommend confirmation of the diagnosis for high risk scores pregnancies using invasive tests.

2827S

First trimester trisomy 18 screening using fetal epigenetic marker and nuchal translucency. *D.E. Lee[1], S.Y. Kim[1], S.Y. Park[1], J.W. Kim[1], D.J. Kim[1], D.W. Kwak[2], H.M. Ryu[1,2].* 1) Laboratory of Medical Genetics, Cheil General Hospital and Women's Healthcare Center, Seoul, Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women's Healthcare Center, Kwandong University College of Medicine, Seoul, Korea.

Objective: The association between fetal trisomy 18 and increased nuchal translucency (NT) thickness during the first trimester of pregnancy is well-established. Recently, the placental-derived maspin (U-maspin) gene was applied as a fetal-specific epigenetic marker for noninvasive detection of trisomy 18 in maternal plasma. This study assessed the accuracy of prenatal screening for trisomy 18 by a combination of U-maspin concentration and fetal NT thickness in the first trimester of pregnancy. Methods: A nested case-control study was conducted using maternal plasma samples collected from 65 pregnant women carrying 11 trisomy 18 and 54 normal fetuses. Using real-time quantitative methylation-specific PCR, the concentrations of U-maspin were measured in first trimester maternal plasma. Fetal NT measurement was performed in the first trimester of pregnancy. Results: U-maspin concentrations were significantly elevated in women with trisomy 18 fetuses compared with controls (27.2 vs 6.6 copies/mL; P<0.001). Fetal NT thickness were significantly elevated in women with trisomy 18 fetuses compared with controls (5.9 vs 2.0 mm; P<0.001).The sensitivities of U-maspin concentration and NT thickness for prenatal screening of fetal trisomy 18 were 90.9% and 90.9%, respectively, with a specificity of 98.1%. Combined analysis of U-maspin concentration and NT thickness had the sensitivity of 100% for prenatal screeningof fetal trisomy 18, with a specificity of 98.1%. Conclusions: The combination of U-maspin concentration and NT thickness is highly efficient for prenatal screening of fetal trisomy 18 in the first trimester of pregnancy.

2828S

Prenatal array CGH and follow up of fetuses with increased nuchal translucency: results from VUmc. *K.E. Stuurman[1,2], S.L. Bhola[1], M.A.J. Engels[1], E.J. Meijers-Heijboer[1].* 1) Clinical Genetics, VU University Medical Center, Amsterdam, Netherlands; 2) Medical Genetics, University of British Columbia, Vancouver BC, Canada, Women's Hospital & Health Centre Site.

Introduction: Increased nuchal translucency (NT) (≥3.5mm) in pregnancy is associated with chromosomal aneuploidy, genetic syndromes and structural defects. Prenatal array CGH for all fetal anomalies increases the percentage of genetic abnormalities with 2-12% compared to standard karyotyping. We assessed the outcome of prenatal array CGH in a consecutive series of fetuses with increased NT only seen in our clinic. We also followed up this cohort postnatally with respect to clinical outcome. Methods and Materials: Included were all women pregnant between January 2011 and August 2013 with a fetus with increased NT, but without additional abnormalities seen on ultrasound. We offered array CGH (Agilent 180k oligoarray) subsequent to QF-PCR and karyotyping when enough fetal material was present. Results: In total 155 fetuses had increased NT only. 124 had invasive prenatal testing (80%), of which 65 had normal QF-PCR results and karyotype (52%). Of these 65 fetuses, 37 received subsequent prenatal array CGH (60%). Two had an abnormal result (11 Mb duplication on chromosome 10q25.1-q26 and 8 Mb duplication on chromosome 2p25)(5%) and these pregnancies were terminated. In total 48 of 65 fetuses, with and without prenatal array CGH, were followed up in further pregnancy and postnatally (75%). Of these 48 fetuses, five had termination of pregnancy (10%); three because of hydrops foetalis (one was diagnosed with RyR1-related congenital myopathy postnatally), two because of low trust in a good outcome. Two resulted in IUFD (4%) and one had postnatal death (2%). Four syndromes and structural defects were diagnosed postnatally (Noonan syndrome, Beckwith Wiedemann syndrome, RyR1-related congenital myopathy and aniridia)(8%), and three had undiagnosed psychomotor retardation (6%). Conclusion: Prenatal array in fetuses with increased NT only increases abnormal findings with 5%. However, more than 25% of the fetuses had an unfavorable outcome. The abnormalities seen postnatally could not have been picked up with array CGH.

2829S

Non Invasive Prenatal Testing and Prevention of chromosomal and Genetic Disorders. *A. Al-Aqeel.* Dapartment of Pediatrics, Prince Sultan Military Medical City, Riyadh, Saudi Arabia.

Background: Noninvasive prenatal testing (NIPT) that will allow genetic testing of a fetus within the first trimester of pregnancy by isolating cell-free fetal DNA (cffDNA) in the mother's plasma raises a range of ethical and legal issues. The goal of this study is to provide an Islamic ethical framework for health care providers and government agencies providing NIPT. Methods: We refer to our previous experience in medical genetics, screening tests and ethics in combination with the Islamic "Shariai'ha" (Figh) principles and authoritative "consensus edicts" "fatwas" of Islamic scholars, literature review and our publications. We developed a set of best practices for the provision of NIPT within an Islamic framework. Results: Applying the Islamic "Shariai'ha" principle "The basic concept in useful matters is permissiveness" which indicates that everything is lawful, as long as it is useful to people, our principal recommendations include promotion of NIPT to high risk pregnant women for the prevention of fetal aneuploidy and in certain cases at high risk of single gene disorders, with the amendment of current informed consent procedures to include attention to the noninvasive nature of this new testing and the potential for a broader range of results earlier in the pregnancy. However, the need for confirmatory testing by amniocentesis must be discussed carefully with the pregnant woman, as abortion in Islam should be done before 120 days of gestation in severe malformations or devastating disorders which are incompatible with life. Conclusions: Pregnant women at increased risk of aneuploidy can be offered cffDNA testing. Its performance in low-risk women and women with multiple gestations is unclear. Such test should be regulated by government agencies. Since limited professional guidance is available clinician performing the test should adopt responsible best practices in the provision of the test within an ethical framework that combines appeal to written precedent with sensitivity to the options of individuals and families dealing with choices and necessities within the laws, norms and traditions of their society.

2830S

Susceptibility loci for neurodevelopmental disorders -prenatal genetic counseling and psychological impact. *K.E.M. Diderich[1], L.C.P. Govaerts[1], J. Verhagen-Visser[1], S.L. van der Steen[1], M. Joosten[1], M.F.C.M. Knapen[2,3], F.A.T. De Vries[1], D. Van Opstal[1], M.I. Srebniak[1], S.R. Riedijk[1], R.J.H. Galjaard[1].* 1) Clinical Genetics, Erasmus MC, Rotterdam, Netherlands; 2) Dept. of Obstetrics and Prenatal Medicine, Erasmus Medical Centre, Rotterdam, the Netherlands; 3) Foundation of Prenatal Screening South-west Netherlands, the Netherlands.

Objectives: Since 2012 genomic SNP array is used as a first-tier prenatal cytogenetic test for all indications in our laboratory. Next to causative and unexpected diagnoses, array may also detect susceptibility loci (SL) for neurodevelopmental disorders, with a yet unquantifiable risk for the fetus. While some may argue that we should not disclose such uncertain results, potentially leading to unnecessary anxiety, the fact that an abnormal phenotype has been described in association with these SL, may be a reason to disclose SL. We classify SL as pathogenic. With this study we evaluated the effect of reporting and counseling SL and explored the psychological impact of hearing about a SL during pregnancy. Methods: To assess the psychological impact of releasing SL 9 couples with a child that was diagnosed with a SL during pregnancy were approached for an interview. From these, 8 women and 4 men were interviewed by phone 3-18 months after the prenatal diagnosis. During their pregnancy, they all received pre-test counseling. The post-test genetic counseling concentrated on the phenotype of the particular SL, its incidence in the normal and affected population and the difference between postnatal and prenatal ascertainment. Targeted parental array testing was offered. Extensive US examination was offered when the SL was associated with physical abnormalities. Results: Hearing about the SL during pregnancy was initially shocking to 8 and concerning for 4 parents. However, at the time of the interview, eleven parents had no worries anymore about the SL while one mother considered it a stigma. All parents considered their child healthy without concern about the SL. They stressed the importance of pre- and posttest counseling and expressed their wish for a choice regarding SL disclosure during pre-test counseling. Eleven parents indicated they wished to learn about a SL again in a next pregnancy. One mother did not know whether she wanted to learn about a SL again. Conclusions: Although our group is small, it provides a preliminary insight into how couples experienced hearing about a SL as the result of prenatal diagnosis. From these preliminary results it seems that couples are able to cope with the uncertainties associated with this type of chromosome aberration. Further studies have to be carried out in order to investigate the clinical significance of SL when found prenatally and the (dis)advantages of their disclosure in a prenatal setting.

2831S

Placenta whispers: Discordant noninvasive prenatal testing (NIPT) results and the role that confined placental mosaicism (CPM) plays. *T. Boomer[1], S. Cherny[2], J. Miles[3], C. Pittore[4], J. Wardrop[1], J. Jesiolowski[5], N. Teed[1], C. Settler[1], N. Dharajiya[1], T. Monroe[5], J.-S. Saldivar[1].* 1) Sequenom Laboratories, San Diego, CA; 2) Central Dupage Hospital Maternal Fetal Medicine, Winfield, IL; 3) Medical University of South Carolina, Charleston, SC; 4) Capitol Health Maternal Fetal Medicine, Pennington, NJ; 5) Sequenom Laboratories, Morrisville, NC.

Background: Noninvasive prenatal testing (NIPT) screening for aneuploidy relies on the presence of circulating cell-free DNA believed to be largely placental in origin. The genetic material in fetal and placental tissue matches in most pregnancies. However, discordance between these tissues can occur due to post zygotic non-disjunction or trisomy rescue, causing uneven distribution of cells between fetus and placenta. A comprehensive study of three confirmed cases of confined placental mosaicism (CPM) identified by positive NIPT are highlighted, detailing invasive diagnostic results, postpartum placental studies, and pregnancy management considerations. Case 1: NIPT performed at 14.6 weeks gestation indicated trisomy 13 in a sample sent for advanced maternal age. Amniocentesis and blood studies at birth were normal. The fetus was small for gestational age in late pregnancy. Postdelivery, chromosome 13 FISH testing revealed three out of four placental sections as mosaic for trisomy 13: 81% (162/200), 34% (67/200), and 76% (152/200). Case 2: NIPT performed at 10 weeks gestation indicated trisomy 13 in a sample sent for advanced maternal age. FISH on chorionic villus sampling (CVS) revealed trisomy 13 in 34/200 cells (17%), though cultured CVS results showed a normal karyotype. Amniocentesis results were normal by FISH (45 cells) and karyotype (37 cells). The pregnancy is ongoing with normal fetal growth. Postnatal studies are pending. Case 3: NIPT performed at 18 weeks gestation indicated trisomy 18 in a case referred for positive biochemical screening (1/11 risk for Down syndrome). Amniocentesis and blood studies at birth were normal. Post-delivery FISH testing on cultured placental tissue revealed 3.2% (8/250 cells) mosaicism for trisomy 18. Conclusions: These cases describe three pregnancies with positive NIPT results, negative amniocentesis results, and confirmed CPM. Postpartum analysis of two placentas and CVS analysis of a third demonstrate a wide spectrum of CPM levels. NIPT provides new and unique insight into the placenta as a whole and may reveal evidence of CPM that would otherwise be unseen or minimized by other test methods. CPM may also be associated with placental dysfunction and present clinically as abnormal biochemical screening, fetal growth restriction, or preterm labor. Cases with discordant 'false positive' NIPT results may warrant more conservative pregnancy management.

2832S

Transcriptome expression analysis of amniotic fluid cell free fetal RNA according to gestational weeks in Korean women. *Y. Jung, S. Shim, S. Shim, S. Sung, J. Park, D. Cha.* CHA University, Seoul, Rep of Korea, Seoul, South Korea.

Objective: To characterize the transcriptome expression patterns and biological pathways in amniotic fluid cell free fetal RNA according to gestational weeks in Korean women. Method: The discarded amniotic fluids of Korean women in terms of gestational weeks were prospectively collected from euploid fetuses for this study. Total RNA was extracted from cell free amniotic fluid supernatant and hybridized to Affymetrix GeneChip Human Arrays. Differentially expressed transcripts among 16-18 gestational weeks, 25-26 weeks and 37-38 weeks amniotic fluids were obtained by The Welch's t-test. Hierachial clustering analyses were performed to visualize overall expression characteristics of all samples used for the study. The biological functions of selected genes were analyzed using various online Gene Ontology databases. Results: There were 2902, 6035 and 6289 genes significantly expressed in early second, late second and third trimesters, respectively. Hierachial clustering showed differential transcriptome expression pattern according to gestational weeks. Comparing with the early second trimester, fetus related specific genes were more highly enriched in late second and third trimesters (471 vs 913 and 949, respectively). Gene expression analysis showed enrichment of brain transcripts in the late second trimester as compared with early second and third trimesters. Only 33 genes were differentially expressed in the first trimesters and pathway analysis revealed that those are related to cell proliferation and apoptosis pathway. The transcripts that are related to fetal brain were enriched in late second trimester. Conclusion: This study provides information regarding gene expression changes during normal fetal development in Korean women. Brain specific transcripts are enriched in the late second trimester rather than early second trimester. This data suggests that brain development occurs in late second trimester.

2833S

Fragile X prenatal studies suggest mothers with >80 repeats transmit the normal allele in 55% of pregnancies. *S.L. Nolin, A. Glicksman, X. Ding, N. Ersalesi, W.T. Brown, C. Dobkin.* New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY.

We have performed fragile X prenatal studies for 1,871 pregnancies of women with intermediate, premutation or full mutation alleles from 1991-2013. The studies included 1521 chorionic villus and 338 amniotic fluid samples. The expanded allele was transmitted in 955 (51%) pregnancies with expansion to a full mutation in 308 (32%). Unexpectedly, mothers with large premutation (>80 repeats) or full mutation alleles transmitted the normal allele in 55% (368/671) of pregnancies while mothers with smaller alleles (45-79 repeats) transmitted the normal allele in 46% (548/1200) of pregnancies (p=.00014, Chi square 1df). These surprising results suggest the presence of intermediate and small premutation alleles may provide some early developmental advantage. For larger alleles, some full mutation embryos may be non-viable from loss of the X chromosome carrying the full mutation. This has already been observed in an increased risk for mosaic Turner syndrome in females (Dobkin et al., AJMG, 2009). Early loss of an X in males and females may lead to skewed distributions of the normal versus the fragile X chromosome.

2834S

Differences of transcriptional profiling analyses between cell free mRNA and mRNA originated from amniocytes in amniotic fluid using GeneChip® PrimeView™ Human Gene Expression Array. *D.H. Cha[1,2], S.S. SHIM[1], S.H. SHIM[2], Y.W. Jung[1], S.R. Sung[2].* 1) Dept OB/GYN, Kangnam CHA Hosp, CHA University, Seoul, South Korea; 2) Genetic Research Center, Kangnam CHA Hosp, CHA University, Seoul, South Korea.

Amniotic fluid is the only body fluid in direct contact with the the fetal oropharynx, lungs, gastrointestinal tract, skin, and urinary system and amniotic fluid cell-free fetal RNA(AF cffRNA) can provide biological information on developing fetal organ system. The aim of this study is to compare RNA profile between amniocytes and cell free fetal RNA in amniotic fluid supernatant. The total 10 AF samples were collected each supernatants and cell pellets. The cell free fetal RNA was extracted from 5-10ml of amniotic fluid supernatant. All supernatant samples were performed using the QIAamp® Circulating Nucleic Acid (Qiagen, germany) kit with an on-column DNase digestion step to remove genomic DNA according to the manufacturer's instructions. The RNA was purified with the RNeasy® MinElute® Cleanup kit (Qiagen, germany). The collected amniocytes were cultured in BIO-AMP-2 complete medium (Biological Industries Ltd., Haemek, Israel) at 37°C in a 5% CO2 incubator. The cultured amniocytes were extracted using QIAamp RNeasy ® mini kit. The extracted RNA was hybridized to GeneChip® PrimeView™ Human Gene Expression Array. The data were analyzed with Robust Multi-array Analysis using Affymetrix default analysis settings and global scaling as normalization method. The normalized, and log transformed intensity values were then analyzed using GeneSpring 12.5. We were identified 2902 transcripts in amniotic fluid supernatant and 15,418 transcripts in the cultured cell. We found 403 genes that were differentially expressed in amniotic fluid supernatant compared with cultured cell. Among these genes, we were identified 123 genes that have been known to tissue-specific transcripts through a search of pathway analysis software and Medline. Also, the 63 genes in amniotic fluid supernatant were expressed in brain and the 60 genes were confirmed expressed in various tissues, such as heart, liver, kidney, lung, eye, etc. The gene expression studies for the development of the fetus have been limited by practical problems, such as the collection of the sample. Most studies of the fetus have obtained from tissue specimens after spontaneous abortions, terminations of pregnancy. Our results demonstrate that amniotic fluid cffRNA is reflected in real-time to the development of the fetus. This study might contribute to the further studies for the discovery of biomarkers for abnormal fetal growth.

2835S

Methods for Isolation and Enrichment of circulating cell-free fetal DNA (ccff DNA) from maternal plasma, for Non-Invasive Prenatal Tests (NIPT), such as the MaterniT21™ PLUS Laboratory Developed Test. *G. DeSantis, L. Chamberlain, A. Kulkarni, D. Wong, N. Agarwal, E. Kaltgrad, P. Whitley, D. van den Boom, M. Ehrich.* R&D, Sequenom, San Diego, CA.

The use of ccf DNA has enabled the introduction of non-invasive prenatal testing (NIPT) methods such as fetal RHD genotyping and fetal aneuploidy detection from maternal plasma. Methods to process and extract ccf DNA from maternal blood that may enrich fetal fraction are desired to facilitate enhanced performance and decreased costs of NIPT. We aimed to identify key physical or biochemical markers that distinguish the nature and packaging of fetal versus maternal DNA in plasma that could be used to enrich fetal fraction. Several collection and processing conditions along with several methods for DNA extraction, including both manual and automated platforms, were evaluated. The fetal DNA, total DNA and fetal fraction for each method as well as purity for downstream assays and applications (including fetal copy and fetal fraction analysis, library preparation, and next generation sequencing) are compared. Methods explored to enhance fetal fraction include: processing changes, centrifugation protocols, digestion with endonucleases, characterization of micro-particle compartments, and bead based strategies. Fetal fraction assessments are made by various approaches including qPCR, methylation sensitive restriction digestion, mass array analysis and sequencing based fetal fraction determinations. The pre-analytical technologies explored affect the size distribution profile of isolated circulating cell free DNA (ccf DNA) and fetal fraction. Distribution of, and enrichment of, various DNA fragment sizes are discussed as well as mechanisms for enrichment. Micro-particle fractions are correlated to fetal and maternal DNA distribution to provide insights to enrichment approaches explored. Fetal mutations and fetal chromosomal abnormalities can be detected by molecular analysis of circulating cell free fetal DNA (ccff DNA) extracted from maternal plasma and classification may be enhanced by fetal fraction enrichment. The collection and processing methods facilitate efficient extraction of circulating cell-free fetal DNA (ccff DNA) from maternal plasma. The enhanced understanding of fetal fraction enrichment approaches and compartmentalization will lead to enhanced performance and decreased costs for NIPT.

2836S

Prenatal diagnosis of mosaic isochromosome 20q detected in amniocentesis. *S. Ito[1], T. Kuchikata[2], H. Yoshihashi[2].* 1) Division of Nursing, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; 2) Division of Medical Genetics, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan.

Mosaicism of abnormal karyotypes in amniocentesis are observed approximately 3-8%, which considered to be either true mosaicism or artifact of cultured amniocytes. When chromosome mosaicism was encountered, the definitive prenatal cytogenetic diagnosis depends on the type of fetal chromosome aberration. However, we often left in a conflict situation and precise information should be offered in prenatal genetic counseling. We describe a case of 34-year-old woman, anxious about the abnormal result of maternal serum screening and underwent the amniocentesis in 17 weeks of gestation. The result of chromosomal analysis was 46,XX,i(20)(q10)[4]/46,XX[11](rate of mosaicism:19%), in which derived from three independent cultures. Whether to continue the pregnancy or not, she was referred for genetic counseling in 19 weeks of gestation. She was informed that the most of fetuses prenatally diagnosed mosaic isochromosome 20q represented normal phenotypic features if not apparent abnormal echo findings, but partly show several malformations and development delay after birth in the previous reports. Decision must be made before 22 weeks of gestation in Japan, and it was impossible to perform further subsequent genetic analysis. After the session, she attended the fetal ultrasonographic examination in 20 weeks of gestation and no major malformation was noted, and decided to carry it to term. At 40 weeks of gestation, female infant was delivered with no distinctive abnormalities. Postnatal chromosome analysis of the lymphocytes cell in infancy showed non-mosaic normal karyotype different from the result revealed in amniocentesis. Although concerns about somatic mosaicism still remained, she showed normal growth and development in good health at 8 months. Discussion: These clinical findings contribute to a better understanding of prenatally diagnosed mosaic isochromosome 20q. The long arm of chromosome 20 is known as gene-rich region, therefore we assume that genetic imbalance due to isochromosome 20q are likely to have no small effect on fetal phenotype and survivability. In this case, the results of ultrasonographic examination were effective and essential for prenatal decision-making. Considerable discrepancy between genotype and phenotype may lead to serial dilemmas during the period of pregnancy and even after birth. In addition to medical information, psychological support in genetic counseling is also necessary.

2837S

Phocomelia in Thrombocytopenia-absent radius (TAR) Syndrome due to compound heterozygosity for a 1q21.1 microdeletion and a RBM8A hypomorphic allele. Report of two cases. *R. Jobling[1], S. Unger[2], P. Shannon[3], A. Toi[4], S. Keating[3], D. Chitayat[1,5].* 1) Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Division of Molecular Pediatrics, Maternité, Clinique Infantile, Lausanne, Switzerland; 3) Mount Sinai Hospital, 3Department of Laboratory Medicine and Pathobiology; 4) Department of Diagnostic Imaging University of Toronto, Toronto, Ontario, Canada; 5) The Prenatal Diagnosis and Medical Genetics Program; University of Toronto, Toronto, Ontario, Canada.

Thrombocytopenia-absent radius (TAR) syndrome is a rare autosomal recessive disorder characterized by megakaryocytic thrombocytopenia and longitudinal limb deficiencies mostly affecting the radial ray. Most patients are compound heterozygotes for a 200 kb interstitial microdeletion in 1q21.1 and a hypomorphic allele in RBM8A, mapping in the deleted segment. We report two cases with TAR syndrome detected prenatally with phocomelia and involvement of the lower limbs. Case 1: The couple had two pregnancies affected with phocomelia and Roberts Syndrome (RS) was first suspected. In their second offspring the fetus had all four limbs affected with upper limb phocomelia, short femurs and contractures with pterygia of the lower limbs' joints. Case 2: During the couple's fourth pregnancy, routine anatomy ultrasound revealed upper limb phocomelia. The couple elected to terminate the pregnancy. Fetal autopsy revealed upper limb phocomelia, single palmer creases, adducted thumbs, bilateral 5th digit clinodactyly and partial syndactyly of finger 2-5 and lower limbs' contractures with pterygia. Both fetuses had a deletion at 1q21.1 encompassing the RBM8A gene inherited from one parent and a point mutation in the RBM8A gene inherited from the other. Fetal ultrasound finding of phocomelia leads to a wide differential diagnosis including maternal thalidomide exposure, RS, Raas-Rothschild Syndrome, Fuhrmann Syndrome among others. , VACTERL association and. An important addition to this differential diagnosis is TAR syndrome.

2838S

Unexpected findings using SNP-array for prenatal diagnosis: benefit or burden? *M. Joosten, M.I. Srebniak, K.E.M. Diderich, L.C.P. Govaerts, S.R. Riedijk, F.A.T. De Vries, R.J.H. Galjaard, D. Van Opstal.* Clinical Genetics, Erasmus MC, Rotterdam, Netherlands.

Background: The array technique is able to detect at least 5% more cytogenetic aberrations than conventional karyotyping in cases of fetal ultrasound abnormalities. Its added value for other indications has also been shown. However, with the increase of resolution the number of pathogenic findings not related to the indication will extend as well. The chance of finding these so-called unexpected diagnoses (UD), which potentially may involve late-onset untreatable disorders, is one of the reasons that the use of this technique in prenatal diagnosis is still controversial. **Methods:** Since 2009 we have performed ~ 2500 prenatal SNP-arrays, initially in pregnancies with ultrasound anomalies, and since 2012 as a first-tier test for all prenatal cytogenetic indications. We retrospectively investigated the prevalence, nature, counseling and pregnancy outcome of UD amongst these cases. **Results:** In approximately 1 in every 200 prenatal SNP-arrays a submicroscopic UD was encountered (n=12). This figure does not include susceptibility loci for mainly neurodevelopmental disorders which can be considered to be UD as well, but which we regard as a separate category of pathogenic findings (Srebniak et al., 2013). The risk of a UD is independent of indication. In most cases (9/12) the UD involved an early-onset disease. In the cases of a severe untreatable disorder (e.g. Angelman syndrome) the UD helped the couples in making a decision about the course of their pregnancy. Prenatal awareness of an early-onset treatable disease (e.g. Leri-Weill dyschondrosteosis) may be considered beneficial for the newborn regarding therapy and follow up. The only late-onset untreatable disorder that we found in 3/12 cases in our cohort, was hereditary neuropathy with liability to pressure palsy (HNPP [MIM162500]), which is generally regarded as a milder disease. None of the dreaded CNVs, such as BRCA1-deletions, were found. **Conclusion:** Due to the added value of detecting extra clinically relevant submicroscopic chromosome aberrations without unmasking any severe late-onset untreatable disease in our cohort of ~2500 prenatal cases, we argue that SNP-array should be the first-tier test for prenatal cytogenetic studies in all indications.

2839S

NIPT in a Clinical Setting: Patient Decisions and Pregnancy Outcomes. *C. Kenyon[1], J. Youngblom[1], B. Blumberg[2], E. Obolenksy[2].* 1) California State University, Stanislaus, Turlock, CA; 2) Kaiser Permanente, Oakland Genetics Department, Oakland, CA.

Noninvasive prenatal testing (NIPT) has recently become clinically available for screening of fetal trisomies, creating an alternative to California State Prenatal Screening (CaPNS) as well as diagnostic tests. The study sought to investigate how NIPT functions in a clinical HMO setting by focusing on the population of women who were eligible for NIPT based on a positive CaPNS result in the Kaiser Permanente Northern California system. Study objectives included identifying the detection rate of screened trisomies, looking at the choice of NIPT based on different factors, describing the range genetic abnormalities not identified by NIPT, and evaluating the use of NIPT as a second-tier test. The Integrated Prenatal Screening database as well as Kaiser Permanente electronic medical records for patients and resulting offspring were retrospectively reviewed for 811 pregnancies eligible for NIPT between October 2012 and June 2013. After receiving a positive CaPNS result, 57.3% of patients chose NIPT as a follow-up test. There were no false-positives or false-negatives and a redraw was requested in 3.87% of all NIPT blood draws. Women who received a trisomy 21 positive CaPNS result were significantly more likely to chose NIPT if they received their CaPNS positive in the first trimester as opposed to the second. CaPNS risk score did not play a significant role in choice of NIPT. A trend that women of Hispanic ancestry were slightly less likely to choose NIPT was observed. Pregnancies with at least one ultrasound abnormality were more often associated with diagnostic testing rather than NIPT, while pregnancies with only one or more soft markers were more likely to choose NIPT. Women age 40 and over were more likely to choose NIPT than women less than 40. As expected, there were some genetic conditions detected in this population that NIPT is not able to screen for. Use of NIPT as a second-tier test was associated with a significantly longer test course than going straight from NIPT to diagnostic testing. These results demonstrate that NIPT has become an important part of prenatal screening and is a popular follow-up choice after a positive CaPNS. However, NIPT is not able to screen for the entire range of genetic conditions that can be screened for by CaPNS or tested for by diagnostic methods. This is important because NIPT has the potential to replace analyte-based prenatal screening as the primary population-based prenatal screening test.

2840S

Prenatal array comparative genomic hybridization (aCGH) in fetuses with structural cardiac anomalies in a medium-sized Canadian Prenatal Genetics Clinic. *J. Lazier[1], D. Fruitman[2,3], J. Lauzon[1,4], F. Bernier[1,2,4], J. Chernos[1,5], R. Simrose[6], M.A. Thomas[1,2,4].* 1) Department of Medical Genetics, University of Calgary, Alberta Children's Hospital, Calgary, Alberta, Canada; 2) Department of Pediatrics, University of Calgary, Alberta Children's Hospital, Calgary, Alberta, Canada; 3) Section of Cardiology, Alberta Children's Hospital, Calgary, Alberta, Canada; 4) Alberta Children's Hospital Research Institute for Child and Maternal Health; 5) Cytogenetic Laboratory, Alberta Children's Hospital, Calgary, Alberta, Canada; 6) Department of Obstetrics and Gynecology, University of Calgary, Calgary, Alberta, Canada.

OBJECTIVES: To determine the yield of aCGH over karyotype in cases of prenatal structural cardiac anomalies in a medium-sized Prenatal Genetics Clinic.

METHODS: 21 cases referred to the Calgary Prenatal Genetics clinic with a structural fetal cardiac anomaly, either isolated or with additional fetal anomalies, were prospectively recruited between December 2011 and March 2013. Cases with a normal karyotype and FISH for deletion 22q11.2 had an array CGH performed at Signature Genomics using the Prenatal Chip OS.

RESULTS: Of the 19 cases who met inclusion criteria, 12 presented with isolated fetal cardiac anomalies and 7 with additional fetal anomalies. The cardiac and extra-cardiac anomalies were varied. Array CGH was successfully completed on 16 cases. One case had a large 8p deletion that was also seen on karyotype and included the GATA4 gene, which has been associated with congenital heart disease. Three cases provided more information than the karyotype, however, were not clearly the etiology for the cardiac anomaly. Two cases had an inherited CNV, including one with a duplication of 16p11.2 and another with a deletion of 15q11.2. One case had the incidental finding of being a carrier of cystinosis, a recessive disease not associated with cardiac anomalies.

CONCLUSIONS: Recent studies into prenatal aCGH have grouped fetal anomalies together but there is a reported need to better define the diagnostic yield of array CGH in individual fetal anomalies, such as cardiac anomalies, with studies that assess the full cohort of cases with that anomaly (deWit, *et al*., 2014). This study prospectively recruited all cases with a fetal cardiac anomaly who had an aCGH performed. We demonstrated a lower than expected diagnostic yield for pathogenic findings, after the exclusion of deletion 22q11.2 deletion. There was a higher than expected detection (12.5%) of variants of unknown significance that could be defined as potentially clinically relevant CNVs. These CNVs are associated with variable penetrance or expressivity for neurocognitive issues and are challenging to counsel in the prenatal genetics clinic.

2841S

Non-invasive Prenatal Diagnosis of Duchenne Muscular Dystrophy: Comprehensive Genetic Diagnosis from Patient to Fetus. *B. Lim[1], J. Chae[1], Y. Lee[2], J. Namkung[2], J. Park[2], S. Yoo[3], S. Lee[3], J. Shin[3], J. Kim[3], J. Seo[3].* 1) Department of pediatrics, Seoul Natinoal University Hospital, Seoul, Seoul, South Korea; 2) Bioinformatics Technology Lab, Healthcare Group, Future Technology R&D Division, SK Telecom, Sungnam, Korea; 3) Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul, Korea.

Background Non-invasive prenatal diagnosis of monogenic disorders using maternal plasma and targeted massively parallel sequencing has been actively investigated. We previously demonstrated comprehensive genetic diagnosis of Duchenne muscular dystrophy patient could be feasible in a single targeted sequencing platform. In the present study, we aimed to demonstrate extended applicability of this approach to carrier detection and non-invasive prenatal diagnosis. Methods Custom solution-based target enrichment was designed to cover whole DMD region. First, targeted massively parallel sequencing was performed using genomic DNAs of four mother and proband pairs to test whether carrier status could be reliably detected. Then, maternal plasma DNAs were collected in two pregnant carrier mothers who also participated in the preceding carrier testing and were sequenced using the same-targeted platform to predict the inheritance of DMD mutation to fetus. Overrepresentation of an inherited allele was determined by comparing average read fraction of two phased haplotypes after examining and correcting recombination event. Results Carrier status of deletion/duplication mutation was reliably detected through coverage plotting and breakpoint analysis. De novo nonsense mutation was also correctly identified in one family. The presence of inherited DMD mutation status was correctly predicted in two families, as early as 6 weeks of gestation. In one of two families, detecting recombination event and reconstructing phased haplotype resulted in correct diagnosis. Conclusions We successfully demonstrated that comprehensive genetic diagnosis of Duchenne muscular dystrophy from patient to fetus would be feasible in a single targeted massively parallel sequencing platform with tiling design. .

2842S

Safeguarding non-invasive prenatal testing with spiked sample tracking barcodes. *K. Neveling, D.T. Thung, L. Beulen, W. Buijsman, I. Gomez, S. van den Heuvel, H. Mieloo, E. Kater-Baats, J. Hehir-Kwa, B.H.W. Faas.* Human Genetics, Radboud university medical center, Nijmegen, Gelderland, Netherlands.

Non-invasive prenatal testing (NIPT) using next generation sequencing (NGS) allows to detect fetal chromosomal aberrations in cell-free fetal DNA present in the mothers' blood. NGS however remains a relatively new technique with not all laboratory processes being automated, hence incidental sample swaps cannot be excluded. For other NGS applications such as exome sequencing, tests are developed that compare exome data and independently derived SNP data, thereby identifying potential sample swaps. However such an approach is not feasible for NIPT due to low sequencing depth. Thus, a reliable sample tracking method for NIPT is currently lacking. We have developed a sample tracking method for NIPT that overcomes the low coverage generally used. Based on a manuscript by Quail et al (2014), we generated PCR products of the bacteriophage PhiX174, including unique 11-mer barcodes. These barcoded amplicons were spiked into DNA isolated from human plasma, followed by library preparation and sequencing. As proof of principle, three different spike-in concentrations (0.1%, 0.5%, and 1%) were tested for three different barcodes. Following sequencing, reads that mapped against PhiX174 (BWA 0.5.9, allowing one mismatch per read) were extracted and inspected for the presence of any of 384 described barcodes. The correct spiked barcode could be unambiguously identified in all samples. A concentration of 1% spiked DNA thereby reflected the best signal to noise ratio (124:1). On average, 24×10^3 PhiX174 reads were assigned to the correct barcode and less than 0.2×10^3 PhiX174 reads were misassigned, thereby leaving enough reads for NIPT (mean 22.0 x 10^6 hg19 mappable reads). In a second experiment, PhiX174 amplicons were directly spiked into human plasma. Again all spiked amplicons could be correctly identified within each sample, for all three different concentrations tested, with an even lower mean signal to noise ratio (3612:1) for the 1% concentration. Consequently, barcoded PhiX174 amplicons can be used for sample tracking in NIPT. The PhiX174 sequences do not map to the human genome, do not interfere with the NIPT data and can be spiked directly into human plasma, without being lost during DNA isolation and library preparation. If these barcodes also remain stable when spiked directly into human blood, the addition of a barcode directly to each NIPT blood sample can be introduced into the diagnostic workflow to facilitate sample tracking and detect sample swaps.

2843S

Maternal copy number variants contribute to the burden of false positive prenatal aneuploidy test results. *M.W. Snyder[1], L.E. Simmons[1,2], J.O. Kitzman[1], J. Henson[2], R. Daza[1], J. Shendure[1], H. Gammill[2,3].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Obstetrics & Gynecology, University of Washington, Seattle, WA; 3) Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA.

Prospective studies of noninvasive prenatal testing (NIPT) for aneuploidy by analysis of circulating cell-free DNA have demonstrated high sensitivity and specificity in both high- and low-risk cohorts. However, the overall low incidence of aneuploidy continues to limit the positive predictive value of these tests for general screening use. To date, the full spectrum of causes of false positive tests is not well understood. From a series of four pregnancies with discordant prenatal tests, we investigated one potential cause: maternal copy number gains. Here, we demonstrate the potential for maternally carried duplications on chromosome 18 to cause false positive results in prenatal testing of cell-free DNA. Using published metrics from one NIPT methodology, we estimated the minimum size of the maternal duplication required to give rise to false positive test results in NIPT frameworks based on counting statistics. The calculated minimum size depends on several factors, including the total number of reads per sample, the coefficient of variation for the chromosome in question, the fetal fraction, and the fetal inheritance of the maternal duplication. Assuming a fetal fraction of 15%; and no inheritance of the maternal CNV, duplications of 1.13 Mb, 452 kb, and 386 kb are sufficient for chromosomes 13, 18, and 21, respectively. We next estimated the probability that a randomly selected woman of European ancestry would carry a non-pathogenic large duplication on chromosomes 13, 18, or 21, and thus be at potentially elevated risk for a false positive result if electing NIPT. We analyzed a set of copy number calls for 19,584 individuals of European descent and selected all CNVs yielding increased copy number in uniquely identifiable genomic regions and of sufficient lengths, as determined by the minimum size analysis described above. Again assuming a 15%; fetal fraction and no fetal inheritance of the duplication, we identified three such events on chromosome 13, 19 on chromosome 18, and 12 on chromosome 21, representing 0.17% of analyzed controls. If published estimates of NIPT false positive rates are representative, our analysis suggests that up to 50% of T18 and 20% of T21 false positive predictions could be due to maternal CNVs.

2844S

Whole-genome prenatal sequencing and integrative genomics: Detection of structural variation from invasive and non-invasive approaches. *M. Talkowski[1,2,5], V. Pillalamarri[1], H. Brand[1], Z. Ordulu[3,4], S. Pereira[3], T. Kammin[3], J. Kitzman[6], J. Shendue[6], J. Gusella[1,2,5], C. Morton[3,4,5].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Departments of Neurology and Genetics, Harvard Medical School, Boston, MA; 3) Departments of Obstetrics, Gynecology, and Reproductive Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 4) Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 5) Broad Institute of Harvard and MIT, Cambridge, MA; 6) Department of Genome Sciences, University of Washington.

Technical innovation has had a dramatic impact in prenatal genetic diagnostics, particularly for congenital anomalies where de novo structural variations (SV) represent a major source of risk but are not considered by non-invasive cell free fetal DNA (cffDNA) studies that are limited to detection of aneuploidy. We recently performed the first example of prenatal clinical diagnosis from whole-genome sequencing (WGS) using large-insert 'jumping libraries' at a cost and timeline comparable to karyotyping. Here, we present real-time prenatal sequencing and integrative genomic interpretation of fetal DNA from an ongoing collection of cases referred for advanced maternal age and harboring de novo balanced SVs as part of the Developmental Genome Anatomy Project. In two subjects, DGAP247 and DGAP248, protein-coding genes were disrupted (KHDRBS3 and RCF3, respectively). Convergent genomic interpretation from public databases and clinical microarray data from over 34,000 cases suggested each variant was of unknown significance, with no associated syndromic phenotype. The parents of DGAP247 chose to receive these findings and carried DGAP247 to term, resulting in a well newborn examination. The parents of DGAP248 chose not to receive the findings. In DGAP258, we identified a pericentric inversion that rearranged the known copy number morbid 6p25.3 locus, as well as a cryptic inversion; neither disrupted genic sequence and the twins received a well newborn examination at birth. We also evaluated the capability of deep WGS to detect a balanced SV from cffDNA by generating 53X coverage of DNA isolated from maternal plasma of DGAP247 during the third trimester, with a measured fetal DNA composition of ~25% of reads. We used a split read algorithm to detect the inversion from 14% of reads, demonstrating feasibility of SV detection from cffDNA. These data highlight the power of prenatal WGS by jumping libraries to detect pathogenic SVs of all classes, however they are reliant on invasive methods that carry substantial risk to the fetus. They also illustrate for the first time the feasibility of balanced SV detection without concomitant dosage imbalance by non-invasive methods, though costs and analytical effort were exorbitant. Collectively, our studies emphasize the unique challenges facing development of a comprehensive prenatal genetic screening strategy that accesses the entire pathogenic mutational spectrum but minimizes risks to the fetus.

2845S

Demonstration of Equivalent Performance for a Noninvasive Prenatal Test (NIPT) using High Output and Rapid Throughput Modes of a Sequencing Instrument. *R.C. Tim, L. Liu, Z. Zhu, P. Liu, B. Sagoe, P. Oeth, M. Ehrich.* Sequenom Laboratories, 3595 John Hopkins Court, San Diego, CA.

The Illumina HiSeq™ 2500 (Illumina®, San Diego, California) is capable of sequencing in two run modes - one for collecting large amounts of data and the other for acquiring data rapidly. This study demonstrates equivalent sequencing performance of a noninvasive prenatal test, the MaterniT21™ PLUS laboratory-developed test, using both the high output and rapid run modes of the instrument. The two run modes of the instrument were compared with the same set of sequencing libraries in a paired fashion. Five library plates, comprised of 385 euploids, 44 trisomy 21 samples, 8 trisomy 18 samples, and 3 trisomy 13 samples, were analyzed on the same HiSeq 2500 using the high output and rapid run modes with 12-plex multiplexing. In a second experiment focusing on the clustering method using rapid run mode, 48 paired libraries were sequenced with on-board cluster generation in 24-plex or following template hybridization on the Illumina cBot in 12-plex. All sequencing libraries were generated from circulating cell-free (ccf) DNA following extraction from 4 mL plasma aliquots. All flow cells were clustered on the Illumina cBot or using on-board cluster generation, and then sequenced on the HiSeq 2500. Sequencing reads were de-multiplexed, aligned to the human genome (hg19) with Bowtie 2, and chromosomal representations were calculated. Although the HiSeq 2500 rapid run mode generates fewer passed filter reads due to a reduced imaging area per flow cell lane, sequencing performance between high output and rapid run mode is remarkably similar. Representations for chromosomes 21, 18, 13, X and Y are highly correlated between the two modes. No statistically significant differences in z-scores were found for the three aneuploidy types and the discriminatory distance between euploids and aneuploids is similar in either run mode. Trisomy 21 sensitivity/specificity was > 99.9% / > 99.9% for the high output mode and > 99.9% / 99.7% for the rapid run mode. The sequencing metrics for onboard cluster generation and for cBot template hybridization using the rapid run mode were also found to be comparable. Therefore, classification performance of the MaterniT21™ PLUS test was shown to be equivalent using both the high output and rapid run modes of the HiSeq 2500. Comparable results were also found in the rapid run mode using either on-board cluster generation or cBot template hybridization.

2846S

False negative NIPT results for trisomy 13, 18 and 21: risk figures derived from cytogenetic investigations in chorionic villi. *D. Van Opstal, M. Srebniak, C. van den Berg, J. Polak, F. de Vries, R-J.H. Galjaard.* Clinical Genetics, ErasmusMC, Rotterdam, Netherlands.

Background: Non-invasive prenatal testing (NIPT) for fetal trisomy detection already revealed that there is a small chance of a false positive and false negative result. This is partly due to the fact that the fetal DNA present in the cell free maternal plasma fraction is derived from the cytotrophoblast of chorionic villi (CV), which is not always representative for the fetus. This we already learned in the eighties from cytogenetic investigations in short-term-cultured CV (STC-villi) in which cells derived from the cytotrophoblast are studied. For this reason, accurate cytogenetic studies in CV should be done by using both STC-villi as well as long-term cultured CV (LTC-villi), the latter investigating cells derived from the mesenchymal core of CV which has the same embryonic origin as the fetus itself. We calculated the risk for a false negative trisomy 13, 18 and 21 NIPT result of a biological nature based on our experience with CV. **Methods** : All cases of fetal trisomy 13, 18 and 21 among ~6000 CV samples that were cytogenetically investigated (STC- and LTC-villi) in our centre between January 2000 and December 2011, were retrospectively studied for the presence of a normal karyotype or mosaicism < 50% in STC-villi. The main indications for CV sampling in these trisomic cases were fetal ultrasound abnormalities, increased nuchal translucency, advanced maternal age > 36 years, and/or abnormal first trimester screening (>1:200). **Results**: 404 (6,7%) cases of trisomies 13, 18 and 21 were found amongst ~6000 samples. Of these 404 cases, 15 (3,7%) had a normal (N=9) or <50% mosaic (N=6) karyotype in STC-villi and therefore would potentially be missed if NIPT was performed in these cases. It involved 6 cases of trisomy 21 and 9 cases of trisomy 18. **Conclusion:** Apart from technical reasons (large BMI, low fetal fraction etc) that may explain false negative NIPT results, in 2 á 3/1000 NIPT samples of patients at high risk, a trisomy 18 or 21 will be missed due to the biological phenomenon of absence of the chromosome aberration in the cytotrophoblast. It is important that patients opting for NIPT are informed about these figures so that an informed choice between invasive an non-invasive testing can be made.

2847S

Prenatal detection of fetal aneuploidy on the Ion Torrent Proton platform. *T. Zwiefelhofer, P. Whitley, K. Roy, M. Saha, T. Burcham, D. van den Boom, M. Ehrich.* Sequenom Laboratories, San Diego, CA.

Noninvasive prenatal testing (NIPT) for fetal aneuploidy detection via massively parallel sequencing has been successfully implemented in a number of high throughput clinical laboratories. Automation and parallelization of the complex workflow has reduced turnaround time and labor while maintaining high sensitivity and specificity. As these developments have improved workflow issues, the availability of new sequencing platforms on the market has introduced additional flexibility in implementation. Platform flexibility should encourage competitive pricing, foster innovation and ultimately, improve patient satisfaction. Here, we examine the performance of an NIPT fetal aneuploidy test on the Ion Torrent Proton platform. We examined the performance of the MaterniT21™ assay using the Ion Torrent™ Proton Sequencer (Life Technologies™, San Diego California). One hundred and fifty-four patient samples, including sixteen from women carrying a known trisomy 21 fetus, as determined by fetal karyotyping, were analyzed. Libraries were prepared and sequenced according to manufacturer's recommendations. Sequenced reads were aligned, filtered for quality and normalized for GC bias. Robust statistics were then applied to identify positive samples with a z-score greater than 3. All patient samples were correctly identified according to their karyotype results. The total number of aligned reads and uniformity of genome coverage was sufficient to generate the necessary discriminatory power to indicate aneuploidy status. Fetal aneuploidy status was correctly determined for 154/154 pregnant females, including 16 carrying a T21 fetus. Though the current Proton workflow requires more labor than is optimal for a production environment, significant improvements in that respect are anticipated in the launch of the Ion Chef template preparation system. Sequencing time was brief at < 3 hours and data analysis consistent with standard platforms. In summary, the performance of the MaterniT21™ assay on the Ion Torrent Proton platform in this limited study suggests the possibility of its suitability for implementation in a clinical environment.

2848S

Utilization of a SNP array in prenatal diagnosis of Ellis van Creveld syndrome in a consanguineous couple; a case report. *B. Suskin[1,2], K. Erskine[1,2], J. Gebb[1,2], S. Klugman[1,2], P. Dar[1,2].* 1) Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine; 2) Montefiore Medical Center, Bronx NY.

Background: Osteochondrodysplasias and dysostoses include more than 350 disorders. As a group, their pattern of inheritance can be dominant, recessive, X-linked, or secondary to imprinting, somatic mosaicism or teratogens. Individually, they are relatively rare leading to challenges in definitive prenatal diagnosis. Case: 25 year old G2P1 who had a 3 week femur length lag on an 18 weeks dating scan presented for detailed anatomical survey at 20 weeks. History was remarkable for gestational diabetes in her prior pregnancy and consanguinity; the patient and her husband are first cousins. The scan revealed rhizo-mesomelia, polydactyly, syndactyly, mildly shortened ribs, hypoplastic left ventricle, aortic coarctation and ventricular septal defect. Chest circumference was above the 10th percentile. After extensive genetic counseling, the couple opted to undergo amniocentesis with microarray analysis with the understanding the test may not have a diagnostic yield. The array revealed multiple regions of loss of heterozygosity which is to be expected when an individual's parents share a common ancestor. A query on these regions was performed to identify possible candidate genes for the phenotype observed and identified EVC and EVC2. Mutations in these genes are associated with Ellis van Creveld (EVC) syndrome, also known as chondroectodermal dysplasia, the most likely condition based on the sonographic features. Following the SNP array results a direct mutation analysis on EVC and EVC2 was sent. Discussion: EVC is a rare disorder with an incidence of 1 in 60,000. Although it has been diagnosed prenatally in cases with a known family history, this case appears to be among the earliest presumed diagnoses in a family without prior history of EVC. The utilization of SNP array as indirect first step assisted in this case to narrow down the differential diagnosis and direct us towards the most adequate mutation analysis test. The large number of skeletal dysplasias and overlapping sonographic features, makes direct mutation analysis in new and rare cases impractical. The utilization of SNP array, especially in cases with consanguinity, can be incredibly helpful first step to guide us in choosing the correct direct mutation analysis studies. This will allow us to provide the patients with accurate diagnosis and improve our counseling on expected outcomes.

2849S

Uniparental origin GWAS of human gestational age implicates genes involved in angiogenesis. *J. Bacelis[1], V. Sengpiel[2], R. Myhre[3], G. Zhang[4], L. Muglia[5], S. Nilsson[6], B. Jacobsson[2,3].* 1) Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, SE-40530, Gothenburg, Sweden; 2) Department of Obstetrics and Gynaecology, Sahlgrenska Academy, Sahlgrenska University Hospital/Östra, SE-41685, Gothenburg, Sweden; 3) Department of Genes and Environment, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; 4) Human Genetics Division, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 5) Center for Prevention of Preterm Birth, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, United States of America; 6) Mathematical Sciences, Chalmers University of Technology, SE-41296, Gothenburg, Sweden.

Background: More than 10% of the pregnancies World-wide end up in preterm delivery (PTD) - the major cause of neonatal death and morbidity. More than 30% of the variation in gestational age can be attributed to genetic factors. Our knowledge of such factors is very limited. **Hypothesis**: Genetic polymorphisms (SNPs) affecting gestational age are located in genomic regions that experience a uniparental silencing (imprinting) and thus are more likely to be detected by genome-wide association study (GWAS) in separated haplotypes of the fetal genome. **Methods**: We used data from 991 mother-child pairs from the Norwegian Mother and Child cohort. Genotypes were phased using family structure data and reference haplotypes from 1000 Genomes Project. SNPs at maternally and paternally derived haplotypes were tested for association with gestational age (in days), while adjusting for child's sex, maternal age and parity. **Results**: None of the 525577 tested SNPs showed a genome-wide significance ($p=2\times10\text{-}6$). The top SNPs from paternal-haploid analysis were located in the introns of angiogenesis-related genes *ZBTB16*, *ANG*, *COL6A1.* Angiogenesis is an essential process during placental vascularisation and can influence nutrient supply to the fetus, thus affecting gestational age. *ANG* gene encodes angiogenin which is induced both by hypoxia and inflammation, and has been shown to be a marker for PTD. *ZBTB16* is known to regulate expression of collagen, metalloproteinases, secretion of interleukins and interferon gamma, all of which are known to be involved in PTD. Moreover, *ZBTB16* showed signs of imprinting in other studies. The top SNPs from maternal-haploid analysis were located in the intron of the gene *ABCC8*, which is involved in insulin (also imprinting-related) secretion. **Conclusion**: Our findings are consistent with the intergenomic-conflict hypothesis, which predicts that imprinting is a way for paternal and maternal genomes to negotiate the distribution of resources between the mother and the fetus. The fact that a hypothesis-free parent-of-origin genotype analysis can detect fetal growth-related and imprinting-related genes while investigating a fetal growth-related phenotype - increases a plausibility of functional involvement of reported genes in PTD. Both top genes (*ZBTB16* and *ABCC8*) are located on chromosome 11, the most imprinted human chromosome. **Replication**: Currently we are carrying out the replication analyses in the Finnish and Danish (DNBC) cohorts.

2850S

Long non-coding RNAs associated with ubiquitin pathway involved in preterm premature rupture of membrane. *N. Zhong.* Human Genetics, New York State Institute for Basic Research in DD, Staten Island, NY.

Preterm birth (PTB) is the leading cause of neonatal death. Eighty percent of cases of PTB are spontaneous (sPTB). About 33% of sPTBs result from preterm premature rupture of membrane (PPROM). Although genetic predisposition, maternal exposure to physiopathological condition(s), or gene-environmental interaction(s) have been identified as etiological contributor(s) to PPROM, the genetic mechanism underlying PPROM is yet unclear. Our earlier studies determined that transcriptional/posttranscriptional regulator long non-coding RNA (lncRNA) is associated with PPROM. In this study, we identified that lncRNAs associated with the ubiquitin pathway are up- or down-regulated in PPROM. A total of 125 hits with 38 lncRNAs were identified to associate with the ubiquitin pathway from 40 placentas of groups A, sPTB; B, full-term birth (FTB); C, PPROM; and D, premature rupture of membrane (PROM). Five ubiquitin proteins, including USP10 (ubiquitin carboxyl-terminal hydrolase 10), UBE2B (ubiquitin- conjugating enzyme E2B), UBE2H (ubiquitin-conjugating enzyme E2H), UBE2Q1 (ubiquitin-conjugating enzyme E2Q1), and UBXN6 (UBX domain-containing protein 6) were found up-regulated; and UCHL5 (ubiquitin carboxyl-terminal hydrolase isoform L5) was down-regulated when the non-membrane rupture (AB) group was compared to the rupture of membranes (CD) group. Differentially expressed lncRNAs involved in the ubiquitin pathway could be up- (labeled as red or pink) or down- regulated. Our findings suggested that ubiquitin pathway might present a pathogenic mechanism underlying the membrane rupture in the PPROM, which deserves for further intensive investigations.

2851S

Association of candidate gene single nucleotide polymorphisms with the clinical subtypes of preterm birth (PTB). *L.G. Gimenez[1,2], A.M. Momany[5], J.A. Gili[1,2], F.A. Poletta[1,2,4], T. Busch[5], B. Comas[1,2], V. Cosentino[2], C. Saleme[3], H. Krupitzki[1], E. Castilla[2,4], E. Gadow[1], J.C. Murray[5], J.S. Lopez Camelo[1,2,4].* 1) Dirección de Investigación, CEMIC (Centro de Educación Médica e Investigaciones Clínicas), Buenos Aires, Argentina; 2) ECLAMC (Estudio Colaborativo Latinoamericano de Malformaciones Congénitas) at CEMIC, Buenos Aires, Argentina; 3) Maternidad Nuestra Señora de la Merced, Tucumán, Argentina; 4) INAGEMP (Instituto Nacional de Genética Médica Populacional) at CEMIC, Buenos Aires, Argentina; 5) Department of Pediatrics, University of Iowa, Iowa City, USA.

Background. Preterm birth (PTB) is the leading cause of perinatal morbidity and mortality worldwide. The etiology of PTB is multi-factorial, heterogeneous, and there is strong evidence of genetic susceptibility. The aim of this work was to investigate the association between 24 single nucleotide polymorphism (SNPs) and the different clinical subtypes of preterm birth: spontaneous (PTB-I), premature rupture of membrane (PTB-PROM) and medically indicated (PTB-M). These SNPs have been previously studied in a heterogeneous group of PTBs. Methods. The sample included 674 triads (proband, mother, father) recruited at the Nuestra Señora de la Merced Maternity Hospital in Tucumán, Argentina. Of these triads, 233 had probands from PTB-I, 241 had probands from PTB-PROM and 200 had probands from PTB-M. We studied 24 SNPs in 18 candidate genes. Genotyping was performed using Applied Biosystems Taqman probes and the Fluidigm genotyping platform. SNPs were chosen based on previous reports of significant association with preterm birth. Data were analyzed using the Transmission Disequilibrium Test (TDT) (Spielman et al. 1993). The p values (<0.05) were uncorrected for multiple comparisons. Preliminary results. We found a significant association (P<0.05) between a SNP in COL4A3 (rs10178458), a SNP in PON1 (rs2272365), and a SNP in CRHR1 (rs4458044) with PTB-I. A SNP in F3 (rs610277) showed significant association (P<0.05) with PTB-PROM and PTB-M. A SNP in KCNN3 (rs883319) showed significant association (P<0.05) with PTB-M. Conclusions. This study suggests different genetic influences between the different clinical subtypes. These findings may have implications in understanding the pathophysiology of clinical subtypes of preterm birth.

2852S

TLR1 SNP associated with preterm birth in a Wisconsin birth cohort. *D. Pillers, S. Tokarz, J. Eickhoff, L. Zyduck, W. Luo, B. Pattnaik, M. Baker.* Pediatrics, University of Wisconsin-Madison, Madison, WI.

Introduction: An important medical challenge is to mitigate the public health impact of preterm birth, both in its cost and complications. More than 30% of extremely premature infants are the products of pregnancies that have been complicated by infection. The fetal unit includes the fetus, the fetal membranes (chorion and amnion) and the placenta. Chorioamnionitis, or infection of the fetal membranes, can lead to preterm birth via rupture of the membranes. Although chorioamnionitis is associated with infants of any gestational age, it is predominantly found in the preterm infant, and its incidence increases as the gestational age decreases. The first line of defense against pathogens is the innate immune system, and the portal of entry is the Toll-like receptor (TLR). TLRs are pattern-recognition receptors that recognize pathogen-associated molecular patterns. Activation of TLRs initiates an inflammation cascade involving downstream signaling moieties and cytokines. Genetic variation in these pathways is associated with a variety of disease processes, including chronic and deregulated inflammation. TLRs are initiators of NF-κB and play a role in mediating the inflammatory events associated with labor. Purpose: We tested the hypothesis that genetic alterations of the fetal innate immune response are associated with an increased risk of prematurity. Approach: We evaluated three TLRs, 1, 2 and 4, using single nucleotide polymorphisms (SNPs). We utilized a unique and valuable set of newborn DNA samples derived from residual newborn screening specimens from over 3000 Wisconsin infants who represent the full spectrum of gestational ages (23 - 42 weeks). A univariate analysis was used to determine whether the TLR1 genotype is a predictor for gestational age and/or birth weight. Results: We found no significant correlation between TLR2 or TLR4 SNPs and preterm birth. There was a correlation between the TLR1 SNP rs4986791 (N248S) genotype in both the birth weight and gestational age analyses when compared to full-term populations (37-42 wks gestational age). There was a significant association between TLR1 SNP (N248S) and preterm birth in the non-Hispanic Wisconsin black population (P< 1×10-8). Conclusions: Our data suggest that TLR1 alterations may be associated with an increased risk of preterm birth in black infants in Wisconsin.

2853S

Whole Exome Sequencing of Hispanic Infants Reveal Novel Pathways Implicated In Spontaneous Preterm Birth. *M.K. Veerapen[1,2], E. Rampersaud[3], L. Pelaez[4], J.E. Potter[5], S. Wu[6], M.M. Rodriguez[4], O.A. Bodamer[1,2,3].* 1) Dr John T. Macdonald Department of Human Genetics; 2) Department of Biochemistry and Molecular Biology; 3) John P. Hussman Institute of Human Genomics; 4) Department of Pathology; 5) Department of Obstetrics and Gynecology; 6) Department of Pediatrics, Division of Neonatology, University of Miami, FL.

Approximately 12% of all infants in the United States are born prematurely with considerable variability based on ancestry, social status and access to health-care. Preterm birth (PTB) is multifactorial with an overall heritability of up to 37% and higher variance in spontaneous PTB (SPTB). Despite an increased risk for SPTB in Hispanic women, few studies have identified genetic factors associated with SPTB in this population. Consequently, much of the heritability remains unexplained. Whole exome sequencing (WES) has the potential to close this knowledge gap through the identification of additional genetic factors. The use of WES on archival tissues is novel, facilitating the study of well characterized cohorts from existing biorepositories.We proposed to utilize a cohort of retrospectively collected umbilical cords from SPTB Hispanic infants for the identification of genetic risk factors to SPTB by WES. Twenty-three Hispanic infants following SPTB were retrospectively identified through an umbilical cord collection at the University of Miami/Jackson Memorial Hospital. gDNA was isolated for ACE-capture WES (Personalis Inc., Menlo Park, CA). Individual exomes were filtered for variants with a genotype quality (GQ)>90 and analyzed using the likelihood ratio test(LRT) against the 1000 Genomes Project as background (VAAST v2.0.4). Significant shared genes (FDR<0.05) and VAAST scores were used for pathway analysis with MetaCore (v6.18 Build 65505). Mean gestational age was 34.1±1.72 weeks; birth weight was 2.36±0.46 kg; and APGAR scores at 10min were 8.8±0.59. gDNA (4.66±1.43 μg) were successfully extracted from archival umbilical cords for WES: average read depth was 44x with a mean coverage of 42 Mb. Individuals exomes were GQ filtered and gene-based VAAST analyses identified 217 shared genes (FDR<0.05). The VAAST scores by LRT were used for pathway analysis which revealed a predominance of novel cytoskeletal and immunological pathways (FDR<0.05): highly scored genes (>20) include *TUBA3E, KRT14*, and *HLA-DQA1*. To the best of our knowledge, this is the first retrospective SPTB study on Hispanic infants utilizing WES to identify novel genes and pathways not previously implicated in SPTB. The genes identified in cytoskeletal and inflammatory pathways likely modulate infant response to inflammatory stimuli contributing to the pathophysiology of SPTB. Our analyses will be further improved once integrated with maternal datasets to increase clinical significance.

2854S

Noninvasive detection of a balanced fetal translocation from maternal plasma. *S. Kim, T. Jensen, D. van den Boom, C. Deciu, M. Ehrich.* Bioinformatics, Sequenom, Inc., San Diego, CA.

Noninvasive prenatal testing based on massively parallel sequencing (MPS) of circulating cell free DNA (ccfDNA) from pregnant plasma offers a powerful tool for detecting fetal chromosomal aneuploidies and other copy number variations; however, copy neutral structural rearrangements have proven challenging. We aimed to detect and characterize a balanced fetal specific translocation event by sequencing ccfDNA from maternal plasma. Simulations were used to develop an algorithm which leverages base incremental changes in mapping characteristics of ccfDNA to identify paired end reads potentially harboring structural rearrangements. We then applied this methodology after performing high-coverage, 100bp paired-end sequencing of ccfDNA isolated from the plasma of a 38 year-old pregnant donor carrying a fetus with a balanced translocation. Our algorithm identified the known translocation (p=1.21e-8) and discounted the likelihood of others, enabling the base specific localization of the breakpoints. Furthermore, while no evidence of chromothripsis existed, we identified a 6bp deletion present within der(8) which is absent from the der(11) reciprocal rearrangement after de novo assembly of 76 chimeric reads. Overall, we have demonstrated here the first proof of concept study detecting and characterizing a balanced fetal specific translocation by non-targeted whole genome sequencing of ccfDNA from maternal plasma.

2855S

Low folate levels and MTHFR polymorphism C677T in case-mothers of children with neural tube defects and control-mothers of Pakistani origin. A Case- control study. *N. Nauman[3], S. Jalali[1], S. Shami[1], S. Rafiq[3], G. Groβe[2], A. Hilger[2], M. Draaken[2], M. Nöthen[2], M. Ludwig[2], H. Reutter[2].* 1) Pathology, Rawalpindi Medical College, Rawalpindi, Pakistan; 2) Institute of Human Genetics, Bonn, Germany Department of Neonatology, Bonn, Germany; 3) Department of Pathology, Holy Family Hospital, Rawalpindi, Pakistan.

Introduction: Neural tube defects (NTDs) are congenital malformations of central nervous system which result from a failure of the neural tube to close during the fourth week of embryogenesis. Low folate levels are associated with increased risk of neural tube defects. Genetic studies examining the gene coding for the folate metabolizing methylenetetrahydrofolate reductase (MTHFR) enzyme suggest that the functional 677C/T thermolabile polymorphism contributes a genetic risk to NTDs. Previous studies suggest that not only the fetal genotype but also the maternal genotype might have an impact on the fetal development. The deleterious effects of this mutation can be overcome if folate levels are adequate. Here we carried out a case-control study of case-mothers and control-mothers of Pakistani origin. Patients and Methods: We examined 109 case-mothers of children born with NTDs and of100 control-mothers without history of NTDs in their offspring. Case-mothers and control mothers were of Pakistani origin. Red blood cell folate and serum folate were determined. Genotype and folate level comparisons were carried out using chi-square analysis. Results: In the case mothers mean RBC folate and serum folate were highly significantly low as compared to control mothers (p-<0.0001). Genotype analysis of case-mothers and control-mothers revealed 11(10.09%) and 2(2%) respectively to be homozygous for the MTHFR 677T allele. There were 32(29.35%) case-mothers and 26(26%) control-mothers to be heterozygous for this polymorphism. Genotype comparison by chi-square analysis of case-mothers and control-mothers was significant (p=<0.0393). Conclusion: We found the homozygous MTHFR 677TT genotype to be more frequent among case-mothers compared to control-mothers. Mean RBC and serum folate levels were significantly low in case mothers as compared to control mothers. Combined analysis of these results with the information folate supplementation during the periconceptional period of case-mothers and control-mothers is warranted to further elucidate the role of the MTHFR genotype and folate supplementation per se in the risk of NTDs in pregnancies of women of Pakistani origin. Key words: Case- control study, MTHFR polymorphism, folate levels, neural tube defects.

2856S

FETAL DEMISE: DIAGNOSIS AND MANAGEMENT. INITIAL EXPERIENCE IN THE STATE OF INDIANA - 129 PATIENTS. *M. Tucker, L. Escobar, A. Hall.* Medical Genetics , St.Vincent, Indianapolis, IN., Select a Country.

Historically up to 60%; of fetal demise situations have no identifiable etiology. Attempting to determine the cause of fetal death remains important because it may influence estimates of recurrence and future preconceptional counseling, pregnancy management, prenatal diagnostic procedures, and neonatal management. We present data on our experience during the last five years of the state wide Perinatal Loss Evaluation Program in our institution. All patients were referred to our center for evaluation of fetal demise with the majority being second or third trimester losses. We either examined patients directly or through photographic material obtained by previously trained staff. We requested cord blood or tissue material for cytogenetic analysis and further history was obtained by our genetic counselor. Cases were evaluated by a pediatric pathologist, a dysmorphologist, and maternal fetal medicine specialist. The etiology of the demise in each case was suggested after reviewing medical records, autopsy, dysmorphologic exam and cytogenetic analysis.

Preliminary results indicated a likely identifiable cause for the demise in 88% of case (N=129). Fetal (33%), multiple risk factors (15%), placental (8%), infectious (8%), cervical (6%), umbilical cord (6%).

We briefly review here the diagnoses seen, particularly those attributed to fetal abnormalities with likely or proven genetic factors. Our data demonstrates the importance of evaluating fetal demise material in order to provide appropriate information to families in terms of cause and recurrence of their loss. We review the value of appropriate management of fetal tissue by pathology and dysmorphology as well as the value of photographic evaluation post mortem when products of conception are not available.

2857S

Molecular and histopathological findings in placentas of newborns with Down Syndrome. *R. García-Robles[1], J. Martinez[2], P. Ayala-Ramírez[2], M. Olaya[3], M. Bermúdez[2].* 1) Instituto de Investigación en Nutrición, Genética y Metabolismo. Universidad El Bosque, Bogotá, Colombia; 2) Instituto de Genética Humana. Pontificia Universidad Javeriana, Bogotá, Colombia; 3) Departamento de Patología. Hospital Universitario San Ignacio - Pontificia Universidad Javeriana, Bogotá, Colombia.

Down Syndrome is the most frequent chromosomopathy with a worldwide incidence of 1/600-700 newborns. Down Syndrome is the main cause of mental retardation and is caused by a trisomy of chromosome 21 involving a specific region named “Down Syndrome critical region”. The locus of the CBS gene is on this region and this gene codes for the enzyme Cystathionine β Synthase that participates in homocysteine metabolism. It has been reported overexpression of this gene in some cells and tissues of individuals with Down Syndrome and it is possible that this abnormal expression could be involved in some phenotypical findings of the disease. However, it has not been described this finding in placentas of affected individuals. The aim of this study was to assess the expression levels of mRNA an protein of Cystathionine β Synthase and histopathological changes in placentas of individuals affected with Down Syndrome and a control group of healthy and normal newborns. Thereby, we studied six placentas of newborns with Down Syndrome and 16 placentas of control group. We evaluated mRNA by real time PCR and protein by immunohistochemistry. An expert pathologist was blinded to diagnosis and evaluated the placentas. We found increased expression of Cystathionine β Synthase mRNA levels (p=0.0465) and protein levels (p=0.0090) in placentas of case group compared to control group. The histopathological changes showed statistically significant differences between groups in the following findings: irregular villous outlines (p=0.0007) and trophoblastic inclusions (p=0.0037). These results suggest overexpression of Cystathionine β Synthase in placentas of individuals with Down Syndrome, possibly causing alteration in homocysteine metabolism locally. Besides, histopathological findings in placenta were associated with trisomy 21. Finally, these molecular and pathological findings could help to identify individuals with Down Syndrome when the placenta is available to be assessed.

2858S

Rising Whole Body Counts of ^{137}Cs (WBC) in Pregnant Women and Persistent Elevated Rates of Neural Tube Defects (NTD) and Microcephaly/Microphthalmia (M/M) in a Chornobyl Impacted Region of Rivne (R) Province in Ukraine. *W. Wertelecki[1], L. Yevtushok[2], N. Zymak-Zakutnia[3], Z. Sosyniuk[2], S. Lapchenko[4], I. Kuznietsov[5], A. Korblein[6].* 1) OMNI-Net USA; 2) Rivne Diagnostic Center, Rivne, Ukraine; 3) Medical Genetics Center, Khmelnytsky, Ukraine; 4) OMNI-Net Ukraine; 5) Eastern European University, Lutsk, Ukraine; 6) OMNI-Net Germany.

Purpose: A population-based surveillance of congenital anomalies in R demonstrated elevated rates of NTD and M/M. Two prevalent teratogenic risks in R are alcohol and ^{137}Cs. Analyses indicated that alcohol is an unlikely primary cause. This study concerns WBC of ambulatory patients and pregnant women. **Methods:** Study of 44,438 WBC patterns recorded from adult ambulatory patients (2001-2013) inclusive of 6,425 pregnant women of known gestational age and body weight (2011-2013). Population-based 2000-2012 rates of NTD and M/M were concurrently computed. **Results:** WBC were statistically significantly higher among patients and pregnant women residing in Chornobyl contaminated regions of R. The average WBC of women rose significantly since 2001. Among pregnant women from the contaminated regions of R, the average WBC of ^{137}Cs were 40.7 and 39.7 Bq/kg among those under the age of 20 or older than 29 years respectively. The corresponding values among women from non-contaminated regions of R were 11.2 and 10.8 respectively. The population-based rates of NTD and M/M in polluted regions of R are persistently elevated and do not differ significantly from those reported earlier. Among the highest NTD-M/M rates are those in two counties in proximity to nuclear power plants. One of these counties is considered to be not contaminated by Chornobyl. **Interpretation:** The patterns of NTD and M/M rates in R may reflect impacts of alcohol and ^{137}Cs although the former is an unlikely primary cause. The impacts of two nuclear power plants on NTD-M/M rates in R should be considered by prospective cause-effect investigations.

2859S

Assessing the causal relationship between maternal height and birth outcomes: A Mendelian randomization analysis. *G. Zhang[1], J. Bacelis[2], C. Lengyel[3], B. Jacobsson[4,5], L. Muglia[3].* 1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 3) Center for Prevention of Preterm Birth, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 4) Department of Obstetrics and Gynaecology, Sahlgrenska Academy, Sahlgrenska University Hospital/Ostra, Gothenburg, Sweden; 5) Department of Genes and Environment, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway.

Background: Epidemiologic studies have shown that maternal height is significantly associated with gestational age and various birth outcomes, such as birth length and birth weight. However, it is unclear whether these associations truly reflect causal effects of maternal height. Alternatively, as an indicator of maternal nutrition status, maternal height could be indirectly associated with those birth outcomes. In this study we investigated the causal influence of maternal height on birth outcomes by Mendelian randomization using height genetic score as instrumental variable.

Methods: We used data of 3,486 mother/infant pairs from three birth cohorts collected from three northern European countries (Finland, Denmark and Norway). Genome-wide SNP genotype data were generated by SNP arrays and imputation. We constructed height genetic score using 171 SNPs to estimate the causal effects of maternal height on birth outcomes.

Results: In observational analysis, maternal height was significantly associated with gestational age ($p=1.4\times10^{-6}$), birth length ($p=3.26\times10^{-13}$) and birth weight ($p<2\times10^{-16}$). The constructed height genetic score in mothers was significantly associated with maternal height ($p<2\times10^{-16}$) and could explain more than 8% of the observed variance in maternal height. In Mendelian randomization analysis, height genetic score in mothers was significantly associated with infant's birth length (p=0.0015) and birth weight ($p=1.49\times10^{-5}$) and the estimated causal effects were directionally concordant with the observed associations (i.e. higher maternal height was associated with longer birth length and higher birth weight). The association between maternal height genetic score and gestational age was not significant (p=0.71).

Conclusion: Our results highly suggest that maternal height, as an important component of prenatal environment, causally affects both birth length and birth weight; however, whether maternal height causally impacts gestational age, or reflects nutritional or other influences on overall maternal health, requires further evidence.

2860S

Maternal and placental genome-wide and candidate gene association studies of placental abruption. *M. Denis[1], D.A. Enquohabrie[2,3], M.G. Tadesse[4], B. Gelaye[1], S.E. Sanchez[5,6], M. Salazar[7], C.V. Ananth[8,9], M.A. Williams[1].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Center for Perinatal Studies, Swedish Medical Center, Seattle, WA; 3) Department of Epidemiology, University of Washington School of Public Health, Seattle, WA; 4) Department of Mathematics and Statistics, Georgetown University, Washington, DC; 5) Sección de Post Grado, Facultad de Medicina Humana, Universidad San Martín de Porres, Lima, Peru; 6) A.C. PROESA, Lima, Peru; 7) Department of Obstetrics and Gynecology, San Marcos University, Lima, Peru; 8) Department of Obstetrics and Gynecology, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY; 9) Department of Epidemiology, Joseph L. Mailman School of Public Health, Columbia University, New York, NY.

Introduction: Available evidence supports the role of genetics in the pathogenesis of placental abruption (PA), the premature separation of the placenta from the uterus, an important cause of perinatal mortality. Associations of maternal genetic variants with risk of PA have been demonstrated. However, PA-related variations in the placental genome and interactions between maternal and placental genomes on PA risk have not been investigated. **Methods**: Maternal blood and placental samples collected from participants of the Peruvian Abruptio Placentae Epidemiology study were genotyped using the Illumina Cardio-Metabochip platform. A total of 118,782 genome-wide SNPs and 333 SNPs in 32 candidate genes from mitochondrial biogenesis and oxidative phosphorylation pathways were examined in placental DNA from 280 PA cases and 244 controls. For interaction analyses in maternal and placental genomes, 325 SNPs from 32 candidate genes were assessed among 222 PA cases and 198 controls. Univariate and penalized logistic regression models were fit to estimate odds ratios. We also examined associations of combinations of SNPs with PA risk using weighted genetic risk scores (WGRS) and haplotype-based analyses. A multinomial model was used to investigate maternal and placental genome interactions. Functions of genes represented by significant SNPs were further examined using pathway analyses. **Results**: Overall, in placental genome-wide and candidate gene analyses, no SNP was significant after false discovery rate correction. The top GWAS hits were *rs544201*, *rs1484464* (*CTNNA2*), *rs4149570* (*TNFRSF1A*) and *rs13055470* (*ZNRF3*) (nominal p-values: 1.11e-05 to 5.00e-05). Genes participating in cell cycle, growth and proliferation were overrepresented by the top 200 SNPs of the GWAS. Subjects in the highest quartiles for WGRS (9 SNPs selected by lasso regression) had a 5.5-fold higher risk (95%CI 3.2-9.7) of PA compared with those in the lowest quartile. A haplotype block in *PPARG* was associated with a 50% (95% CI: 0.29-0.88) lower risk of PA compared to the referent haplotype. We found a strong interaction between maternal and placental genomes on PA risk for two SNPs in *PPARG* (*chr3:12313450* and *chr3:12412978* with p-values 8.62E-08 and 7.89E-10). **Discussion**: Variations in the placental genome and interactions between maternal-placental genetic variations can contribute to PA risk. Larger studies in this area can help advance our understanding of PA pathogenesis.

2861S

Comprehensive genotype phenotype correlations reveal NLRP7 role in regulating the balance between embryonic tissue differentiation and trophoblastic proliferation. *NMP. Nguyen[1], L. Zhang[1], P. Sauthier[2], C. Dery[1], K. Rahimi[3], R. Reddy[1], J. Arseneau[1], A. Cheung[4], U. Surti[5], L. Hoffner[5], M. Seoud[6], G. Zaatar[6], R. Bagga[7], R. Srinivasan[7], M. Brequet[2], W. Buckett[1], P. Coullin[8], A. Ap[1], R. Slim[1].* 1) Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Service de Gynécologie Oncologique, Montréal, Québec, Canada; 3) Department of Pathology, Hôpital Notre-Dame, Montréal, Québec, Canada; 4) Department of Pathology, University of Hong Kong, Queen Mary Hospital, Hong Kong, China; 5) University of Pittsburgh, Pittsburgh, Pennsylvania, USA; 6) American University of Beirut, Beirut, Lebanon; 7) Post Graduate Institute of Medical Education and Research, PGIMER, Chandigarh, India; 8) Endocrinologie et Génétique de la Reproduction et du Développement, Clamart, France.

Hydatidiform mole is a human pregnancy with excessive trophoblastic proliferation and abnormal embryonic development. NLRP7, a nucleotide oligomerization receptor, pyrin containing 7, is a major gene responsible for recurrent hydatidiform moles. Recurrent HMs are caused by recessive mutations, mostly in NLRP7, and have diploid biparental genomes To better understand the effect of NLRP7 on the mechanism leading to hydatidiform moles, we used a combination of five different approaches, flow cytometry, fluorescent in situ hybridization, microsatellite DNA genotyping, and p57KIP2 and Ki-67 immunohistochemistry to comprehensively characterize 103 hydatidiform moles from patients with or without NLRP7 mutations and variants. We confirm that most molar tissues from patients with two NLRP7 defective alleles are diploid biparental with a single cellular population. importantly, our data demonstrate significant genotype-phenotype correlation between the nature of the mutations and the histopathological features of thehydatidiform moles. Protein-truncating mutations were associated with more severe molar phenotype such as excessive trophoblastic proliferation and absence of embryonic tissues while missense mutations were associated with milder phenotype and were more permissive for embryonic tissue development. Our data suggest that NLRP7, depending on the severity of its mutations, regulates the time at which embryonic development arrests.

2862S

Kidney disease genes linked to Hyperemesis Gravidarum. *M. Schoenberg-Fejzo[1], K. MacGibbon[2], J. Sinsheimer[3], P.L. Reddy[4], P. Pajukanta[4], K. Tabsh[1].* 1) Department of Obstetrics and Gynecology, David Geffen School of Medicine, University of California, Los Angeles, CA; 2) Hyperemesis Education and Research Foundation, Leesburg, Virginia; 3) Departments of Biostatistics, Biomathematics, & Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA; 4) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA.

Hyperemesis Gravidarum (HG), severe nausea and vomiting of pregnancy, often results in dehydration and undernutrition and is the 2nd leading cause of hospitalization in pregnancy after preterm birth. Current treatments are largely ineffective and the cause is unknown. Evidence for a genetic component comes primarily from a classic twin study and reports of familial aggregation. This study was undertaken to elucidate the etiology of HG by identifying predisposing genes. HG reduces reproductive fitness, which makes whole exome sequencing to find rare variants a reasonable approach. We performed exome sequencing of 18 individuals from 5 families with HG. Using a series of filtering steps, we identified rare, damaging, biologically relevant variants exhibiting co-segregation with affection. Potentially causative variants were screened in available family members, additional families, and a replicate group of over 570 cases with HG and over 425 unaffected controls using Sanger Sequencing, Fluidigm, and Taqman. Three of five families carry biologically related rare damaging variants in genes involved in kidney function: *PKHD1*, *PKD1*, and *LAMA5*. None of over 425 controls carried these variants. Thus, a common pathway may be responsible for the majority of cases of familial HG. This novel discovery of rare variants relating to HG susceptibility is the first step in understanding the biology of HG; may be relevant to understanding genetic susceptibility to nausea and vomiting in general; and provides a new path for development of more effective therapies.

2863S

Interbirth interval varies according to HLA inheritance of first and second born siblings. *C.A. GENTIL[1], H.S. GAMMILL[1,3], S.B. KANAAN[1], C.T. LUU[1], T.M. AYDELOTTE[1], J.L. NELSON[1,2].* 1) Clinical Research Division, Fred Hutchinson Cancer Research Center, SEATTLE, WA; 2) Division of Rheumatology, University of Washington, SEATTLE, WA; 3) Department of Obstetrics and Gynecology, University of Washington, SEATTLE, WA.

Background: Interbirth interval (IBI), the time between births of siblings, may reflect disparate evolutionary pressures from mother and each child. Fetal-maternal exchange during pregnancy results in long-term persistence of low levels of exchanged cells in respective individuals, referred to as microchimerism (Mc). Haig recently proposed that maternal acquisition of fetal Mc from an earlier birth could affect the length of time before a subsequent birth (IBI), for example by interfering with implantation of subsequent embryos, thereby maximizing maternal resources for him/herself.[1,2] We hypothesized that for a 2nd child the maternal environment would be most hospitable if the child was HLA-identical with his/her older sibling (shorter IBI) and least hospitable if fully HLA-mismatched (longer IBI), with HLA haploidentical siblings of intermediate IBI. **Methods:** We HLA-genotyped 242 children of 121 healthy mothers, and 96 children of 48 women with systemic sclerosis (SSc). All children were 1st born (C1) or 2nd born (C2). Haplotypes were assigned after determining HLA-DRB1, DQA1 and DQB1 alleles by PCR and probe-specific hybridization, with additional HLA class I typing when needed to assign haplotypes. C2 was classified as HLA-identical (ID), haploidentical (Hap) or fully mismatched (MM) with C1. IBI was calculated for each family. Kruskal Wallis and Mann Whitney were used to compare IBI according to sibling HLA relationships. **Results:** All 169 families were combined for analysis as IBI did not differ between healthy and SSc families. Median IBI (in days) was shortest in the ID group (796 [interquartile interval 626-1086]) and longest in MM (1068 [747-1323]), with Hap intermediate (961 [701-1482]), P=0.03. Pairwise comparisons showed significant differences between ID vs MM (P=0.03) and ID vs Hap (P=0.02) but not between Hap vs MM (P=0.76). Within the Hap group, IBI did not significantly differ considering whether the shared haplotype was maternally or paternally inherited (P=0.67). **Conclusions:** IBI differed significantly according to whether C2 was HLA-identical, mismatched or haploidentical with C1 in normal and SSc families. The finding of shorter IBI between HLA-identical and longest between mismatched siblings supports the hypotheses proposed by Haig. To our knowledge this is the first HLA haplotype analysis of the concept of sibling rivalry (or cooperation). [1]Haig D, Evol Med Public Health 2014, 12-17. [2]Haig D, Chimerism 2014, 1-3.

2864S

The genetic basis of preeclampsia in a high altitude indigenous Andean population. *C.R. Gignoux[1], K. Sandoval Mendoza[1], G.L. Wojcik[1], A. Moreno Estrada[1], P. Ortiz-Tello[1], A.F. Adams[1], C. Eng[2], S. Huntsman[2], V. Villanueva Dávalos[3], J. Manzaneda[4], F. Manzaneda[4], M. Hurtado[5], V. Villegas[5], E. González Burchard[2], C. Gallo[5], L. Enrique Lencinas[3], J.C. Baker[1], C.D. Bustamante[1].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA; 3) Dept of Obstetrics and Gynecology, Hospital Regional "Manuel Núñez Butrón", Puno, Peru; 4) Universidad Nacional del Altiplano, Puno, Peru; 5) Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia.

Preeclampsia is the largest killer of pregnant women worldwide. It is a disorder characterized by the sudden onset of maternal hypertension, edema, and proteinuria, with a global prevalence of 3-8% of all pregnancies. The condition in particular has elevated rates in Hispanic/Latino women. Preeclampsia also poses a risk to newborns, particularly in areas of the world where resources are limited. The heterogeneity of the disorder has posed a challenge in our understanding the molecular basis of the preeclampsia. However, the trait is known to be exacerbated in high altitude regions such as the Andes. In order to investigate the genetic basis of preeclampsia and its relationship to altitude we have begun a collaboration with researchers and clinicians in Puno, Peru, a city nearly 4,000 meters above sea level. The local population is primarily of Quechuan and Aymaran descent. Based on population genetic analysis then we expect these groups to have inherited a unique pattern of genetic risk factors from other populations previously studied. We have collected data on almost 500 individuals in affected trios (including maternal, paternal, and cord blood) and population controls, with sections of placenta alongside pertinent medical records. The trio design allows for measuring the contributions of maternal, paternal and proband effect on the expectant mother. We have generated genome-wide data at over 800,000 sites across the genome with the Affymetrix Axiom LAT array on all samples to date. Most individuals have very high levels of Native American ancestry (median value >98% as measured by ADMIXTURE). Using trio-aware genome-wide association analyses we identify several top candidates, including a group of SNPs in LD over a cluster of blood clotting factor genes on 13q34 (*PROZ*, *F7*, *F10*), with a relatively high odds ratio for preeclampsia (OR 3.4, 95% CI 2.0-5.7,$p<4\times10^{-6}$). Although this region has not been previously identified in genome-wide studies, preeclampsia is known to cause changes in clotting levels and blood thinners can be used as preventative treatment. PROZ in particular has been extensively studied as a biomarker in multiple pregnancy disorders. This finding of a novel association in a relevant biological pathway and other genetic discoveries in this study demonstrates the increased power of relevant study design and underscores the importance of more genetic analysis in diverse, understudied populations worldwide.

2865S

Molecular and pathological abnormalities in placentas of pregnancies complicated by preeclampsia. *P.A. Ayala-Ramírez[1], R. García-Robles[2], M. Olaya[3], J.L Rodriguez[3], T. Buitrago[2].* 1) Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogotá, Bogotá D.C., Colombia; 2) Instituto de Investigación en Nutrición, Genética y Metabolismo, Universidad El Bosque, Bogotá D.C., Colombia; 3) Departamento de Patología, Pontificia Universidad Javeriana-Hospital Universitario San Ignacio, Bogotá D.C., Colombia.

Preeclampsia (PE) is a hypertensive disorder of pregnancy of unknown etiology with impact in maternal and fetal health. Its clinical manifestations occur in the second half of pregnancy with a world frequency of 2- 8% and in developing countries occurs in 10%. In Colombia, PE causes 42% of maternal mortality. It is proposed that an alteration in placental homeostasis may be involved in the pathophysiology of this disease. The aim of this study was to compare levels of mRNA and protein of Tissue Factor (F3) and Thrombomodulin (THBD) as well as histopathological findings in placentas with PE. We studied 18 term placentas of patients with diagnosis of PE and 18 term placentas from pregnancies without any complication and healthy newborn. It conducted an evaluation by an expert blinded pathologist of the presence of thrombosis, fibrin deposits, decidual arteriopathy, necrotic villi, leukocyte stasis, hyperplasia of syncitiotrophoblast and edema. The assessment of mRNA and protein levels of F3 and THBD was performed by Real time PCR and ELISA, respectively. The results showed statistically significant differences between cases and controls for these variables: decidual arteriopathy (p=0.027), necrotic villi (p=0.001) and hyperplasia of syncitiotrophoblast (p=0.0017). In addition, it found statistically significant differences in levels of F3 mRNA (p=0.011) and protein (p=0.0001) as well as in mRNA THBD (p=0.0001) and protein (p=0.0001) levels, with an increase in case group. We found an abnormal expression of F3 and THBD with increased levels of protein and mRNA in placenta of pregnancies complicated by PE, suggesting a possible role of these molecules in the pathophysiology of this disease. Also, these results agree with the existence of alterations in hemostatic mechanisms and hystopathological changes in placenta of pregnancies affected by PE.

2866S

A 'conditional-ON' mouse model of Fibrosyplasia Ossificans Progressiva (FOP). *A.N. Economides[1,2], L. Huang[2], L. Xie[2], K. Nannuru[2], K. Feeley[2], T. Persaud[2], P. Yang[2], V. Idone[2], A. Lee[3], P. Yu[4], C. Schoenherr[2], S.J. Hatsell[2], A.J. Murphy[2].* 1) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY; 2) Regeneron Pharmaceuticals., Tarrytown, NY; 3) National Center for Advancing Translational Sciences, NIH, Rockville, MD; 4) Brigham and Women's Hospital, Boston, MA.

FOP (MIM 135100) is an autosomal dominant disorder characterized by early onset, episodic and progressive ossification of skeletal muscle and associated connective tissue (Pignolo et al., 2011). FOP is driven by mutations in the intracellular domain of ACVR1 (also known as ALK2), with the great majority altering Arginine 206 to Histidine (R206H). In order to enable the development of therapeutic approaches for FOP, as well as mechanistic studies of the disease process, we have engineered a Cre-regulated 'conditional-ON' allele of ACVR1[R206H] in the mouse - *Acvr1*$^{[R206H]COIN}$ - using a FlEx- (Schnutgen et al., 2003) or COIN-like design (Economides et al., 2013). A conditional approach was necessitated because an unregulated *Acvr1[R206H]* knock-in allele cannot be transmitted through the germline in mice (Chakkalakal et al., 2012). *Acvr1*$^{[R206H]COIN}$ was generated by introducing the R206H mutation in exon 5 of *mmuAcvr1*, placing the mutated exon and flanking intronic sequence in the antisense strand. The corresponding region from human ACVR1 was inserted into in the sense strand of *mmuAcvr1* upstream of the inverted exon. To enable Cre-dependent replacement of the wild type exon 5 with its R206H version, the introduced human sequence was flanked by a 'FlEx' array of *lox*2372 and *lox*P sites in parallel orientation, whereas the mutant exon is followed by *lox*2372-*lox*P placed in antiparallel orientation with the *lox*2372 and *lox*P sites flanking human exon 5. This arrangement allows Cre-mediated inversion of the mutant exon into the sense strand and simultaneous deletion of the wild type human exon 5. Body-wide activation of the FOP allele in *Acvr1*$^{[R206H]COIN/+}$; *Gt(ROSA26)Sor*$^{CreERt2/+}$ adult mice resulted in progressive ossification resembling FOP. Heterotopic ossification was evident by μCT on the axial skeleton, as well as long bones, as early as 2 weeks after dosing with tamoxifen, whereas untreated mice were normal, displaying no abnormal bone formation. Heterotopic ossification was spontaneous, and did not require experimentally induced inflammation. The FOP phenotype was obtained in adult mice, enabling the uncoupling of the process observed in these mice from development. Heterotopic ossification was prevented by treatment with the ACVR1 inhibitor LDN-212854 (Mohedas et al., 2013), demonstrating that this physiologic mouse model of FOP can be used for testing candidate therapeutic regimens. This and other novel therapeutic regimens will be presented in detail.

2867M

Knock-in human FGFR3 achondroplasia mutation as a mouse model for human skeletal dysplasia and potential therapy evaluation. *Y. Lee[1], S. Chen[2], Y. Chen[2, 3].* 1) Institue of Cellular and Organismic Biology, Taipei, Taiwan; 2) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 3) Department of Pediatrics, Duke University Medical Center, Durham, NC.

Activating mutations of fibroblast growth factor 3 (FGFR3) lead to several autosomal dominant human skeletal dysplasias with varying degrees of skeletal deformities. Achondroplasia, the most common genetic dwarfism in human, results from a point mutation, G380R, in FGFR3. To facilitate the study of skeletal dysplasias and to assess the potential therapeutic approaches designed to targeting the activating human FGFR3, we have generated a mouse model of which the endogenous mouse FGFR3 gene has been replaced with human FGFR3 G380R cDNA (Fgfr3hG308R). Mice carry the heterozygotes (Fgfr3$^{-/hACH}$) and homozygotes (Fgfr3$^{hACH/hACH}$) for the human FGFR3 G380R mutation have varying degrees of disproportionate, proximal shortening of the limbs, mid-face hypoplasia and relative macrocephaly due to a failure in endochondral proliferation and differentiation. This mouse model offers a generally applicable approach for the assessment of the potential therapeutic approaches designed to the specific protein targets for human diseases.

2868T

Retinal vascular lesions associated with mutations in *Col4a1*. *M. Alavi, D. Gould.* Department of Ophthalmology, University of California, San Francisco, School of Medicine, San Francisco, CA.

Type IV collagen alpha 1 (COL4A1) is a major component of almost all basement membranes and mutations in *COL4A1* cause multisystem disorders in humans and mice. Notably, patients and mice harboring mutations in *COL4A1* exhibit ocular defects including anterior segment dysgenesis, optic nerve hyperplasia and retinal vascular tortuosity. Here we present our findings studying the retinal phenotype of mice carrying a dominant-negative *Col4a1* mutation (*Col4a1*$^{\Delta ex41}$).

We examined retinas of *Col4a1*$^{+/\Delta ex41}$ C57BL/6J.129S1/SvImJ mice and *Col4a1*$^{+/+}$ littermate controls at different ages from postnatal day (P) 21 to P720 *in vivo* using fluorescein angiography (FA), fundoscopy and optical coherence tomography. Histology, immunohistochemistry and quantitative PCR analyses complemented our analyses.

FA showed fully penetrant retinal vascular tortuosity and abnormal ramification. Retinal examinations revealed serous chorioretinopathy, retinal hemorrhages, fibrosis or signs of pathogenic angiogenesis with anastomosis of the choroid and the retinal vasculature in approximately 80% of all *Col4a1*$^{+/\Delta ex41}$ eyes at various ages. Retinal hemorrhages and anastomosis were observed as early as P21. We assessed expression levels of candidate genes involved in angiogenesis and found increased vascular endothelial growth factor (VEGF) expression in *Col4a1*$^{+/\Delta ex41}$ retinas compared to *Col4a1*$^{+/+}$ retinas.

Our findings suggest that patients carrying mutations in *COL4A1* may be at risk for sudden vision loss resulting from retinal vascular insults. What is more, retinal exams may help to identify individuals that are at greatly increased risk of suffering from intracerebral hemorrhages. Elevated VEGF expression may be a primary cause or consequence of vascular lesions that then leads to further retinal damage. Currently VEGF is a major target for preservation of vision in patients with age-related macular degeneration. This suggests that anti VEGF therapy may also be an effective strategy to prevent pathogenic vascular defects that result from *COL4A1* mutations.

2869S

Inflammatory demyelination in a duplication mouse model of Pelizaeus-Merzbacher Disease. *G.M. Hobson[1,2], L. Sakowski[1,2], K. Clark[1,2], K. Sperle[1].* 1) Nemours Biomedical Res, A I duPont Hosp Children, Wilmington, DE; 2) Dept Biological Sciences, University of Delaware, Newark, DE.

Pelizaeus-Merzbacher disease (PMD [MIM 312080]) is an X-linked leukodystrophy affecting the myelin sheath in the central nervous system (CNS). It is caused by alterations in the proteolipid protein 1 gene (*PLP1* [MIM 300401]) that encodes the most abundant protein in CNS myelin. Our lab has developed the *Plp1*dup mouse model that is a genetically faithful model of the most common cause of PMD, a genomic duplication. Previous characterization on 129/Bl6 mixed background mice has shown disorganized myelin in white matter tracts, abnormal mRNA levels, and impaired gait. Further characterization on an inbred C57Bl/6 strain using the Noldus CatWalk has shown that two major phases of the gait cycle, stand and swing, are affected. *Plp1*dup mice also demonstrated a decreased latency to fall in accelerating trials on the Stoelting Rotarod, indicating a progressive effect on motor coordination and balance. Light microscopy on brain sections of *Plp1*dup mice showed degenerating nerve fibers and abnormal myelin organization with evidence of gliosis and activated microglia, suggesting an inflammatory response. Analysis of markers of gliosis and inflammation in brain mRNA showed upregulation of several key inflammatory markers, including TNF-α, and GFAP. Global transcriptome profiling revealed significant upregulation of 45 genes, many of which are involved in an inflammatory response. Further characterization of the *Plp1*dup mouse model will allow us to better understand the underlying mechanisms for PMD and develop rational treatments for a currently untreatable disease.

† Trainee Award Finalist

2870M

***Poc1a*, a component of the centriole and cilia, causes skeletal dysplasia and male infertility: a mouse model.** *K.A. Geister[1], M.L. Brinkmeier[2], D.L. Burgess[3], J. Cavalcoli[4], L.Y. Cheung[2], J.M. Oatley[5], M. Oatley[5], J. Wendt[3], S.A. Camper[1, 2].* 1) Graduate Program in Cellular and Molecular Biology, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 3) Roche-Nimblegen, Inc., Madison, WI; 4) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 5) School of Molecular Biosciences, College of Veterinary Medicine, Washington State University, Pullman, WA.

We report the identification of the lesion responsible for skeletal dysplasia and male infertility in the spontaneous, recessive mouse mutant *chagun* (1). We carried out regional sequence capture and determined that *Poc1a* (protein of the centriole 1a) is disrupted by the insertion of a processed cDNA into exon 8, with a target site duplication suggestive of LINE-1 mediated insertion. This disruption causes exon skipping, but the reading frame is maintained, and the mutant protein is detected in tibial extracts, suggesting that this is a partial loss of function allele. A BAC transgene containing *Poc1a* is sufficient to correct the growth and testis defects. Embryonic fibroblasts from mutant embryos have impaired cilia formation and multipolar spindles. The defects in cell division and cilia likely account for the disorganized proliferative zone of the epiphyseal growth plate because chondrocytes are not polarized properly and undergo increased apoptosis. This explains the mechanism underlying primordial dwarfism in humans with two different mutations of *POC1A* (2, 3). Male infertility is caused by arrested spermatogenesis at the pachytene stage and progressive germ cell loss. Spermatogonial stem cell transplantation studies reveal that *Poc1a* is essential for normal function of both Sertoli cells and germ cells. We hypothesize that partial POC1A function is necessary for viability in mammals. Mutations in a number of genes associated with centrosome function are now known to cause skeletal dysplasias and primordial dwarfisms. This mouse model will provide information about the global functions of POC1A, clarify how centrosome dysfunction leads to defects in skeletal growth and development, and serve as a means to uncover potential therapies. 1. Cha et al., Skeletal dysplasia and male infertility locus on mouse chromosome 9. Genomics 83:951-960, 2004. 2. Shaheen et al. *POC1A* truncation mutations causes a ciliopathy in humans characterized by primordial dwarfism. Am J Hum Gen 91:330-336, 2012. 3. Sarig et al. Short stature, onychodysplasia, facial dysmorphism, and hypotrichosis syndrome is caused by a *POC1A* mutation. Am J Hum Genet 91:337-342, 2012.

2871T

Using electroporation as a model of degeneration/regeneration to investigate the regenerative potential in neuromuscular disorders (NMD). *M. Vainzof, C.F. Almeida, R. Ishiba, A. Martins-Bach, A.L.F. Santos, L. Nogueira.* Human Genome Research Center, University of Sao Paulo, São paulo, Sao Paulo, Brazil.

After injury, the healthy muscle is able to repair and grow new fibers. However, in neuromuscular diseases, the process of regeneration is not completely efficient and its constant activation leads to exhaustion of this ability. Understand all the steps involved in muscle degeneration and regeneration is of major importance for a better comprehension of the genetic muscle diseases. Mouse models for NMD are good molecular models, but not always they reproduce the histopathological and phenotypic characteristics associated to the mechanisms of the respective disease. Protocols for in vivo gene therapies have used electrotransfer to optimize transfection efficiency. However, most of them have reported histological evidence of muscle degeneration and complete regeneration within 15 days. Therefore, we are proposing the use of this model of very fast muscle degeneration and regeneration processes to characterize different patterns of muscle regeneration in NMD. We are studying electroporated lower limb of normal mice and dysferlin deficient SJL mouse model for Limb Girdle Muscular Dystrophy type 2B, with the main objective to study the pattern of alterations along time, both through histological analysis and quantitative mRNA expression of genes involved in the degeneration/regeneration pathways. The animals were anesthetized and received 8 pulses of 100 V, duration of 20 milliseconds and interval of 0.5 second at the calf. Each group of animals was analyzed after 3, 5, 10, 15, 21 and 30 days. In control animals, histological analysis showed small areas with foci of degeneration starting at d3 and d5, and signals of regeneration starting at d5. After d10, large areas of centrally nucleated fibers were observed, continuing in d21 and d30. In SJL model, the degeneration was more intense in d3 and d5, and was also observed in d10. In d21 and d30, large areas of centronucleated fibers were observed. Quantitative mRNA expression of the gene Myod, Myf5 and Myog, showed the activation of the regeneration genes in SJL mice, since d3 up to d10. This new system of acutely induced degeneration and regeneration in mouse models is useful to study and quantify alterations in satellite cells and regeneration impairment in some mildly affected myopathic model, in which the histopathological pattern of the muscle is not comparable to the pattern observed in human diseases. FAPESP-CEPID,CNPq-INCT, CAPES-COFECUB, FINEP.

2872S

Definitive Implication of Innate Immunity in the Pathogenesis of Scleroderma. *E. Gerber[1], J. Beckett[1], D. Huso[2], H. Dietz[1,3].* 1) Institute for Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Comparative Medicine, Johns Hopkins University, Baltimore, MD; 3) Howard Hughes Medical Institute, Chevy Chase, MD.

The most common and severe form of scleroderma (systemic sclerosis; SSc) shows sporadic adult onset of skin and visceral fibrosis in association with autoantibody production and other markers of autoinflammation. In contrast, stiff skin syndrome (SSS) shows congenital dermal fibrosis caused by heterozygous mutations in the domain of fibrillin-1 that mediates cell-matrix attachment via integrin binding. Both SSS and SSc show excessive deposition of fibrillin-1 that serves to concentrate the profibrotic cytokine TGFβ. We previously showed that knock-in mouse models of SSS recapitulate dense dermal fibrosis and show all of the autoinflammatory manifestations of SSc including antinuclear and anti-topoisomerase-1 antibodies, Th2 and Th17 skewing and B cell activation in the skin in association with recruitment and activation of plasmacytoid dendritic cells (pDCs), a cell type that performs a surveillance function for pathogens and responds via IL6 and IFNα production upon presentation of RNA or DNA to TLR7/9 receptors, respectively. All therapeutic interventions that prevent fibrosis in SSS mice, including agents that downregulate integrin expression on pDCs and TGFβ and ERK inhibitors, associate with prevention of pDC recruitment and activation. This prompted the question regarding the relative contribution of innate vs. adaptive immunity in disease pathogenesis, an unaddressed question in SSc as well, and the mechanism of cross-talk between the matrix and pDCs. Using a transgenic allele that expresses diphtheria toxin receptor specifically on pDCs, we now show that postnatal pDC depletion is sufficient to prevent fibrosis in SSS mice, while introduction of the null state for Rag2 (no T or B cells or antibodies) had no effect on disease expression. These data unequivocally point to a predominant, if not exclusive, contribution of innate immunity. In other autoimmune diseases such as lupus and psoriasis, abnormal protein aggregates present nucleic acids to TLRs, with the naturally occurring antimicrobial peptide LL37 serving as a critical intermediate. We wondered whether the macroaggregates of fibrillin-1 might be initiating similar events in SSS (and perhaps SSc). In keeping with this hypothesis, introduction of haploinsufficiency for Ll37 or treatment with a synthetic TLR7/9 blocking peptide abrogated fibrosis in SSS mice. These data offer many novel therapeutic strategies in scleroderma and perhaps other more common presentations of fibrosis.

2873M

Human-Mouse:Disease Connection, new pathway to discovery. *J.T. Eppig, C.L. Smith, P. Frost, K. Stone, L.E. Corbani, J. Campbell, K.L. Forthofer, J.E. Richardson, S.M. Bello.* Jackson Laboratory, Bar Harbor, ME.

The Human-Mouse:Disease Connection (www.diseasemodel.org) is a translational tool affording simultaneous access to human-mouse genomic, phenotypic, and genetic disease information. Researchers can explore phenotypes and disease relationships, identify candidate genes, and evaluate mouse mutants displaying a spectrum of indicative phenotypes. Within the web display, links are provided to supporting mouse model publications and to repositories supplying mouse resources. Three primary approaches give users flexibility to search from a human or mouse perspective, using (1) genes or gene IDs for either species, (2) genome location(s) from either species, and (3) mouse phenotype or human disease terms. Data also can be uploaded from VCF files. Thus, exploration can begin with a single gene or set of genes, a region for QTL, multiple deletion regions, or using phenotype/disease searches, such as “Crouzon Syndrome”, “neurofibromatosis”, or “cardiomyopathy”. All search methods initially return an interactive grid that presents a visual overview of results facilitating comparison of phenotypes and diseases across multiple genes, phenotypes, and diseases. The grid features color cues reflecting depth of annotated human and mouse data, and grid cells are active links leading to more detailed information, including availability of mouse models from repositories worldwide. Alternate web page displays with gene and disease-focused information are one click away. Integration of mouse and human data related to gene homologs, genomic locations, mutations, phenotypes, and diseases is ongoing in the Mouse Genome Informatics resource (MGI, www.informatics.jax.org), the primary data source for the Human-Mouse:Disease Connection portal. These data include over 1300 OMIM-defined human genetic diseases with at least one experimentally defined mouse model and mouse phenotypes for more than 51,600 unique genotypes. Data (to be released in June 2014) from the systematic screens of knockout mutations analyzed in the International Mouse Phenotype Consortium (IMPC) project will further augment phenotype-genotype associations for mouse mutations. We will discuss the challenges of human and mouse data integration and new developing capabilities of the Human-Mouse:Disease Connection portal to ease translation between human and mouse data and to promote best matching of human disease and mouse model resource data. Supported by NIH grant HG000330.

2874T

Social and maternal behaviours are affected by a mutation in *Gtf2ird2* in Williams-Beuren Syndrome mouse model system. *N. SHARMIN[1,2], M. Tassabehji[2], R. Hager[1].* 1) Faculty of Life Sciences, University of Manchester, Manchester, Lancashire, United Kingdom; 2) 1.Genetic Medicine, Faculty of Medical and Human Studies, St Mary's Hospital, University of Manchester, Lancashire, United Kingdom.

Human social behaviour is affected by both environmental and genetic factors. The human genetic disorder Williams-Beuren Syndrome (WBS) is characterised by neurological deficits that impact social and cognitive behaviour. Our hypothesis is that deletion of the *Gtf2ird2* gene, from the Gtf2i-family of transcription factors, affects social and neuropsychological development and consequently affects maternal behaviour during early development. By phenotyping maternal behaviour in a single-gene knockout mouse model of WBS, we show, for the first time, that deletion of *Gtf2ird2* influences key maternal and social behaviours in homozygous and heterozygous animals, affecting offspring development. In addition, *Gtf2ird2* heterozygous mothers exhibit high rates of filial cannibalism compared to either knockout or wild type mothers. These results show that this single gene mutation has an important role in the hyper-social behavioural symptom seen in WBS patients. The results are in line with existing literature on *Gtf2i* gene family.

2875S

Canine developmental disorder maps to the critical region of human 22q11.2 deletion syndrome. *M. Hytönen[1,2], A. Lappalainen[3], H. Lohi[1,2].* 1) Molecular Neurology, Research Programs Unit and Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland; 2) Folkhälsan Institute of Genetics, Helsinki, Finland; 3) Equine and Small Animal Medicine, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland.

Dogs have emerged as clinically and genetically relevant large animal models for human inherited disorders. We present here a genetic study of a congenital canine syndrome characterized by typical craniofacial features, patellar subluxation, microphthalmia and mental disturbances. We have performed genome wide association study to map the disease to canine chromosome 26. The associated locus overlaps human DiGeorge syndrome region characterized by a 22q11.2 deletion. Our ongoing targeted resequencing efforts aim to discover the causative mutation in the critical region. Comparative phenotypic analyses reveal similarities in canine and human phenotypes and genetic characterization of the canine locus may assist in the genetic dissection and understanding of the DiGeorge syndrome in both species. This study also enables the development of a genetic test for breeding purposes in dogs.

2876M

A transgenic zebrafish model for facioscapulohumeral dystrophy. *A. Lek[1,2], H. Mitsuhashi[1,2], F. Rahimov[1,2], C. Mosimann[1,2], L. Zon[1,2], L. Kunkel[1,2].* 1) Genetics, Boston Children's Hospital, Boston, MA; 2) Pediatrics and Genetics, Harvard Medical School, MA.

Facioscapulohumeral muscular dystrophy (FSHD) is one of the most common forms of muscular dystrophy, characterized by asymmetric weakness of the facial, shoulder and upper arm muscles, accompanied by hearing loss and retinal vasculopathy. Although the causative gene for FSHD remains controversial, the primate specific retrogene, DUX4, is a leading candidate. There is currently no mammalian model that recapitulates the human FSHD pathology. Previously, our lab has shown that misexpression of very low levels of human DUX4 in zebrafish development recapitulates the phenotypes seen in FSHD patients. It was demonstrated that microinjection of human DUX4 mRNA into zebrafish eggs caused asymmetric abnormalities of the eyes and ears, and disorganization of fin and trunk muscles. Using a tamoxifen-controlled CreER^{t2}-loxP system, we have now generated a transgenic DUX4 line that successfully reproduces the mosaic, low-level expression of DUX4. We show that activating DUX4 expression during development results in a degenerative muscle phenotype by day 7 post-fertilization. This stable line will enable us to control when, where and how much DUX4 is expressed to best model FSHD pathogenesis in zebrafish, allowing for better functional studies.

2877T

Role of palmitoylation in alopecia and skin abnormality in genetic deficiency of Zdhhc13 in Mice. *K. Liu[1,2], Y. Chen[3], L. Shen[1], A. Saleem[1], I. Song[1,5], L. Chen[1], Y. Chen[3], J. Wu[1], J. Yen[1], Y. Chen[1,2,4].* 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 2) Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan; 3) Institute of Chemistry, Academia Sinica, Taipei, Taiwan; 4) Department of Pediatrics, Duke University Medical Center, Durham, NC, USA; 5) Graduate Institute of Life Sciences, National Defence Medical Center, Taipei, Taiwan.

Palmitoylation, a post-translational modification of 16-carbon of palmitic acid, regulates protein folding, trafficking, stability, and protein-protein interactions. Defects of palmitoylation are associated with certain human neurological disorders and cancers. We previously reported mice with genetic deficiency in palmitoyl acyl transferase, Zdhhc13, had alopecia and hyperkeratosis, amyloidosis and osteoporosis. Here we investigated the pathogenic mechanism of hair and skin abnormalities. Mutant mice had hair loss prematurely occurred in each of Catagen of hair cycle, instead of Telogen and resulted in cyclic alopecia phenotype. Scanning electron microscope showed the broken and dilapidated hair cuticle structure. Furthermore, immunohistology showed hyperproliferation of the epidermis, and enhanced and delayed terminal differentiation of the keratinocytes (also known as cornification or keratinization). Biochemically, we found a palmitoylation substrate, cornifelin, was not present in the epidermis and hair cuticle. We also identified the 5 palmitoylation sites of cysteine on cornifelin. Interestingly, overexpression of cornifelin in XB2 cell, a mouse teratoma keratinocyte cell line, resulted in reduced amount of cornifelin protein when each of 2 predicted palmitolyated-cysteines of C-terminal was mutated to serine. Our data suggested that 1. Cornifelin is a specific substrate of ZDHHC13. 2. Defective palmitoylation causes loss of cornifelin protein. 3. Because of unique localization of cornifelin in hair cuticle and in skin cornified layer, it is likely that absence of cornifelin in hair and skin of mice causes the defective hair cuticle structure and disruption of cornification which results in cyclic alopecia and hyperkeratosis.

2878S

Mouse Models: Effective Therapeutics Development and Comparative Genomics. *C.L. Smith, S.M. Bello, J.T. Eppig, Mouse Genome Informatics Staff.* Mouse Genome Informatics, Jackson Laboratory, Bar Harbor, ME.

Advances in genome sequencing technology have revolutionized human molecular genetics. Coupling sequence, phenotypes, pathogenicity variant calls, copy number variation, and other biological evidence with use of mouse model data can accelerate the identification and testing of therapeutic candidates. Examples include use of losartan to treat aortic aneurysms in Marfan syndrome and other heritable conditions, enzyme replacement therapy in Pompe Disease and ipilimumab, an antibody therapy that extends life in patients with metastatic melanoma. Other potential therapies developed using mouse model data and awaiting clinical trials include progesterone antagonist therapy (Lonaprisan) in Pelizaeus-Merzbacher hypomyelinating disease and C-type natriuretic peptide (CNP) analog (BMN 111) in the treatment of Achondroplasia.

Exome sequence data from patients with diseases with unknown etiology have provided several hundred new candidate genes for many diseases, and targeted re-sequencing technologies have provided gene variation information on dozens of these candidate genes in tens of thousands of individuals with high specificity and sensitivity. Comparison of gene, phenotype and genetic disease data provided by mouse models aids in the refinement of potential therapeutic targets. Mouse Genome Informatics (MGI, www.informatics.jax.org) provides data sets to correlate mouse phenotype with human clinical signs and symptoms. MGI catalogs all mouse mutant alleles with genotypes annotated to phenotype and OMIM disease model descriptions, and includes links to other supporting gene information including sequence, polymorphism, spatiotemporal expression, genomic location, biochemical function and process, sub-cellular topology, and mammalian gene homology. Standardization of nomenclature and application of bio-ontologies, including the Mammalian Phenotype (MP) Ontology, ensure that data are consistently annotated, making robust data mining possible. The model data at MGI aids in understanding biochemical pathways and pathological processes, and thus assists in determining the underlying mechanisms of human genetic disease. We will show MGI query results utilizing comparative human and mouse data that assist in finding novel candidate genes for human disease, and we will provide examples of therapeutic targets identified using data generated from mouse models. Supported by NIH grant HG000330.

2879M

Rapid Approaches in Functional Validation of Candidate Disease Genes Identified from Structural Rearrangements and Next-Generation Sequencing. *R. Greenlees[1,2], S. Yousoof[1,2], E. Semina[3], R.V. Jamieson[1,2].* 1) Eye Genetics Research Group, Children's Medical Research Institute, The Children's Hospital at Westmead, Save Sight Institute, Sydney, Australia; 2) Sydney Medical School, University of Sydney, Australia; 3) Medical College of Wisconsin, Milwaukee, US.

The application of next-generation sequencing (NGS) has provided unparalleled opportunities for gene discovery; however, there is a bottleneck in functional validation of the candidate disease genes. This study presents functional approaches that can be utilized to elucidate the role of candidates using the example of a novel candidate disease gene in eye disease. The candidate disease gene was identified in a patient presenting with anterior eye and lens abnormalities, and a de novo balanced chromosomal translocation disrupting the candidate disease gene. NGS in an extended cohort of patients with genetic eye disease has identified other patients with likely pathogenic variants in this gene. The candidate disease gene contains domains with predicted roles in cell adhesion. 2-dimensional epithelial cell assays showed colocalisation of the candidate with markers of adherens and tight junction proteins, as well as cytoskeletal filamentous actin (F-actin). In another patient with anterior segment eye disease, NGS led to a predicted pathogenic missense variant in this gene, and transfection of a construct with this variant led to abnormal F-actin localisation in the 2D assay. A 3-dimensional assay using Caco2 cells resulted in formation of epithelial cysts and shRNA knockdown of the candidate disease gene resulted in abnormal cysts with multilayering and aberrant expression of polarity markers. Animal models, including zebrafish and mouse, were utilised to investigate the function of the candidate disease gene in vivo. Decreased expression of the candidate disease gene in both these models produced abnormal lens development. The knockout mice exhibited a cataract phenotype and demonstrated abnormalities in the expression of cell adhesion and polarity markers at both embryonic and postnatal ages. The cell-based assays we have utilised provide a rapid, inexpensive model system for functional validation of novel candidate disease genes and variants affecting the lens, identified from NGS discovery projects. The animal-based approaches have provided more detailed mechanistic understanding of the role of this gene in epithelial morphogenesis.

2880T

Natural genetic modifiers of autosomal dominant retinitis pigmentosa and ER stress. *C.Y. Chow, M.F. Wolfner, A.G. Clark.* Dept Molec Biol & Gen, Cornell Univ, Ithaca, NY.

Retinitis pigmentosa (RP) is a hereditary disease characterized by progressive loss of vision due to degeneration of rods and cones in the retina. Autosomal dominant retinitis pigmentosa (ADRP) makes up 30-40% of all RP cases. Dominant mutations in the *rhodopsin* gene (*RHO*) comprise 25% of all ADRP cases and represent the most common cause of RP. Dominant mutations in the *Drosophila melanogaster* ortholog of *RHO*, *Rh1* (or *ninE*), provide an important model for dissecting the pathophysiology of ADRP. The $Rh1^{G69D}$ mutation in *Drosophila* closely resembles many human mutations in *RHO*. This mutation results in a misfolded Rh1 protein that is retained in the endoplasmic reticulum (ER), and induces the ER stress response, leading to apoptotic cell death and retinal degeneration. Previous studies demonstrated that mutations in genes in the ER stress and apoptosis pathways can alter the phenotypic presentation of $Rh1^{G69D}$. However, these previous genetic screens relied entirely on loss-of-function (LOF) mutations. In the human population, it is unlikely that severe LOF mutations contribute appreciably to variability in ADRP phenotypes. We take advantage of natural genetic variation in *Drosophila* to identify dominant modifiers of $Rh1^{G69D}$ induced retinal degeneration. We crossed the $Rh1^{G69D}$ mutation into ~200 strains from the *Drosophila* Genetic Reference Panel (DGRP). The DGRP is a collection of wild-derived *Drosophila* strains that harbor genetic polymorphisms present in a natural population. To assess the effect of DGRP backgrounds in modulating the phenotypic impact of $Rh1^{G69D}$, we measured eye size to quantify the extent of degeneration and apoptosis. We found that eye size varied by more than ten phenotypic standard deviations, presenting a more than two-fold difference in eye size due to background modifiers. This dramatic phenotype allowed us to perform an association study to identify natural genetic polymorphisms that enhance or suppress the $Rh1^{G69D}$ retinal degeneration. We used functional studies to validate the novel candidate genes found through this approach. The candidate polymorphisms and genes nominated from this study may more accurately identify potential human modifiers of ADRP, as they are drawn from existing natural variation. These results have important implications for identifying modifiers of the ER stress response and identifying loci that modify human ADRP.

2881S

***AIPL1* mutation in Persian cats defines a new model for Leber's Congential Amaurosis.** *B. Gandolfi[1], H. Alhaddad[2], R.A. Grahn[3], D.J. Maggs[4], H-C. Rah[5], N.C. Pedersen[6], L.A. Lyons[1].* 1) Department of Veterinary Medicine & Surgery, University of Missouri - Columbia, Columbia, MO; 2) College of Science, Kuwait University, Safat, Kuwait; 3) Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California - Davis, Davis, CA USA; 4) Department of Surgical and Radiological Sciences, School of Veterinary Medicine, University of California - Davis, Davis, CA USA; 5) College of Medicine, Chungbuk National University, Chongju, Chungbuk Province, South Korea; 6) Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California - Davis, Davis, CA USA.

Leber's Congential Amaurosis (LCA) is a rare inherited eye disease that has onset at birth or early in life. LCA Type 4 (LCA4, MIM: 604393) is caused by mutations in the gene, *aryl-hydrocarbon-interacting protein-like 1 (AIPL1)* and is considered the severe end of the spectrum for LCA. Newborns with *AIPL1* mutations generally have severely impaired vision or blindness, nystagmus, and an abnormal or flat electroretinogram. A variety of mutations in *AIPL1* have been identified to cause LCA in humans, presenting with varying ages of onset and degree of cone - rod dystrophies. Persian cats have been demonstrated to have an autosomal recessive progressive retinal atrophy (PRA) similar to LCA. Onset of photoreceptor loss is approximately 5 weeks of age with severe loss by 16 weeks of age. The cats often have divergent strabismus but do not suffer from bilateral ectasia with central thinning of the cornea as the cat ages. A series of association studies, including linkage analysis and transmission distortion tests suggested that the casual gene for Persian PRA was localized to cat chromosome E1, which is homologous to human chromosome 17. A partial analysis of *AIPL1*, did not readily identify a mutation, as the cat genome assembly was in complete for *AIPL1*. Whole genome sequence to 30x coverage of a trio of cats, including one carrier and affected parent and one affected offspring, was performed using 350 and 550 bp PCR-free libraries and HiSeq 100bp paired end reads. Variant calling was performed using Freebayes by Maverix Biomics. A stop gain was identified at in *AIPL1* at c.577C>T producing a p.Arg193*, as predicted by the human sequence and knocking out approximately 33% of the normal protein. Data from the sequencing reads and an improved cat genome assembly supported the confirmation of the mutation in the extended pedigree of Persians with PRA. Two mouse models of *AIPL1* deficiency include the Aipl1-hypomorphic (h/h) mouse (with reduced Aipl1 levels and a relatively slow degeneration), and the Aipl1-null mouse (with no functional Aipl1 and a very rapid retinal degeneration). Gene therapies using the sc-Y733F-AAV2/8 viral vector have shown rescue vision loss in the null mouse model. Cats have a longer life span than mice and can be evaluated for repeated and longer term gene and stem cell therapy trials. The cat can now become an efficient and effective model for gene and stem therapies for LCA.

2882M

Genome-wide linkage analysis in conjunction with whole exome sequencing for identification of deafness-causing genes in multi-generational families. *M. Grati[1], R. Mittal[1], J. Qing[1,3], DH. Xie[3], Q. Ma[1], LL. Du[1], Q. Liu[1,4], QJ. Wang[4], SM. Yang[4], P. Dai[4], Y. Feng[5], D. Yan[1], S. Masmoudi[6], S. Blanton[2], XZ. Liu[1,2].* 1) Departments of Otolaryngology-Head and Neck Surgery, University of Miami Miller School of Medicine, Miami, Florida 33136, USA; 2) Dr. John T. Macdonald, Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL 33136, USA; 3) Department of Otolaryngology- Head and Neck surgery, Institute of Otology, the Second Xiangya Hospital, Central South University, Changsha, Hunan, 410011, China; 4) Department of Otolaryngology, Head and Neck Surgery, Chinese PLA, Beijing, 100853, China; 5) Department of Otolaryngology, Xiangya Hospital, Central South University, Changsha, Hunan, 410008, China; 6) Microorganisms and Biomolecules Laboratory, Centre of Biotechnology of Sfax, Sfax University, Sfax, Tunisia.

Hearing loss (HL) is the most common sensory defect, affecting 1 in every 500 newborns around the world which half of it is due to genetic causes. Nonsyndromic HL (NSHL) accounts for more than 70% of all hereditary HL. Gene isolation for NSHL has been successful in isolated populations and consanguineous families. However, even after the gene has been localized to a region on a chromosome, the process of positional cloning by Sanger sequencing to analyze each selected gene in the linked interval, is costly and time consuming and can be impractical, in the case of large candidate intervals. Now whole exome sequencing (WES) offers a valuable tool for the many challenges in the identification of mutations in heterogenic diseases such as deafness. In this study we report the genome-wide linkage in conjunction with WES analyses of a collection of 14 multi-generational families including 10 Chinese families with autosomal dominant NSHL and 4 consanguineous Tunisian families with autosomal recessive NSHL for finding deafness causative genes. The SureSelect human all exon 50Mb kit (Agilent Technologies) and the Hisq2000 instrument (Illumina) were used for WES and the Genomes Management Application (GEMapp; (https://secureforms.med.miami.edu/hihg/gem-app) was applied for data filtering. Computational functional prediction algorithms and conservation scores were also applied to evaluate the possible impact on protein function of the detected variations. By linkage analysis, we determined the chromosomal location of the disease to prioritize candidate genes for mutation screening. On average, 97% of RefSeq exons were captured and approximately, each exome in the 14 families had 97%, 85% and 70% of mappable bases of the Gencode defined exome represented by coverage of at least 2, 10 and 20 reads, respectively. By our filtering strategy under the autosomal dominant or recessive model, we have rapidly identified novel genes with homozygous or heterozygous missense variants common to affected siblings subjected to WES. Co-segregation of the variants with the disease phenotype is validated using Sanger sequencing in each family and In vitro studies to establish the pathogenetic nature of the variants are being performed. Our study shows that combination of genome wide linkage analysis with WES is a powerful strategy for identification of causative mutations in genetically heterogeneous condition such as deafness.

2883T

***ILDR1*: Novel mutation and a rare cause of congenital deafness in the Saudi Arabian population.** *K. Ramzan[1], K. Taibah[2], A. Tahir[1], N. Al-Tassan[1], A. Berhan[1], A. Khater[2], S. Al-Hazzaa[3], M. Al-Owain[4], F. Imtiaz[1].* 1) Department of Genetics, King Faisal Specialist Hospital & Research Centre, Riyadh, Riyadh, Saudi Arabia; 2) ENT Medical Centre, Riyadh 11333, Saudi Arabia; 3) Department of Ophthalmology, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 4) Department of Medical Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia.

Hearing impairment is the common human sensorineural disorder and is a genetically heterogeneous phenotype for which more than 100 genomic loci have been mapped so far. *ILDR1* located on chromosome 3q13.33, encodes a putative transmembrane receptor containing an immunoglobulin-like domain. We used a combination of autozygosity mapping and candidate gene sequencing to identify a novel mutation in *ILDR1*, as a causative gene for autosomal-recessive non-syndromic hearing loss (arNSHL) in a consanguineous Saudi family with three affected children. Autozygosity mapping identified a shared region between the affected individuals encompassing *ILDR1* on chromosome 3q13.12-3q22.1. Sequencing revealed homozygous 9 base pair duplication, resulting in an in-frame duplication of three amino acids p.(Asn109_Pro111dup). The mutation was segregating with the disease phenotype and is predicted to be pathogenic by SIFT and PROVEAN. The identified mutation is located in the immunoglobulin-type domain of the ILDR1 protein. In silico analysis using I-TASSER server and PyMOL offers the first predictions on the structural and functional consequences of this mutation. To our knowledge, this is the first ILDR1 mutation identified in a Saudi family. Identification of *ILDR1* mutation in only one of 100 Saudi familial and sporadic individuals with hearing loss suggests that this mutation is unique to this family and that *ILDR1* should be considered as a rare cause of congenital deafness among Saudi Arabian population. Our data also confirms the evidence for *ILDR1* allelic heterogeneity and expands the number of familial arNSHL-associated *ILDR1* gene mutations.

2884S

Homozygosity mapping of families with autosomal recessive intellectual disability and examination of WWOX, GFRA3, and PTBP1 genes. *A. Alkhateeb[1,2], S. Aburahma[2], W. Habbab[1].* 1) Qatar Biomedical Research Institute, Doha, Qatar; 2) Jordan University of Science and Technology, Irbid, Jordan.

Intellectual disability is a relatively common disorder affecting all populations. In Arab countries, where consanguinity has a high prevalence, there is an increase in autosomal recessive intellectual disability. We tried to identify the genetic causes of intellectual disability in four consanguineous local Arab families. Families were recruited in the study after giving informed consent and obtaining institutional review board approval. All families had multiple affecteds, at least one of the affecteds was a female. Samples were genotyped by high-density HumanOmniExpress (700K, Illumina Inc). Genotypes were analyzed by HomozygosityMapper and multiple regions of homozygosity were found in the four families. No copy number variants were found. A single overlap between two families harbored WWOX (WW domain-containing oxidoreductase) gene. Woxx was found previously to be mutated in mice with seizures and epilepsy and the two families had seizures and epilepsy as part of their phenotype. Sequencing WWOX gene detected three commons SNPs with no indication of pathogenicity. In the third family, two other candidate genes (GFRA3 and PTBP1) localized within the homozygosity interval were sequenced; no variants were found in the coding region. In conclusion, we recruited four local consanguineous families with multiple affecteds of intellectual disability. Those families showed different homozygosity intervals illustrating the heterogeneity of the phenotype. We examined the candidacy of WWOX, GFRA3, and PTPP1 genes within the homozygosity intervals but no pathologic variants were found. We are now in the process of doing whole exome sequencing for one patient of each family.

2885M

Santos syndrome is caused by homozygous mutation in *WNT7A*. *L.U. Alves[1], S. Santos[2], R.S. Thiele-Aguiar[1], P.A. Otto[1], R.C. Mingroni-Netto[1].* 1) Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil; 2) Departamento de Biologia, Universidade Estadual da Paraíba, Campina Grande, PB, Brazil.

Santos et al. (Am J Med Genet 2008) described a new syndrome in six relatives living in a remote area in Northeastern Brazil. This syndrome is characterized by fibular agenesis/hypoplasia, hypoplastic femora and grossly malformed/deformed clubfeet with severe oligodactyly, upper limbs with severe ungual hypoplasia/anonychia sometimes associated with mild bractactyly and occasionally pre-axial polydactyly. This syndrome, named as "Santos syndrome" (OMIM 613005), was classified as a distinct syndrome from other previously described conditions exhibiting fibular agenesis/hypoplasia. An autosomal dominant inheritance model with incomplete penetrance was suggested to be more likely by Santos et al. (2008), but autosomal recessive inheritance was not discharged. Linkage analysis was performed based on results of genotyping by SNP-array (500 K - Illumina/Agilent) in samples from the six affected individuals. After multiple point LOD score calculations, two candidate regions on chromosome 3 were found to be statistically significant under the dominant (k=0,324) and recessive model hypothesis, in 3p26.1-p25.2 and in 3p13-q12.3. The maximum values of LOD score calculated under recessive model hypothesis were 2,856 and 3,235 for these two regions, respectively. Under the dominant model, positive LOD scores were observed in the same regions, but with smaller values than in the recessive model. The best candidate gene in the mapped regions was *WNT7A*, in which homozygous mutations had already been associated with two other limb defect syndromes (Fuhrmann syndrome, #228930 and AARRS, #276820). Sequencing of *WNT7A* revealed a novel homozygous c.934G>A (p.Gly312Ser) mutation, in five of six affected individuals; the remaining affected individual in heterozygous as to the mutation, and his phenotype is accordingly much less severe than his relatives. The glycine 312 at WNT7A protein is highly conserved (Uniprot) and the mutation was predicted to be probably damaging with score of 1,0 by PolyPhen-2. Thus, Santos syndrome can be explained by mutation in the *WNT7A* gene. We hypothesize that the affected heterozygous individual either has a different condition or his phenotype results from a mild clinical manifestation of Santos syndrome.

2886T

Mutation Screening in *PRPF31* in an Autosomal Dominant Retinitis Pigmentosa (ADRP) Family with Incomplete Penetrance. *S. Bhatia[1], S. Goyal[1], I.R. Singh[2], V. Vanita[1].* 1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India; 2) Dr. Daljit Singh Eye Hospital, Amritsar.

Purpose: The objective of present study was to perform mutation screening in *PRPF31* in a four generation ADRP family with incomplete penetrance. **Material and Methods:** Present study involved a large four generation ADRP family with incomplete penetrance which was diagnosed and collected at the Dr. Daljit Singh Eye Hospital, Amritsar with six affected individuals and few asymptomatic carriers. Asymptomatic carrier passed the disease to next generations. Age of onset of the disease in this family was by birth. Ophthalmic examinations included visual acuity testing, intraocular pressure, fundus testing alongwith fundus photography and Optical Coherence Tomography (OCT) after pupil dilation and Electroretinography (ERG) testing. We undertook mutation screening in the *PRPF31* gene that is well reported to be linked with ADRP with incomplete penetrance. Bi-directional sequence analysis of amplified products of all the 14 *PRPF31* exonic regions including splice sites, was performed in this family. **Results:** Mutation screening revealed identification of a novel missense mutation in one of the exonic region of *PRPF31* in this family. The identified missense mutation segregated completely with the disease in all the six affected members of this family as well as carriers of the disease. However, 19 tested unaffected members of this family and 100 control samples from the same population didn't carry the identified *PRPF31* substitution, excluding its possibility as a polymorphism. **Conclusions:** Present study identifies a previously unreported mutation in *PRPF31*. These findings further demonstrate the role of *PRPF31* in relation to ADRP with incomplete penetrance.

2887S

Molecular Genetic Analysis in an Autosomal Recessive Retinitis Pigmentosa Family of Indian origin. *S. Goyal, V. Vanita.* Department of Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

Purpose: Present study aimed to identify underlying genetic defect in a nonsyndromal retinitis pigmentosa family (arRP) of Indian origin by whole genome homozygosity mapping. **Material and Methods:** It was a large 4 generation arRP family with 3 affected individuals and their parents being first cousins. Age of onset of the disease varied from 2 to 17 years. Ophthalmic examinations that included visual acuity, intra ocular pressure testing, electroretinography (ERG), ocular coherence tomography (OCT), fundus testing and fundus photography after pupil dilation, confirmed status of 3 individuals being affected by non-syndromic typical RP. Whole genome homozygosity mapping was conducted on 7 members (3 affected and 4 unaffected) using 8-sample HumanOmniExome Beadchip, consisting of >2,40,000 functional exonic markers. Genotyping data was subsequently used to identify the identical homozygous regions present in affected individuals using the online homozygosity mapping tool Homozygosity Mapper. Further, mutation screening in candidate genes at mapped homozygous region was done by bi-directional sequence analysis of the amplified products on an automated DNA sequencer. Sequences were assembled with SeqA6 sequencing analysis software (ver.6.0) and analyzed. **Results:** The homozygous region at 14q31.3 was identified that harbored the *SPATA7* and Tetracopeptide domain 8 gene (*TTC8*). Sequencing of all the coding regions of *SPATA7* indicated c.913-23T>G an already known SNP in all the 3 affected members as well as unaffected members of the family. However, mutation screening in all the 15 coding exonic regions including splice donor and acceptor sites of *TTC8* revealed a novel missense substitution. This novel substitution segregated in all the three tested affected members in homozygous form whereas unaffected members of this family were either heterozygous or homozygous for the wild-type allele. Further, 50 tested control samples (100 chromosomes) from the same population did not carry the identified substitution, excluding its possibility as a polymorphism. **Conclusion:** We identified a novel disease-linked missense mutation in *TTC8* in an arRP family of India origin with nonsyndromic RP. *TTC8* is previously reported to be linked with Bardet-Biedl syndrome (BBS) as well as with nonsyndromic RP. Our findings thus further expand the mutation spectrum of *TTC8* and inherent genetic and phenotypic heterogeneity for arRP.

2888M

Achromatopsia Genetic Determinants in Palestinian Families. *H. Shahin[1], F. Sheibat[1], Y. Ashhab[2].* 1) Dept Life Sci, Bethlehem Univ, Bethlehem, Palestinian Territory; 2) Biotechnology Research Center, Palestinian Polytechnic University, Hebron, Palestine.

Achromatopsia is a rare autosomal recessive disorder that leads to color blindness, and can include photophobia, nystagmus, cataracts, reduced visual acuity and eccentric fixation. Achromatopsia can result from mutations in any of four genes: CNGB3, CNGA3, GNAT2, and PDE6C. The purpose of our study was to discover the genetic causes in consanguineous Palestinian families with achromatopsia. In four multiply affected families, where affected individuals had complete color blindness, reduced vision, photophobia and nystagmus, we first tested linkage to each of the four known genes, then fully screened any linked gene by Sanger sequencing. The first family evaluated includes 7 individuals with achromatopsia and 11 unaffected sibs from 3 generations. Parents of the affected individuals are members of the same extended kindred. CNGA3, which codes for the cyclic nucleotide-gated cation channel alpha-3 cone photoreceptor, is most likely responsible for achromatopsia in this family. Markers flanking CNGA3 were homozygous in affected family members. Sequencing of all 8 coding exons of CNGA3 revealed an in-frame 3 bp deletion at chr2:99,012,573-99,012,576 delATC, resulting in CNGA3 c.1357delATC (p.Ile312del). Deletion of isoleucine at residue 312 is predicted by bioinformatics tools to perturb the transmembrane structure of this photoreceptor. Genotyping this mutation in the four families revealed that all 17 affected family members were homozygous for the deletion and all 25 unaffected family members tested were either heterozygous or wildtype at this site. Multiple mutations in CNGA3, including single amino acid deletions, have been identified in persons with achromatopsia in the Middle East and elsewhere.

2889T

Mutations in ALDH1A3, FOXE3 and VSX2 cause ocular abnormalities in consanguineous Pakistani families. *E. Ullah[1,2], MA. Saqib[2], N. Shah[2], S. Sajid[2], R. Lao[3], E. Wan[3], PL. Tang[3], P. Kwok[3], M. Ansar[2], A. Slavotinek[1].* 1) Department of Pediatrics, University of California San Francisco, San Francisco, CA; 2) Department of Biochemistry, Quaid-i-Azam University, Islamabad, Pakistan; 3) Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA.

Anophthalmia/microphthalmia (A/M) is a genetically heterogeneous birth defect for which the etiology remains poorly understood, with less than 50% patients receiving a molecular genetic diagnosis. We hypothesized that exome sequencing in A/M patients would be an efficient method to provide data to find novel candidate genes. We have collected 6 families with A/M but without any other phenotypic abnormalities from Pakistan. All families had multiple affected members and evidence of consanguinity, suggestive of autosomal recessive (AR) inheritance. We therefore performed Sanger sequencing to exclude mutations in known A/M genes with AR inheritance (ALDH1A3, RAX, VSX2, FOXE3) and two genes with autosomal dominant (AD) inheritance (SOX2, OTX2). In the first family with non-syndromic bilateral anophthalmia, we found a novel frameshift mutation c.1310_1311delAT, p.Tyr437Trpfs*44 in ALDH1A3, which segregates in the family with disease phenotype. The Mutation Taster predicted this mutation as disease causing. In a second family with non-syndromic bilateral anophthalmia, we found a nonsense mutation c.598C>T, p.Arg200* in VSX2, that has been published in a family with bilateral anophthalmia in the medical literature. In a third family with bilateral microphthalmia, corneal opacity and anterior segment dysgenesis, we found a novel missense mutation c.289A>G, p.Ile97Val in FOXE3. This mutation is predicted as disease causing mutation by SIFT, PolyPhen-2 and Mutation Taster. In a fourth family with bilateral microphthalmia, corneal opacity and anterior segment dysgenesis we found a frameshift mutation c.21_24delGGAT, p.Met7Ilefs*216 in FOXE3. This mutation has previously been reported in a Pakistani family with bilateral microphthalmia, sclerocornea, corneal opacity, buphthalmosis and primary aphakia. The remaining two families were negative for the aforementioned disease genes and therefore were subjected to whole exome sequencing. We selected several autosomal recessive sequence variants as potential candidates that were predicted to be deleterious after exome data analysis by Ingenuity Variant Analysis. Although none of the candidate sequence variants segregated with the disease phenotype, we are currently analyzing whole exome data of both families to identify disease causing variant.

2890S

Two novel mutations in *ABCG5* and *ABCG8* genes in a Mexican family with sitosterolemia. *A. COLIMA[1, 2], E. WONG[3, 4], M. MAGAÑA[1].* 1) LABORATORIO DE BIOQUIMICA I, DIVISION DE GENETICA, CIBO, IMSS, guadalajara, JA., Mexico; 2) DOCTORADO EN GENETICA, CENTRO UNIVERSITARIO DE CIENCIAS DE LA SALUD, UNIVERSIDAD DE GUADALAJARA; 3) HOSPITAL GENERAL DE TEPIC; 4) UNIVERSIDAD AUTONOMA DE NAYARIT.

Sitosterolemia (STSL[MIM 210250]) is an autosomal recessive inherited disorder characterized by enhanced intestinal absorption of vegetal sterols and cholesterol and a slow removal of them through the liver. Clinical features include xanthomas, arthralgias, arthritis, thrombocytopenia, accelerated atherosclerosis and premature coronary heart disease. Sitosterolemia is caused by mutations in *ABCG5* and *ABCG8* genes. We studied a 12 years old girl from Nayarit, Mexico with biochemical and clinical characteristics of sitosterolemia. She showed slightly high total cholesterol values (295mg/dL), levels extremely high of sitosterol 681 µg/ml (normal range: 2.16±2.20 µg/ml) and campesterol 324 µg/ml (normal range: 3.09±1.65 µg/ml). She also had soft lumps on elbows and knees, thickening of the Achilles tendon and thrombocytopenia. Her treatment consisted of pravastatin (40 mg) and cholestyramine. One of the two patient's brothers had thrombocytopenia and high values of sitosterol (192µg/ml) and campesterol (125µg/ml). DNA was obtained with the CTAB-DTAB method for a molecular analysis. The primer pairs to PCR were design using the Oligo 6 software. Mutational screening of the *ABCG5* and *ABCG8* genes was performed by direct sequencing using an ABI PRISM310 genetic analyzer. . In *ABCG5* gene was detected a heterozygous mutation c.1523 delC, rendering an amino acid change H510T and L511X, resulting in a truncated protein. This mutation was also present in the mother and in the two brothers of the patient. In addition, deletion of exon 2 of the *ABCG8* gene in the index case and in her mildly affected brother was detected, in homozygous state. We found no information in the literature about the two mutations observed so that they could be considered as new mutations related to sitosterolemia.

2891M

Identification of hemizygous loss-of-function mutations in *OFD1* in two unrelated male patients with a clinical phenotype of primary ciliary dyskinesia (PCD). *W.B. Hannah[1,2], A.S. Rali[1], S. Strausbaugh[1,3], B. Gaston[3], M. Rosenfeld[4], W.E. Wolf[5], M.R. Knowles[5], M.A. Zariwala[6].* 1) Department Internal Medicine, University Hospitals Case Medical Center, Cleveland, OH; 2) Department of Genetics, University Hospitals Case Medical Center, Cleveland, OH; 3) Department of Pediatrics, University Hospitals Case Medical Center, Cleveland, OH; 4) Seattle Children's Hospital, School of Medicine, University of Washington, Seattle, WA; 5) Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC; 6) Department of Pathology & Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC.

This report describes two male patients with hemizygous *OFD1* variants and typical symptoms of primary ciliary dyskinesia (PCD). PCD is a rare, genetically heterogeneous, usually recessive disorder characterized by ciliary dysfunction and sino-pulmonary disease, otitis media, bronchiectasis, situs abnormalities, and male infertility. Rarely, PCD can be present in the setting of other syndromes such as X-linked mental retardation (XLMR). Usually, mutations in *OFD1* are observed in patients presenting with X-linked oral-facial-digital type 1 syndrome; however, in one family an *OFD1* mutation has been reported in male subjects presenting with recurrent respiratory tract infections, macrocephaly, intellectual disability, and ciliary dysfunction (PMID# 16783569). We outline the clinical presentation and genetic profile of two male individuals with *OFD1* variants and concurrent PCD symptoms.

Case 1 was a 32 year old male with chronic cough, sinus disease, bronchitis/pneumonia, and bronchiectasis. His nasal nitric oxide was low (54.7nl/min versus ~300nl/min for unaffected control subjects), consistent with PCD. Nasal ciliary biopsy showed normal ultrastructure. Whole exome sequencing followed by the analysis of 30 genes associated with PCD (including *OFD1* and *RPGR*, mutations in which have been seen in PCD associated with other syndromes) revealed a hemizygous frameshift mutation (c.2868delT [p.Pro957Leufs*2]) in *OFD1* predicted to result in premature translation termination signal. Case 2 was a 30 year old male who presented with a history of recurrent sinusitis and pneumonias, bronchiectasis, intellectual delay, and polydactyly. His nasal ciliary biopsy showed normal ultrastructure. Whole exome sequencing revealed a hemizygous *OFD1* deletion (c.2789_2793del5 [p.Ile930Lysfs*8]) resulting in a frameshift followed by a premature stop codon.

In summary, mutations in *OFD1* were identified in two unrelated male subjects who presented with a phenotype that resembles PCD and in one case also presented with intellectual delay and polydactyly.

This abstract was funded by 5U54HL096458-06 (NIH/ORDR/NHLBI), 5R01HL071798 (NIH/NHLBI), UL1 TR000083 (NIH/NCATS), and 1X01HL115246-01 (NIH/NHLBI).

2892T

A Novel Homozygous LRP5 Splice-site Deletion Mutation Causes Syndromic Autosomal Recessive Familial Exudative Vitreoretinopathy. *V. Chini[1], Y. Al-Sarraj[1], MT. Trese[2], H. El-Shanti[1,3], M. Kambouris[1,4].* 1) Qatar Biomedical Research Institute, Medical Genetics Center, Doha, Qatar; 2) Associated Retinal Consultants, Royal Oak, MI, USA; 3) University of Iowa, Pediatrics, Iowa City, IA, USA; 4) Yale University School of Medicine, Genetics, New Haven, CT, USA.

A consanguineous Saudi Arabian family with two female siblings affected by an autosomal recessive condition resembling Familial Exudative Vitreoretinopathy [FEVR], but also with short stature, bone fragility with thin and wasted appearance was studied by homozygosity mapping and positional candidate gene screening to identify the offending gene and mutation. The gene was mapped to three possible homozygous genomic regions [[2q, 4q, 11q], as the family structure did not allow identification of a single interval with a significant LOD score. Mutations in three genes (FZD4, TSPAN12, NDP and LRP5) have been associated with FEVR. The LRP5 gene localizes within the 11q13.2 homozygosity interval in this family rendering it the positional candidate of choice. Screening by Sanger sequencing identified a novel homozygous one-base splice-site deletion mutation c.3236+1 delG in exon 14. LRP5 is a low-density lipoprotein receptor (LDLR) a transmembrane protein that binds and internalizes ligands in the process of receptor-mediated endocytosis. The cDNA encodes a 1,615-amino acid protein containing conserved modules including a putative signal peptide, four epidermal growth factor (EGF) repeats with associated spacer domains, three LDLR repeats, a single transmembrane-spanning domain, and a cytoplasmic domain. The extracellular domain contains 6 potential N-linked glycosylation sites. LRP5 has a unique organization of EGF and LDLR repeats compared to other LDLR family members and in addition to FEVR, mutations in the gene have been associated with Hyperostosis corticalis OMIM 144750; Osteopetrosis, autosomal dominant 1 OMIM 607634; Osteoporosis-pseudoglioma syndrome OMIM 259770; Osteosclerosis OMIM144750; van Buchem disease, type 2 OMIM 607636; Bone mineral density variability 1 OMIM 601884; Osteoporosis OMIM 166710. Only missense mutations and splice site substitutions in LRP5 have been associated with autosomal dominant and recessive FEVR. This is the first report of an autosomal recessive LRP5 splice-site deletion mutation causing a syndromic form of FEVR.

2893S

A Mutation in SORBS2 Actin filament Adapter, Cell Adhesion, Migration & Intracellular Signaling Protein Causes Autosomal Recessive Hand & Foot Malformation Syndrome. *H. El-Shanti[1,2], Y. Al-Sarraj[1], H. Shaath[1], F. Alshaban[1], RI. Thompson[1], V. Chini[1], M. Kambouris[1,3].* 1) Qatar Biomedical Research Institute, Medical Genetics Center, Doha, Qatar; 2) University of Iowa, Pediatrics, Iowa City, IA, USA; 3) Yale University School of Medicine, Genetics, New Haven CT, USA.

A consanguineous family of Qatari ethnicity with two male siblings, 24 and 6 year old, affected by possibly an autosomal recessive disorder of hand and foot malformations and dysmorphic facial features, was studied by homozygosity mapping and whole Exome Next Generation Sequencing [NGS] of the two affected siblings and one parent to identify the responsible gene and mutation. The right foot of the younger patient was split with preaxial polydactyly and the left foot showed syndactyly between the 3rd and 4th toes and the 5th and 6th toes with postaxial polydactyly. Both feet showed nail hypoplasia and malpositioned halluces. The hands showed scars of the removed postaxial polydactyly. In addition, there is ptosis, hypertelorism, thin and tented upper lip with a smooth filtrum. The older brother had similar facial features and similar hand and foot malformations. The offending gene was mapped to three possible genomic intervals [4q, 19p, 21q]. Comparative analyses of the NGS data for autosomal recessive as well as X-linked inheritance and data mining for damaging variants within the homozygosity intervals and the shared X-chromosome genomic regions identified a homozygous single nucleotide frame shift deletion mutation c.143delC / p.Pro48fs in the SORBS2 gene, at 4q35.1. The mutation co-segregates with the disease phenotype within the family and it is absent in 400 ethnically matched control chromosomes. SORBS2 [sorbin and SH3 domain containing 2] an 824 amino acid protein is involved in actin filament organization, cell adhesion, cell migration and intracellular signaling. It has three C-terminal SH3 domains and an N-terminal sorbin homology (SoHo) domain that interacts with lipid raft proteins. It functions as an adapter protein to assemble signaling complexes in stress fibers, acting as link between Abl family kinases and the actin cytoskeleton and it binds the synapse-associated protein 90 / postsynaptic density-95-associated protein (SAPAP). The frame shift deletion mutation is in exon 1 at amino acid 48 and results in incorporation of 18 erroneous amino acids before a stop codon is encountered. It affects one of the nine possible isoforms [isoform 6] as it is the only isoform containing exon 1. No data on SORBS2 isoform 6 expression or role during embryonic development exists but it appears the isoform is essential for normal fetal development. This is the first association of SORBS2 to a human malformation disease.

2894M

A Mutation in MYO1A Causes Autosomal Recessive Autism Spectrum Disease. *M. Kambouris[1,2], V. Ilyin[3], Y. Al-Sarraj[1], H. Shaath[1], F. Alshaban[1], M. Tolefat[4], V. Chini[1], H. El-Shanti[1,5].* 1) Qatar Biomedical Research Institute, Medical Genetics Center, Doha, Qatar; 2) Yale University School of Medicine, Genetics, New Haven, CT, USA; 3) Carnegie Mellon University-Qatar, Doha, Qatar; 4) Shafallah Center for Children with Special Needs; 5) University of Iowa, Pediatrics, Iowa City, IA, USA.

A consanguineous family of Pakistani ethnicity with two female siblings (22 and 19 years of age), affected by a possibly novel autosomal recessive disorder, was studied by homozygosity mapping and whole Exome Next Generation Sequencing [NGS] of the two affected siblings and one parent to identify the responsible gene and mutation. The disorder is marked by intellectual disability, speech and motor delay, congenital malformations and possibly autism spectrum disorder (ASD). The malformations include microcephaly, microphthalmia, micrognathia and arachnodactyly with hyperextensibility and persistent fetal pads in fingers and toes. The offending gene was mapped to five possible homozygous genomic regions [[6q, 12q, 17p, 20p, 22q], as the family structure did not allow identification of a single interval with a significant LOD score. Comparative analyses of the NGS data for autosomal recessive inheritance and data mining for damaging variants within the homozygosity intervals identified a damaging homozygous c.C1675T / p.R559C mutation in the MYO1A gene, at 12q13.3. The mutation co-segregates with the disease phenotype within the family, is absent in known polymorphism databases and in 400 ethnically matched control chromosomes. Myosins are molecular motors that, upon interaction with actin filaments, utilize energy from ATP hydrolysis to generate mechanical force. The N-terminal motor domain contains both ATP-binding and actin-binding sequences. Following the motor domain is a light-chain-binding 'neck' region containing 1-6 copies of a repeat element, the IQ motif that serves as a binding site for calmodulin and other members of the EF-hand superfamily of calcium-binding proteins. The C terminus has a distinct tail domain that serves in dimerization, membrane binding, protein binding, and/or enzymatic activities and targets each myosin to its particular subcellular location. Heterozygous mutations in MYO1A have been found in patients with sensorineural hearing loss, speculated to cause autosomal dominant sensorineural hearing loss but co-segregation to the phenotype has never been demonstrated. Two rare heterozygous MYO1A mutations [c.G2021A / p.G674D and one in the 3'UTR] have been found in patients with autism but their clinical significance is unknown. The association of MYO1A to autosomal recessive ASD without deafness in this family, elevates the importance of MYO1A both as causative and contributive gene for Autism Spectrum Disease.

2895T

Linkage Analysis and Gene Identification in Consanguineous Pakistani Families with Autosomal Recessive Retinal Dystrophy. *M. Ansar[1,2], K. Lee[1], RL. Santos-Cortez[1], E. Ullah[2], Z. Ravesh[2], MA. Saqib[2], X. Wang[1], JD. Smith[3], J. Shendure[3], MJ. Bamshad[3], DA. Nickerson[3], W. Ahmad[2], SM. Leal[1], University of Washington Center for Mendelian Genomics.* 1) Center for Statistical Genetics, Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan; 3) Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA.

Retinal dystrophy constitutes a phenotypically and genetically heterogeneous group of inherited diseases. Retinal dystrophies can be grouped into different categories depending on age of onset, disease progression, additional clinical features and mode of inheritance. The clinical signs and symptoms of retinal dystrophies are complex and overlapping, which challenges clinicians to make accurate diagnoses. The Pakistani population provides an opportunity to document large families with multiple affected individuals and has been used to extensively study many inherited disorders in the last two decades. In this study, we performed homozygosity mapping, linkage analysis and DNA sequencing using samples from seven consanguineous Pakistani families with autosomal recessive retinal dystrophy. Seven multiplex autosomal recessive retinal dystrophy families were ascertained from different regions of Pakistan. DNA samples from these families were genotyped using 500,000 SNP markers which were analyzed using homozygosity mapping and linkage analysis. Two families are linked to regions containing known autosomal recessive retinal dystrophy genes; FAM161A and RDH12; one family mapped to a region with the autosomal dominant retinal dystrophy gene, PRCD; two families map to regions with syndromic genes e.g. CDH3 which include an eye phenotype and two families mapped to novel regions. For the FAM161A gene a novel variant c.782delA (p.Asp261ValfsX39) was found. No potentially pathogenic variants were identified in either PRCD or CDH3. One of the families for which the autosomal recessive retinal dystrophy locus is linked to a novel region, a DNA sample from a single affected family member underwent exome sequencing and a novel gene which segregates with the RD phenotype was identified on chromosome 1q. The variant was not found in 360 Pakistani control chromosomes. The identified gene was also screened in samples from 24 Pakistani probands with autosomal recessive retinal dystrophy, but no putatively causal variants were found. Functional studies and a search for additional families with a variant in the same gene are ongoing. The five families for which the causal variant has not been identified will undergo exome sequencing.

2896S

Genetic linkage analysis of familial PFAPA in Finland. *E. Einarsdottir[1], U. Lantto[2], J. Kere[1,3,4,5], M. Uhari[2], T. Tapiainen[2], P. Koivunen[2], M. Renko[2].* 1) Department of Biosciences and Nutrition, Karolinska Institute, SE-14183 Huddinge, Sweden; 2) Departments of Pediatrics and Otorhinolaryngology, Institute of Clinical Medicine, University of Oulu, FI-90014 Oulu, Finland; 3) Folkhälsan Institute of Genetics, Helsinki, Finland; 4) Haartman Institute, Medical Genetics, University of Helsinki, FI-00014 Helsinki, Finland; 5) Research Programs Unit, Molecular Neurology, University of Helsinki, FI-00014 Helsinki, Finland.

We report on six multi-case families from Finland with Periodic Fever, Aphthous Stomatitis, Pharyngitis, Adenitis Syndrome (PFAPA). PFAPA is characterized by recurrent spontaneous fevers lasting of up to a week, followed by 3-6 weeks of no apparent symptoms.

PFAPA has previously been considered mainly a sporadic disease, but we have identified six families from the Oulu region of Finland with 2-4 confirmed cases and available DNA samples. We performed a genome-wide linkage analysis of PFAPA in these families, assuming a highly penetrant recessive model and a very rare mutation. We followed up interesting linkage peaks through haplotype analysis and studying potential candidate genes in the regions of interest. We also looked for indications of shared founder risk haplotypes between the different families.

We found a linkage peak with a parametric lodscore of 2,811 (using all families), contributed to mainly by three of the families. This region contains a plausible candidate gene for PFAPA, involved in innate immunity, and we are currently searching this gene for possible causative mutations. We find no indications that previously described possible PFAPA genes might contribute to the disease in the Finnish population.

2897M

Haploinsufficiency of a novel gene on 3p26.1, SMDD1, cause autosomal-dominant dentin dysplasia type I. *F. Xiong[1], Y. Liu[2], Z. Ji[1], Z. Yu[1], L. Hu[1], Z. Liu[3], Z. Tian[4], Q. Qiu[1], D. Chen[5], L. Zhang[4], X. Shang[1], W. Zhang[6], H. Yuan[7], X. Xu[1].* 1) Department of Medical Genetics, Southern Medical University, Guangzhou, China; 2) Department of Prenatal Diagnosis Center, Dongguan Maternal and Child Health Hospital, Dongguan, Guangdong, China; 3) Department of Stomatology, Dongguan Donghua Hospital, Dongguan, Guangdong, China; 4) Department of Stomatology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, China; 5) Department of Stomatology, Fourth affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China; 6) Department of Cell Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong, China; 7) Institute of Otolaryngology, Chinese PLA General Hospital, Beijing, China.

The hereditary dentin defects, dentinogenesis imperfecta (DGI) and dentin dysplasia (DD) comprise a group of autosomal-dominant genetic conditions and the molecular basis of such dental disorders in all sub-groups except for DD Type I (DDI) is linked to mutations in dentin sialophospho-protein gene (DSPP). DDI shows an exclusive character in phenotypes featuring late-onset missing teeth with short dental roots and unknown genetic etiology, thus providing a useful model to unravel the mechanism for dentin formation involved in tooth development. Here we study the gene mapping and the molecular pathogenic mechanism of DDI. Using a cohort of a large Chinese family with 20 normal members and 14 DDI patients, we mapped the gene locus responsible for DDI to 3p26.2-3p24.3 by combining use of whole genome-wide SNP array and STR linkage analysis. We further identified a novel missense mutation, c.353 C>A (P118Q) in SMDD1 gene on 3p26.1 through targeted sequencing of 125 candidate genes by capture-based next-generation sequencing, and followed by Sanger sequencing and co-segregation analysis for confirmation. We showed that the mutant P118Q in dental pulb stem cells expressed at 50% of wild-type levels with exhibiting haploinsufficiency. In vivo zebrafish functional assay of a homology of human SMDD1, we determined that missing teeth similar to that of the index case phenotypes were replicated in the RNAi knock-down zebrafish and could be partially rescued by injection of normal human SMDD1 mRNA to the mutant zebrafish. We also observed that SMDD1 depletion in the zebrafish negatively regulates the expression of two major genes (bmp2 and pitx2) involved in odontogenesis. In addition, we generated the P118Q mutant knock-in transgenic (TG) mice and investigated the role of the P118Q mutation in heterozygous and homozygous TG mice. We found that dentin defects were radiographically evident in all teeth with homozygotes, to show incomplete obliteration of the nonmineralized pulp in TG mice, consistent with clinical characteristics in those patients with DDI. Moreover, high-resolution radiography, micro-computed tomography and scanning electron microscopy revealed a reduced zone of mineralized dentin with anomalies in the number and organization of dentinal tubules in TG mice. Our observations demonstrate that haploinsufficiency of SMDD1 disrupts dental formation and that this novel gene, together with other odontogenesis genes are involved in tooth development.

2898T

Mutations in *CCNO* identified in patients with a clinical phenotype consistent with primary ciliary dyskinesia (PCD) and defective mucociliary clearance reflecting reduced motile cilia generation. *M.A. Zariwala[1], M.W. Leigh[2], J.E. Pittman[2, 5], J.L. Carson[2], M.J. Hazucha[3], J. Wallmeier[4], N.T. Loges[4], H. Olbrich[4], H. Omran[4], M.R. Knowles[3].* 1) Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC, USA; 2) Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC, USA; 3) Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC, USA; 4) Department of General Pediatrics and Adolescent Medicine, University Hospital Muenster, Muenster, Germany; 5) Pediatric Pulmonology, Washington University School of Medicine, St. Louis, MO, USA.

Primary ciliary dyskinesia (PCD) is clinically manifested by oto-sino-pulmonary symptoms, situs abnormalities and male infertility. It is a rare, recessive, genetically heterogeneous disorder caused by defective ciliary structure, function and/or biogenesis. Diagnosis has traditionally relied on documenting defective ciliary motility or ciliary ultrastructure, either using electron microscopy, or immunofluorescence analysis. Not all subjects with PCD present with defective cilia; thus, defining biallelic mutations in one of the 30 known or novel cilia-associated genes has become a useful diagnostic approach in challenging cases. A subset of patients with a clinical phenotype consistent with PCD have reduced generation of multiple motile cilia (RGMC). Indeed, mutations in *CCNO* that affect centriole amplification and migration have been recently reported in such cohort (PMID# 24747639). We have accrued four unrelated families with a clinical phenotype consistent with PCD and low levels of nasal nitric oxide (an adjunct marker for PCD) and ciliary "aplasia and/or oligoplasia" on multiple nasal/bronchial biopsies. Mutation screening of all three coding exons of *CCNO* revealed biallelic mutations in all four unrelated affected probands. Segregation analysis revealed that mutations were inherited in trans. Briefly, an affected male (UNC-79) harbored a novel homozygous frameshift mutation (c.875_897del23 [p.Asp292Alafs*71]). In another family (UNC-178), an affected female presented with a novel, evolutionarily conserved, missense variant (c.851C>T [p.Ala284Val]) and a previously known frameshift mutation (c.248_252dupTGCCC [p.Gly85Cysfs*11]). An affected sib-pair (UNC-136) harbored a previously known homozygous frameshift mutation (c.248_252dupTGCCC [p.Gly85Cysfs*11]), and an affected identical twin-pair (UNC-468) harbored previously known compound heterozygous (c.248_252dupTGCCC [p.Gly85Cysfs*11]) and (c.258_262dupGGCCC [p.Gln88Argfs*8]) mutations. To conclude, mutations in *CCNO* appear to be a common cause of a clinical phenotype indistinguishable from PCD, reflecting RGMC. Large scale screening will be carried out in subjects suspected to suffer from PCD without a known genetic diagnosis to decipher the prevalence of *CCNO* in a large cohort. This abstract was funded by 5U54HL096458-06 (NIH/ORDR/NHLBI), 5R01HL071798 (NIH/NHLBI), UL1 TR000083 (NIH/NCATS), BESTCILIA (EU), and SYSCILIA (EU).

2899S

Identification of New Genes and Pathways for Rare Infantile Forms of Spinal Muscular Atrophy and Neuromuscular Disorders. *J.M. Hunter[1,2], C.D. Balak[1,2], J. Kiefer[2], M.E. Ahearn[1], G. Lambert[3], D. Duggan[3], B. Wirth[4], W. Tembe[5], C. Legendre[5], W. Liang[6], L. Cuyugan[6], J. McDonald[6], J. Adkins[6], A. Kudoglu[7], J. Corneveaux[7], M. Russell[7], M. Huentelman[7], D. Craig[7], J. Carpten[1], S.M. Bernes[8], J. Hall[9], L. Baumbach-Reardon[1,2].* 1) Integrated Cancer Genomics, TGen, Phoenix, AZ; 2) Pharmaceutical Genomics, TGen, Phoenix, AZ; 3) Genetic Basis of Human Disease, TGen, Phoenix, AZ; 4) Institute of Human Genetics, University Hospital of Cologne, Cologne, Germany; 5) Center for Bioinfomatics, TGen, Phoenix, AZ; 6) Collaborative Sequencing Center, TGen, Phoenix, AZ; 7) Neurogenomics, TGen, Phoenix, AZ; 8) Division of Neurology, Phoenix Children's Hospital, Phoenix, Az; 9) Departments of Pediatrics and Medical Genetics, University of British Columbia, Vancouver, Canada.

Spinal muscular atrophy (SMA) (MIM# 600354) is the leading genetic cause of infantile death. Mutations in the SMN1 gene account for a large percentage of SMA. However, numerous cases with SMA or related disorders do not have mutations in SMN1. Our goal is to use next-generation sequencing to identify the genetic cause of undiagnosed cases of motor neuron disease and muscular dystrophies in infants and children. Our journey began many years ago with the study of a rare form of X-linked (XL) form of SMA. In 2008 Dr. Lisa Baumbach-Reardon's group discovered mutations in UBA1 as the cause of XL-SMA (MIM# 3018300). Since that discovery, we have developed a custom sequencing panel for all coding regions of UBA1. Using Ion-torrent sequencing we have screened 10 affected cases and 18 relatives in our cohort. As expected, UBA1 sequencing revealed common variants in each individual, but surprisingly no new disease associated variants were identified. The molecular etiology of disease in these cases has remained undiagnosed for many years. At TGen, we have now been able to performed whole exome sequencing on these individuals as well as numerous other cases. We are excited to report that in many of these cases we have identified novel probable pathogenic mutations in genes known to cause disease as well as in genes not previously associated with disease. We highlight results from four cases. First, a family with two affected boys revealed novel compound heterozygous deleterious CHRND mutations that cause a form of lethal congenital myasthenic syndrome (MIM# 100720). Second, a novel start loss M1V mutation in SCML2 (MIM# 300208) was detected in the proband and mother of a family with an X-linked history of fetal and neonatal deaths. SCML2 encodes a component of the polycomb transcriptional repressor complex, and has not been previously associated with any human disease. In a second large family with X-linked infantile deaths, we identified another novel missense mutation in SCML2 in obligate carriers. Lastly, we identified a novel de novo mutation in DYNC1H1 known to cause autosomal dominant SMA lower extremity predominance (SMA-LED MIM#158600). These results, along with identification of unique and interesting variants in other families, shed light on new mechanisms and pathways associated with SMA and related diseases. We are privileged to be part of the effort to translate next-generation sequencing into answers for these rare and devastating diseases.

2900M

GeneSEARCH - Diagnostic testing and a tool for research in the Iranian Population. *E.G. Ozkan[1], B. Chioza[2], J. Aragon-Martin[1], G. Harlalka[2], M. Patton[1], K. Everett[1], R. Sharifi[1].* 1) St. George's University of London, London, UK; 2) University of Exeter Medical School, Devon, UK.

The GeneSEARCH programme is a collaboration between Iranian research centres and St George's University of London. The aim is to identify genetic causes of inherited diseases in the Iranian communities. Here we report the identification of one previously reported and three novel mutations in genes causing retinopathies in consanguineous Iranian families. Homozygosity mapping was performed using data generated from the Cyto12 SNP array pertaining to affected individuals from each family. Candidate genes were screened using massively parallel sequencing and/or Sanger sequencing. Validation of putative causal variants and segregation analysis was performed using Sanger sequencing. All described variants segregated with disease. In silico tools demonstrated that all segregating variants were likely to be damaging. All described variants were absent from 100 ethnically matched controls. Family 1: a 3 generation family with 4 individuals presenting with retinitis pigmentosa (RP) and polydactyly. The largest region of homozygosity (ROH) was on chromosome 2q23.3-2q31.1 and contains the Bardet-Biedl syndrome (BBS [MIM 209900) gene, BBS5 (MIM 603650). RP and polydactyly are key diagnostic features of BBS. Screening of BBS5 identified a novel substitution in exon 5; c.382C>G (p.H128D). Family 2: a 6 generation family in which 3 males have been diagnosed with Leber Congenital Amaurosis (LCA [MIM 204000).The largest ROH was on chromosome 17p13. This region contains a known gene for LCA: GUCY2D (MIM 600179). Screening of this gene identified a novel variant in exon 12; c.2348T>C (p.L783P). Family 3: a 4 generation family with 7 individuals affected by RP. The largest ROH was on chromosome 1q31.3. This region contains a known gene for LCA: CRB1 (MIM 604210). Screening of this gene identified a novel frameshift mutation in exon 2 (c.361delC). Family 4: a 6 generation family with 18 individuals affected by RP. The largest ROH was on chromosome 15q23-q25.1. The segregating variant rs28937873 (c.932G>A, p.R311Q) in NR2E3 (MIM 604485) has been previously reported in a Jewish family with RP. GeneSearch is a unique opportunity to diagnose rare genetic disorders within the Iranian community. The initial aim of this work is diagnostic and will benefit families and the community for genetic and carrier testing. This study provides the first example of this work. Genetic diagnosis is a crucial component of treatment and counselling for heterogeneous disorders such as these.

2901T

***HIVEP2*: A New Causative Gene for Intellectual Disability?** *A.M. Zink[1,2], S. Srivastava[3,4], K. Cremer[1], M. Menzel[6], T. Wieland[7], M. Schuhbach[6], S. Biskup[6,8], M. Kreiß-Nachtsheim[1], S. Endele[9], D. Wieczorek[10], T.M. Strom[7], M. Zenker[5], J. Cohen[3], H. Engels[1], I. Schanze[5], S. Naidu[3,4].* 1) Institue of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics, Life and Brain Center, Rheinische Friedrich-Wilhelms-University, Bonn, Germany; 3) Department of Neurogenetics, Kennedy Krieger Institute, Baltimore, Maryland, USA; 4) Departments of Neurology and Pediatrics, The Johns Hopkins Hospital, Baltimore, Maryland, USA; 5) Institute of Human Genetics, University Hospital Magdeburg, Magdeburg, Germany; 6) CeGaT GmbH, Tuebingen, Germany; 7) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 8) Hertie Institute for Clinical Brain Research, German Center for Neurodegenerative Diseases, Tuebingen, Germany; 9) Institute of Human Genetics, University of Erlangen-Nuremberg, Erlangen, Germany; 10) Institute of Human Genetics, Universitätsklinikum Essen, Essen, Germany.

Intellectual disability (ID) has an estimated prevalence of 2-3%. Due to its extreme heterogeneity, the genetic basis of ID remains elusive in many cases. Recently, whole exome sequencing (WES) studies revealed that a large proportion of sporadic cases are caused by *de novo* gene mutations. Applying WES to ID patients at different centers, we identified three patients carrying *de novo* mutations in *HIVEP2* (human immunodeficiency virus type I enhancer binding protein 2). Two of the mutations are nonsense mutations and one is a 1 bp deletion resulting in a premature stop codon. In silico prediction programs (SIFT/ Mutation Taster) predict nonsense mediated mRNA decay as the consequence of all three mutations, pointing to a loss of function and haploinsufficiency as the common disease-causing mechanism of the three *HIVEP2* mutations. None of the three mutations was present in the Exome Variant Server or in 2500 in-house exomes. All three patients presented with moderate ID, muscular hypotonia, mild dysmorphic signs and variable mild structural brain anomalies. Growth parameters were in the normal range except for slight microcephaly at birth in one patient. Two of the patients presented with behavioral anomalies including hyperactivity and aggression. *HIVEP2*, also known as *Schnurri-2*, belongs to a family of zinc finger-containing transcriptional proteins involved in growth and development. Many of the genes regulated by *HIVEP2* are implicated in brain development, e.g. *SSTR-2*, *c-Myc* and several genes of the NF-κB pathway. *HIVEP2*- knockout mice have been shown to exhibit several behavioral features suggestive of schizophrenia such as working memory deficits, as well as increased anxiety and hyperactivity. Based on the genotype-phenotype correlation and the previously published functional data we propose *HIVEP2* as a novel causative ID gene.

2902S

DUX4 induces FRG2 expression by directly activating its promoter in facioscapulohumeral muscular dystrophy. *P.E. Thijssen[1], J. Balog[1], Z. Yao[2], T.P. Pham[1], R. Tawil[3], S.J. Tapscott[2], S.M. Van der Maarel[1].* 1) Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America; 3) Neuromuscular Disease Unit, Department of Neurology, University of Rochester Medical Center, Rochester, New York, United States of America.

Facioscapulohumeral muscular dystrophy (FSHD) is the third most common muscular dystrophy and the major form is associated with a genetic contraction of the polymorphic D4Z4 macrosatellite repeat on chromosome 4q35. Several candidate genes have been proposed as pathogenic entities in FSHD, including *FRG2* located proximal to the D4Z4 macrosatellite repeat. For long it has been known that expression of *FRG2* is induced specifically in differentiating myoblast cultures derived from FSHD patients, which was explained by a position effect emanating from the contracted D4Z4 repeat array. However, a contribution of FRG2 upregulation to FSHD pathology was never shown and the function of FRG2 remains unclear at this time. More recently, sporadic activation of the *DUX4* retrogene, encoded within the D4Z4 unit, has emerged as the most prominent disease model. DUX4 is a double homeobox transcription factor and its expression has been shown to be toxic in myogenic cultures. We now show that FRG2 expression in FSHD derived differentiated myoblasts is a direct consequence of DUX4 protein activity. Overexpression of DUX4 in proliferating myoblasts and primary fibroblasts results in a steep upregulation of FRG2 transcription. Moreover, we identified DUX4 binding at the *FRG2* promoter by chromatin immunoprecipitation followed by deep sequencing. Next, we confirmed the transactivation effect of DUX4 on the *FRG2* promoter by luciferase reporter assays. Activation of luciferase was dependent on DUX4 expression and the presence of the DUX4 consensus binding sites identified in the *FRG2* promoter. In conclusion we show that the FSHD specific upregulation of *FRG2* can be explained by the protein activity of DUX4, confirming its central role in FSHD pathogenesis.

2903M

Expansion of the spectrum of nuclear envelopathies: mutation in TOR1AIP1 associated with muscular dystrophy. *P. Dincer[1], G. Kayman-Kurekci[1], B. Talim[1], P. Korkusuz[1], N. Sayar[2], T. Sarioglu[3], I. Oncel[1], P. Sharafi[1], H. Gundesli[1], B. Balci-Hayta[1], N. Purali[1], P. Serdaroglu-Oflazer[3], H. Topaloglu[1].* 1) Hacettepe University, Ankara, Turkey; 2) Bilkent University, Ankara, Turkey; 3) Istanbul University, Ankara, Turkey.

A consanguineous family with three individuals affected by a myopathic phenotype with joint contractures, proximal and distal weakness and atrophy with cardiomyopathy and respiratory involvement was analyzed by genome-wide homozygosity mapping using 250K NspI array. A single homozygous haplotype shared by the three affected individuals was detected. Homozygous c.186delG mutation in torsin A-interacting protein 1 (TOR1AIP1) gene encoding lamina-associated polypeptide 1B (LAP1B) was shown to cause a frameshift resulting in a premature stop codon (p.E62fsTer25). TOR1AIP1 mRNA level in the patient skeletal muscle was 5,88-fold lower than in the control sample. Expression of LAP1B protein was absent in the patient skeletal muscle fibres. Ultrastructural examination showed an intact sarcomeric organization but alterations of the nuclear envelope including nuclear fragmentation and degeneration, and altered chromatin condensation. LAP1B is an integral protein of the inner nuclear membrane that binds to A-type and B-type lamins, and is involved in the regulation of torsin A ATPase. Moreover, overexpression of luminal domain-like LAP1 (LULL1)-the endoplasmic reticulum-localized regulatory partner of torsin A-was overexpressed in the patient's muscle. This suggests a compensatory effect between LAP1 and LULL1. The absence of LAP1B in muscle might influence the structural and mechanical stability of the nuclear envelope due to the impaired binding to the nuclear lamina. The muscle-restricted phenotype underlies a critical role for LAP1B in striated muscle and this study expands the spectrum of nuclear envelopathy causing genes.

2904T

Association of IFRD1 gene polymorphisms with nasal polyposis in Cystic Fibrosis. *A. Baldan[1], F. Belpinati[1], A.R. Lo Presti[1], L. Xumerle[1], M.D. Bettin[1], C. Castellani[2], P.F. Pignatti[1], G. Malerba[1], C. Bombieri[1].* 1) Sec. of Biology and Genetics, Dpt of Life and Reproduction Sciences, University of Verona, Verona, Italy; 2) Veneto Regional CF Center, Verona Hospital, Verona, Italy.

Cystic fibrosis (CF), the most frequent autosomal recessive disease in the Caucasian population, is a lethal multi-system disorder caused by mutations in the CFTR gene. Severity of clinical manifestation, particularly the respiratory phenotype, is widely variable, even among patients with the same CFTR genotype or within the same family. This variability is only partially explained by allelic heterogeneity at the CFTR gene, therefore many genes have been investigated as possible CF phenotype modifier contributing to disease severity heterogeneity. Nasal polyposis (NP) is a manifestation of a chronic inflammatory disease of the upper nasal airways that affects 1-4% of the general population worldwide. In CF the prevalence of NP is reported between 6 to 48% of cases with no association with a particular CFTR gene mutation. Literature data reported the IFRD1 (Interferon-Related Developmental Regulator 1) gene as a possible modifier of cystic fibrosis lung disease severity (Nature 458:1039-42, 2009). The IFRD1 gene encodes for a histone-deacetylase-dependent transcriptional co-regulator, expressed during terminal neutrophil differentiation, that could act as CF phenotype modifier through the regulation of neutrophil effector function. In this study, we investigate a possible association with NP of three IFRD1-SNPs (rs7817, rs3807213, rs6968084), previously described in association with a more severe CF lung phenotype, in a cohort of 146 North-East Italian CF patients, diagnosed according to international criteria and clinically evaluated for respiratory and gastrointestinal parameters. An association study was performed by logistic regression analysis; a p value of less than 0.05 was considered to indicate statistical significance. All analysed IFRD1-SNPs were in Hardy-Weinberg equilibrium. No evidence of association between CFTR genotype and respiratory phenotype was observed. A logistic regression model showed a 4-fold higher probability of NP in patients with CT genotype (p=0.02), and a 7.3-fold higher in patients with TT genotype (p= 0.003) when compared to the homozygote wild-type CC patients for the SNP IFRD1-rs7817. In conclusion, these results show the association of IFRD1-rs7817 polymorphism with NP in CF. Further investigation may be accomplished to test IFRD1 polymorphism association in other population samples.

2905S

A missense mutation in hexokinase 1 (HK1) causes autosomal dominant retinitis pigmentosa (adRP). *S.P. Daiger[1,7], S.J. Bowne[1], L.S. Sullivan[1], D.C. Koboldt[2], D.G. Birch[3], D.K. Wheaton[3], S.H. Blanton[4], R.K. Koenekoop[5], C.F. Chakarova[6], R.S. Fulton[2], R.K. Wilson[2], G.M. Weinstock[2], C.A. Garcia[7], C.E. Avery[1], E.D. Cadena[1], R.A. Lewis[8].* 1) Human Genetics Center, Univ. of Texas Health Science Center, Houston, TX; 2) The Genome Institute, Washington Univ., St. Louis, MO; 3) The Retina Foundation of the Southwest, Dallas, TX; 4) Hussman Institute of Human Genomics, Univ. of Miami, Miami, FL; 5) McGill Ocular Genetics Laboratory, Depts of Paediatric Surgery, Human Genetics and Ophthalmology, McGill Univ. Health Center, Montreal, Quebec, Canada; 6) Institute of Ophthalmology, University College London, London, United Kingdom; 7) Dept. of Ophthalmology and Visual Sciences, Univ. of Texas Health Science Center, Houston,TX; 8) Depts of Ophthalmology and Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Whole-genome linkage mapping in a seven-generation family with autosomal dominant retinitis pigmentosa (adRP) identified a linkage region on chromosome 10q21.3-q22.1 with a maximum LOD score of 3.6 at 0% recombination. All known adRP genes and X-linked RP genes were excluded in the family by a combination of linkage exclusion, Sanger sequencing, and next-generation sequencing (NGS). Whole-exome NGS revealed a missense mutation in hexokinase 1, HK1 c.2539G>A, p.Glu847Lys, tracking with disease in all affected individuals. One severely-affected family member who was homozygous for this region by linkage analysis has two copies of the mutation. No other potential mutations were detected in the linkage region nor were any likely candidates identified elsewhere in the genome. Subsequent testing revealed the identical mutation in four additional, unrelated families with adRP, for a total of five mutations in 459 probands tested (1.1%). Taken together, the families have a combined maximum LOD score of over 8.0. No further instance of the mutation was found in publically-available databases or in control samples tested in our laboratory. Of the five families, three are from the Acadian population in Louisiana, one is French Canadian and one is Sicilian. Analysis of the affected chromosome in each family and in the homozygous individual revealed multiple haplotypes with a maximum possible overlap of 600 kb, suggesting either independent origins of the mutation or an ancient founder mutation. HK1 is a widely-expressed gene, with multiple, abundant retinal transcripts, coding for hexokinase 1. Hexokinase catalyzes phosphorylation of glucose to glucose-6-phospate, the first step in glycolysis. The Glu847Lys mutation is in a highly-conserved site, outside of the active site or known functional sites. Bioinformatic analysis of pathogenicity is inconclusive because of the large number of HK1 isoforms and homologs. Rare recessive null mutations in HK1 cause non-spherocytic hemolytic anemia, which was not observed in these families. However, a discrepancy between the phenotypes associated with dominant versus recessive mutations is common among inherited retinopathies. We conclude that a mutation, or mutations, in HK1 cause roughly 1% of adRP cases in these populations, revealing a novel retinal disease gene and possibly a common founder-effect mutation.

2906M

A novel disease-causing gene for Pelizaeus-Merzbacher disease. *M. Nafisinia[1], N. Sobreira[3], WA. Gold[1,2], L. Riley[1,2], R. Ouvrier[4,5], C. Boehm[3], J. Christodoulou[1,2,6].* 1) Genetic Metabolic Disorders Research Unit, Western Sydney Genetics Program, the Children's Hospital at Westmead, Sydney, NSW, Australia; 2) Discipline of Paediatrics & Child Health, Sydney Medical School, University of Sydney, Sydney, NSW, Australia; 3) Institute of Genetic Medicine, Johns Hopkins Univ School Med, Baltimore, MD, USA; 4) Neurosurgery, Children's Hospital at Westmead, Sydney, NSW, Australia; 5) Children's Hospital at Westmead, Sydney, NSW, Australia; 6) Discipline of Genetic Medicine, Sydney Medical School, University of Sydney, Sydney, NSW, Australia.

Background: Pelizaeus-Merzbacher disease (PMD) is a rare Mendelian disorder characterized by central nervous system hypomyelination. The disease is caused by mutations in several genes, which are associated with delay in motor abilities and intellectual function. Aim: To identify the genetic cause of disease in two siblings with a clinical picture of PMD but negative for mutations in the known PMD genes PLP1 and GJC2. Patients & Methods: Two sibs with a clinical diagnosis of PMD were born to non-consanguineous parents of Maltese background. The affected brother had motor and expressive delay, bilateral horizontal nystagmus, an action tremor, head titubation, an ataxic gait and cognitive impairment. When last examined at 17 years of age, he had normal muscle bulk with slightly reduced strength, brisk deep tendon reflexes, ankle clonus and upgoing plantar reflexes bilaterally. His brain MRI scan with fluid attenuated inversion recovery sequence confirmed a dys- or demyelinating process in the central and non-myelination in the peripheral white matter. His sister was less severely affected than her brother with mild nystagmus, ataxic gait and anxiety attacks. When last examined at age10, she had increased tone in her lower limbs, mild cerebellar signs, moderate cognitive impairment, truncal ataxia and pendular nystagmus. Whole exome sequencing (WES) was used to screen for likely pathogenic variants followed by western blotting, and temperature sensitivity and auxotrophic studies of patient fibroblasts to confirm pathogenicity. Results: WES uncovered a homozygous mutation in the arginyl-tRNA synthetase (RARS) gene (c.5A>G, p.Asp2Gly), which has recently been associated with a hypomyelination syndrome. Protein levels of RARS, and those of a binding partner, KARS (Lysyl-tRNA synthetase), were found to be significantly reduced by 80%; and 65%; respectively using western blotting of patient fibroblast extracts. As RARS is involved in protein synthesis where it attaches arginine to its cognate tRNA, patient cells were tested to determine their ability to function without this essential amino acid. Patient fibroblasts cultured in limited arginine at 30°C, showed a significant reduction ($p<0.001$) in viability compared to control cells indicating the inefficiency of protein synthesis in the patient cells. Conclusion: Here we provide evidence for a novel PMD-causing gene. Screening of a larger PMD cohort is in progress to further delineate the RARS phenotype.

2907T

Identification of a homozygous CLN5 mutation p.S312N in a family with adult-onset cerebellar ataxia. *A. Brusco[1,2], S. Nassani[3], Y. Guo[4], E. Giorgio[1], A. Calcia[1], X. Liu[5], E. Di Gregorio[2], S. Cavalieri[2], E. Pozzi[1], A. Brussino[1], Y. Xie[5], F. Wang[4], L. Tian[4], W. Chen[5], B. Nmezi[6], Q.S. Padiath[6], H. Jiang[5,7], A. Kyttala[8], N.R. Pizio[3], H. Hakonarson[4,9,10], C. Mancini[1].* 1) Department of Medical Sciences, University of Torino, Italy, Torino, Italy; 2) A.O. Città della Salute e della Scienza, S.C.D.U. Medical Genetics, Torino, Italy; 3) U.O. Neurology Ospedale di Lavagna, Genova, Italy, Genova, Italy,; 4) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, United States,; 5) BGI-Shenzhen, Shenzhen, China,; 6) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 7) The Guangdong Enterprise Key Laboratory of Human Disease Genomics, BGI-Shenzhen, Shenzhen, China,; 8) National Institute for Health and Welfare (THL) Public Health Genomics Unit, FIN-00251, Helsinki, Finland,; 9) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, United States,; 10) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States.

The NCLs or Batten Diseases are a group of progressive neurodegenerative disorders characterized by rapid death of cortical neurons and intracellular accumulation of autofluorescent lipopigment material (lipofuscin) in the body's tissues. Fourteen genetically distinct human NCLs were identified. Clinically NCLs are characterized by a combination of cerebellar ataxia, visual impairment, and variable association of seizures, behavioral disturbances, and cognitive deterioration. CLN5 mutations (13q21.1-q32) are reported in severe autosomal recessive forms of variant Late Infantile Neuronal Ceroid Lipofuscinoses (vLINCL, OMIM*608102). The clinical course includes progressive intellectual and motor deterioration, seizures, and visual failure. Typical CLN5 forms have an onset between 4-7 yrs. Here, we describe two italian siblings affected by an adult form of ataxia, who presented with a homogeneous phenotype characterized by onset between 54-56 yrs, walk difficulties, dysarthria. Progressive cognitive decline associated to visual loss, ascribed to glaucoma, appeared only after few years of disease. Neurological evaluation at 61yrs. showed nystagmus and head/truncal tremor, and no sensory-motor neuropathy. MRI showed severe cerebellar atrophy and mild cortical atrophy of both hemispheres. Exome sequencing identified a homozygous c.935G>A (p.S312N) transition in CLN5 gene, never reported before nor present in Exome variant Server (http://evs.gs.washington.edu/). Alignment of orthologous sequences and in silico softwares predict the substitution Ser>Asn to be disease causing. We studied the subcellular localization of the mutated protein, transfecting the CLN5*S312N-pCMV plasmid in Hek293 cells: as reported for vLINCL mutations, p.S312N seemed to affect trafficking and maturation of the CLN5 protein, resulting in its accumulation in ER compartment and indicating that misfolding may be the major cause for its retention. We speculate this mutation acts as mild or modifier genes/epigenetics factors may contribute to this late clinical onset. Our results further demonstrated the power of exome sequencing in identifying genes disease causing, especially when linked to atypical phenotypes.

2908S

Search for missing regulatory region mutations at the DFNB1 locus in GJB2 heterozygotes with deafness. *J. Foster, G. Bademci, M. Tekin.* Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL.

Approximately 30-40% of congenital profound hearing loss is attributed to autosomal recessive mutations in the DFNB1 locus that contains both GJB2 and GJB6 genes which code the proteins Cx26 and Cx30, respectively. Recent reports show that loss of GJB2 expression in deaf probands can result from a gross deletion in the DFNB1 locus upstream of GJB6 when in trans with heterozygous mutations in GJB2. These results support the existence of the putative cis-regulatory element upstream of GJB6. We identified 21 deaf probands who remained heterozygous for a single variant after sequencing both exons of GJB2 in our deafness repository that includes 957 families. In two large families with multiple affected and unaffected children, all affected children were heterozygous for a GJB2 mutation (family 1:c.35delG, family 2: c.-23+1G>A), suggesting that their deafness is related to the DFNB1 locus. We excluded previously reported gross deletions in the DFNB1 locus (GJB6-D13S1830 and D13S1845, chr13:19,837,343-19,968,698, and a >920kb deletion involving both GJB2 and GJB6) via CNV analysis with quantitative PCR. Here we propose these two families as strong candidates for the investigation of the putative cis-regulatory element predicted to be in the 100kb region upstream of GJB6. We are utilizing a custom enrichment in the DFNB1 locus upstream of GJB6 for targeted sequencing. We are also performing whole exome sequencing to rule out mutations in other genes. Our results will help to further characterize the regulatory mechanism through which GJB2 is expressed.

2909M

Identifying causative gene variants for hearing loss using a target enrichment/next generation strategy. *D. Tekin[1], D. Yan[1], S.H. Blanton[2], M. Tekin[2], X.Z. Liu[1,2].* 1) Department of Otolaryngology, University of Miami Miller School of Medicine, Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL.

The extensive genetic heterogeneity of hearing loss makes single gene testing inefficient, except for the most common gene GJB2. Therefore, target enrichment/next generation strategy is more time and cost-efficient considering the translation of research data into clinical care. We have developed a custom capture panel (Miami Hearing Genes) of 146 known deafness genes with a target size of approximately 1MB (Agilent Sure Select DNA Design). The genes were identified by searching databases including Hereditary Hearing Loss Homepage, RefSeq, Ensembl, The Human Genome Mutation Database (HGMD), Online Mendelian Inheritance in Man (OMIM) and recent peer-reviewed publications related to the genetics of deafness. The design covered all coding exons, 5' and 3' untranslated regions (UTRs) and 25 bases of intronic flanking sequences for each exon. The overall in silico coverage of the design was 97.47%. The mitochondrial genome with 16,520 kbp was also blended into the design. To validate our panel, we used a set of 16 DNA samples obtained from patients with hearing loss. There were 8 positive controls with known mutations and 8 unknown samples. Analysis detected all known variants in nuclear and mitochondrial genes. These results prove the accuracy and reliability of the custom capture experiment. The search for the causative variations in candidate genes for the 8 unsolved samples and the screen on additional 200 deaf probands from nonsyndromic hearing loss families are in progress.

2910T

Application of whole exome sequencing for identification of deafness causative genes in small families. *D. Yan[1], M. Grati[1], D. Tekin[1], R. Mittal[1], Q. Ma[1], LL. Du[1], Q. Liu[1,5], J. Qing[1,4], Z. Ahmed[3], XZ. Liu[1,2,4].* 1) Departments of Otolaryngology-Head and Neck Surgery, University of Miami Miller School of Medicine, Miami, Florida 33136, USA; 2) Dr. John T. Macdonald, Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL 33136, USA; 3) Abrahamson Pediatric Eye Institute, Otolaryngology Head & Neck Surgery, Cincinnati Children's Hospital Medical Center 3333 Burnet Avenue ML 7003, Cincinnati, OH 45229; 4) Department of Otolaryngology, Head and Neck Surgery, Institute of Otology, the Second Xiangya Hospital, Central South University, Changsha, Hunan, 410011, China; 5) Department of Otolaryngology, Head and Neck Surgery, Chinese PLA, Beijing, 100853, China.

Hereditary deafness is a genetically heterogeneous disorder, with wide range of symptoms through clinical variation, including difference in frequencies affected, hearing threshold changes and age of onset. The causes of hearing loss (HL) may be genetic, environmental, or multifactorial. Although tremendous progress has been made in our understanding of the molecular basis of hearing and HL, the identification of genes and gene defects that affect the process of hearing remains challenging. Up to date, there are more than 60 genes known to be involved in non syndromic HL (NSHL). Nevertheless, it is estimated that there are many more genes to be discovered. Furthermore, for countless families too small for the conventional linkage analysis, the genetic cause of their HL is still an unsolved problem. For those cases, now the availability of whole-exome sequencing (WES) involving the targeted sequencing of the protein-coding subset of the human genome, makes it possible to efficiently identify novel causative genes and mutations. In this study we perform WES on 4 families negative for common deafness genes mutations that are not amendable to conventional approaches. The SureSelect human all exon 50Mb kit (Agilent Technologies) was used for exons and flanking intronic sequences enrichment and the sequencing was performed on the Hisq2000 instrument (Illumina). The Genomes Management Application (GEMapp; https://secureforms.med.miami.edu/hihg/gem-app) was applied for data filtering. Computational functional prediction algorithms (PolyPhen, SIFT, MutationTaster) and conservation scores (PHASTCO, GERP, PHYLOP) were also applied. Aproximately, 92,000 single nucleotide variants and 9,206 INDELS per sample were obtained before variant filtering. Coverage of targeted exons for >10 reads were ranged from 90.3% to 93.5% and >20 reads from 80% to 83.5%. By our filtering strategy, we have rapidly identified homozygous and heterozygous missense variations in novel genes in families presenting autosomal recessive and dominant NSHL, respectively. Sanger sequencing confirmed cosegregation of the variants with the disease phenotype in each family. *In vitro* studies for confirmation of the pathological nature of the sequence changes are being performed and identification of new NSHL genes in multiplex families that are negative for all known deafness genes is in process. Our study shows the great potential of novel deafness gene discovery in small families utilizing WES.

2911S

Genetic analysis of Rubinstein-Taybi Syndrome. *S. de Boer[1], F. Rossello[2], S. White[1].* 1) MIMR-PHI Institute of Medical Research, Monash University, Clayton, Victoria, Australia; 2) Victorian Bioinformatics Consortium, Monash University, Clayton, Victoria, Australia.

Rubinstein-Taybi syndrome (RTS) is a rare congenital disorder. It affects ~1 in 100,000 individuals and is characterized by intellectual disability, growth delay, broad thumbs and big toes, organ malformations, behavioural problems, and specific facial features. Diagnosis of RTS is essentially based on clinical presentation and can be confirmed by genetic screening. Most cases are due to *de novo* mutations, and causative mutations can be identified in <60% of RTS patients, affecting a single copy of either *CREBBP* or *EP300*. The mutation distribution in RTS is uneven, with *CREBBP* mutations being significantly more frequent than *EP300* mutations (~50% and 5% of cases respectively). The significant percentage of RTS cases without a genetic diagnosis suggests that RTS is a genetically heterogeneous disorder, and that there are additional genes involved in RTS. There are other syndromes characterised by clinical features overlapping with RTS, and some are caused by mutations in genes encoding CBP and/or P300-interacting proteins. Next generation sequencing (NGS) on the HiSeq 1500 (Illumina) or the Ion Proton (Life Technologies) has been used to search for RTS-causing mutations by sequencing the entire protein-coding DNA sequence (exome) of six RTS patients. Data analysis showed one patient with a stop/gain SNV in *CREBBP*, which was confirmed by Sanger sequencing. No *EP300* mutations were found in our patient cohort, and analysis is ongoing to generate a candidate gene list based on the Interactome and genes involved in diseases with similar characteristics to RTS.

2912M

The utility of clinical exome sequencing in identifying the genetic origins of eight unclassified developmental disorders in unique Canadian populations. *S.M.K. Farhan[1, 2], J. Wang[1], J.F Robinson[1], V.M. Siu[2,3,4], C.A Rupar[2,3,4], R.A. Hegele[1,2] FORGE Canada Consortium.* 1) Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University; 2) Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University; 3) Medical Genetics Program, Department of Pediatrics, London Health Sciences Centre; 4) Children's Health Research Institute, London Health Sciences Centre.

We are a part of two large national rare disease initiatives: FORGE and Care for Rare Canada, in which >300 rare diseases have been ascertained and diagnosed. Specifically, our group is investigating the genetic etiology of 8 developmental disorders, namely: (1) infantile mitochondrial complex II and III deficiency (IMC23D), (2) seizures, delay, and unusual facies (SDUF) syndrome, (3) epilepsy with ataxia, (4) osteopetrosis, (5) Angelman-like syndrome, (6) ichthyosis-microcephaly, (7) ataxia, dystonia, and mental retardation, and (8) nocturnal seizures with developmental delay. Collectively, these affect 27 patients in 11 families from different Canadian communities. Their clinical presentations vary in severity, symptoms and tissues affected. Six disorders were found in communities with high rates of consanguinity; consequently, we applied homozygosity mapping to generate candidate loci. Next, whole exome sequencing (WES) was performed on DNA from 19 patients across the 8 disorders. Using a non-synonymous, rare variant analysis, WES with in silico analyses, was applied to identify mutations for each disorder. For some disorders, we performed population screening and functional studies to supplement the clinical description. Using our approach, we have identified the likely causative mutations of 6 disorders thus far. While each disorder is clinically distinct, the process used to identify the causative gene is the same. Importantly, we have identified 5 new genes, heretofore never implicated in a human disease, which when mutated, can lead to developmental disorders. These include: (1) NFS1, an iron-sulfur cluster protein, which causes IMC23D; (2) EXT2, a heparan sulfate biosynthesis enzyme and a tumour suppressor, which leads to SDUF syndrome; (3) TMTC3, a protein whose function is still unknown, is depleted in nocturnal seizures with developmental delay; (4) FSD1, a centrosome associated protein with specific brain tissue expression, is mutated in patients with Angelman-like syndrome; and (5) NUDCD2, which interacts with lissencephaly associated protein PAFAH1B1, is mutated in patients with ataxia, dystonia, and mental retardation. We have also identified a known disease causing gene, KCTD7 as the cause of progressive myoclonus epilepsy. These studies allow for an academic and a clinical focus by allowing us to better understand the dynamic function and mechanistic significance of these genes and importantly, how they can underlie human disease.

2913T

Protein-altering rare variants in candidate genes in patients with Biliary Atresia. *R. Rajagopalan[1], E.A. Tsai[2], C.M. Grochowski[1], A. Falsey[1], K.M. Loomes[3,4], M. Devoto[3,6,7,8], N.B. Spinner[1,5,9].* 1) Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA., USA; 2) Genomics and Computational Biology Graduate Group, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA; 3) Department of Pediatrics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA; 4) Division of Gastroenterology, Hepatology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 5) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia PA, USA; 6) Division of Human Genetics,; 7) Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 8) Department of Molecular Medicine, University La Sapienza, Rome, Italy; 9) Division of Genomic Diagnostics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA.

Biliary atresia (BA) is a pediatric liver disease resulting in necroinflammatory obliteration of the extrahepatic biliary tree,presenting within the first few months of life. The incidence of BA is 1 in 14,000 in Caucasians. The etiology of BA is unknown to date, with evidence for immunologic, infectious, environmental, and genetic factors described. We hypothesized that rare variants in key genes in hepatobiliary development would be present in infants with BA. METHODS: Using a well-characterized cohort of patients with isolated BA, enrolled in an IRB-approved protocol through the NIDDK-funded Childhood Liver Disease Research and Education Network (ChiLDREN), we performed whole-exome sequencing in 100 unrelated Caucasian patients and looked for rare variants in 21 candidate genes. These genes were chosen based on previous reports of mutations found in the syndromic form of BA, animal models with hepatobiliary defects, and/or genome-wide association studies. We hypothesized that genetic susceptibility would be caused by rare variants and we therefore filtered for rare protein-altering sequence variants (nonsense, splice, non-synonymous variants, and small insertions/deletions) that were present at a less than 1% frequency threshold in any of the public datasets (ESP6500 and 1000 Genomes). RESULTS: A total of 20 missense variants (4 novel) and 2 in-frame insertions were identified in 13 genes in a total of X patients. We found more than one variant in LGR4, ZEB2, DNMT1, INVS, JAG1 & PRICKLE4 and some of these variants were shared among more than one proband. CONCLUSIONS: This suggests that these candidate genes may be enriched for protein-altering variants in BA patients. We are currently in the process of validating the candidate variants with Sanger sequencing for both confirmation of their presence and the mode of inheritance. These findings suggest that the underlying etiology of BA is likely to be highly heterogeneous and a single variant or gene may not explain the phenotypic complexity of the disease.

2914S

Whole genome sequencing of mummy DNA shows significant association with human disease phenotype. *S. Bhattacharya[1], J. Li[2], H. Lam[2], R. Lachman[3], N. Asadi[2], A. Butte[1], G. Nolan[4].* 1) Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA.,94305 USA; 2) Department of Bioinformatics, Bina Technologies, Redwood City, CA; 3) Stanford University Medical Center, CA; 4) Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford California 94305.

In 2012 an analysis was initiated of a mummified specimen (Ata) found in the Atacama Desert of Chile, South America. The 6-inch specimen with human or primate features manifested only 10 pairs of ribs (humans have 12), a skull shape that did not appear human, and apparently prematurely ossified growth plates suggesting a greater age at time of death than the size of the specimen would indicate. To determine possible genetic drivers of the observed morphology, DNA from the specimen was subjected to whole genome sequencing using the Illumina HiSeq platform-- 377,333,714 reads passed the set of filtering criteria. Using the BINA secondary analysis pipeline (ver 1.5.0-dev-217-ga8038cc), 97% reads successfully mapped; of these, 89.77% uniquely mapped to the human reference genome (GRCh37) with a set of decoy sequences; 7.03% mapped to multiple locations, and 3.20% were unmapped. The sequence reads were also aligned to the Chimpanzee genome (88.01% uniquely mapped). In total, 3,356,569 single nucleotide variations (SNVs) were found as compared to the human reference genome, 518,365 insertions and deletions (INDELs) detected, of which 81.5 % and 77.5% are in dbSNP, respectively. To identify putative causal variants associated with Ata's phenotype, the SNVs were run through ANNOVAR pipeline and confirmed by BINA's analytic platform. The SNV and INDEL mutations fell within the range of "human normal", ruling out a non-human primate origin for the specimen. A preliminary phenotype enrichment analysis of a subset of variants in genes that carry one or more deleterious mutations, using the hypergeometric test, showed a significant association with several phenotypes, including: proportionate short stature (adjP=9.5e-03), Aplasia involving forearm bones (adjP=3.15e-02), 11 pairs of ribs (adjP= 2.96e-02) and premature osteoarthritis (adjP=2.96e-02)—gene-disease phenotype knowledgebase information was obtained from Human Phenotype Ontology database. Disease enrichment analysis indicates that mutations in genes associated with dwarfism and bone diseases significantly enriched. These findings suggest there might be one or more Mendelian traits influencing the observed phenotype, strengthening the molecular basis for deriving understanding how genomic variations in this specimen lead to phenotypic consequences.

2915M

Excess of de novo variants in genes involved in chromatin remodeling and regulation of transcription in patients with marfanoid habitus and intellectual disability. *L. Faivre[1,2], F. Daoud[1], C. Cabret[1], J. St-Onge[1,2], E. Gautier[2], P. Callier[1,2], D. Lacombe[3], S. Moutton[3], MA. Delrue[3], C. Goizet[3], F. Morice-Picard[3], A. Munnich[4], S. Lyonnet[4], V. Cormier-Daire[4], M. Holder[5], F. Petit[5], B. Leheup[6], J. Thevenon[1], S. Odent[7], PS. Jouk[8], P. Sarda[9], P. Collignon[10], D. Maryin-Coignard[11], A. Jacquette[12], G. Collod-Beroud[13], C. Boileau[14], G. Jondeau[14], C. Thauvin-Robinet[1,2], Y. Duffourd[1,2], JB. Rivière[1,2].* 1) EA4271, Université de Bourgogne, Dijon, France; 2) FHU TRANSLAD, CHU Dijon, France; 3) Département de Génétique, CHU Bordeaux, France; 4) Institut Imagine, Hôpital Necker Enfants Malades, Paris, France; 5) Département de Génétique, CHU Lille, France; 6) Département de Génétique, CHU Nancy, France; 7) Département de Génétique, CHU Montpellier, France; 8) Département de Génétique, CHU Rennes, France; 9) Département de Génétique, CHU Grenoble, France; 10) Service de Génétique, CH Toulon, France; 11) Service de Génétique, CH Le Mans, France; 12) Institut Fédératif de Génétique, Hôpital Pitié Salpétrière, Paris, France; 13) INSERM UMR_S910, Faculté de Médecine Marseille, France; 14) Centre de Référence Maladie de Marfan, Hôpital Bichat, France.

Marfanoid habitus (MH) combined with intellectual disability (ID) is a genetically and clinically heterogeneous group of overlapping disorders. In 2013, we reported our genetic investigations of 100 patients with MH and ID. A custom 244K array CGH and targeted sequencing of 6 genes (MED12, ZDHHC9, UPF3B, FBN1, TGFBR1 and TGFBR2) led to identification of the underlying genetic cause in 20% of subjects, which included non-overlapping chromosomal anomalies, mutations of FBN1, and a mutation of MED12 in fourteen, three and one individuals, respectively. In addition, two subjects carried both a FBN1 mutation and a chromosomal rearrangement, thus suggesting digenic inheritance for at least a fraction of individuals with MH and ID. To further decipher the genetic basis of MH with ID, we performed exome sequencing in 35 affected individuals with a systemic score of at least seven, using a trio approach in 15 subjects, whereas the remaining 20 were sequenced without their parents. The trio-based analysis identified a higher than expected rate of non-synonymous de novo changes (32 mutations, average of 2.1 per trio), thus adding further support to the hypothesis of a polygenic inheritance for MH and ID. Searching for rare variants of genes identified by the trio-based analysis in the remaining 20 index cases combined with segregation analysis in available relatives led to identification of six genes with de novo mutations in at least two unrelated individuals. Overall, at least one likely pathogenic mutation was identified in ~50% of subjects (17/35), with mutations in eight genes previously associated with a Mendelian disease. Among them, two are known to be mutated overgrowth syndromes. Strikingly, a half of the identified mutated genes belong to a highly interconnected network of proteins involved in chromatin remodeling and regulation of transcription, thus highlighting molecular pathways that may be targets for future research efforts. In conclusion, these findings provide additional information about the genetic and molecular basis on MH and ID, with an increased de novo rate, an enrichment of de novo mutations in several genes, and the identification of chromatin remodeling and regulation of transcription as key cellular pathways in the pathogenesis of these phenotypes. Finally, our data demonstrate the added value of studying sub-phenotypes of ID for identifying genes associated with this extremely heterogeneous and complex group of disorders.

2916T

Exome analysis of 116 patients supposed to be autosomal recessive hereditary spastic paraplegia established molecular diagnoses of 49 patients with broad genetic heterogeneities. *H. Ishiura[1], K. Koh[2], H. Shimazaki[3], J. Mitsui[1], Y. Takahashi[4], J. Goto[1], K. Yoshimura[5], K. Doi[5], S. Morishita[5], H. Sasaki[6], Y. Takiyama[2], S. Tsuji[1,7], JASPAC (Japan Spastic Paraplegia Research Consortium).* 1) Department of Neurology, The University of Tokyo, Tokyo, Japan Tokyo, Tokyo, Japan; 2) Department of Neurology, Yamanashi University, Yamanashi, Japan; 3) Department of Neurology, Jichi Medical University, Tochigi, Japan; 4) Department of Neurology, National Center of Neurology and Psychiatry, Tokyo, Japan; 5) Department of Computational Science, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan; 6) Department of Neurology, Hokkaido University Graduate School of Medicine, Hokkaido, Japan; 7) Medical Genome Center, The University of Tokyo Hospital, The University of Tokyo, Tokyo, Japan.

[Background] Hereditary spastic paraplegia (HSP) is a clinically and genetically heterogeneous neurodegenerative disorder. To date, SPG1-72 was identified as the disease loci, of which 56 causative genes have been identified. Autosomal recessive HSP (AR-HSP) is more clinically heterogeneous than autosomal dominant HSP (AD-HSP). A previous study revealed that causative mutations were found in only 17.2 % of patients who was suspected to have autosomal recessive (AR) inheritance (Ishiura et al. 2014). [Methods] We recruited 116 HSP patients who were supposed to have AR inheritance because 1. they had parental consanguinity with or without affected sibling(s), 2. they had affected sibling(s) without affecteds in the previous generations, or 3. they were sporadic but had complicated form HSP. The previous analyses with resequencing microarray for 13 genes, aCGH for 16 genes, and Sanger sequencing for 3 genes were performed in 46 patients. Subsequently, one hundred and four patients underwent exome analysis. Variants with minor allele frequency of < 0.2 % in 56 causative genes for HSP and 203 causative/susceptibility genes for other neurodegenerative diseases were considered to be candidate mutations. Nonsense, insertion/deletion, splice site, and known missense mutations were considered to be pathogenic. In the case of novel missense mutations, we considered them as pathogenic if 1. more than one pathogenic missense mutations have been described. and 2. the identified mutations are located in the vicinity of the region in genes such as *SPAST* or *KIF5A* where pathogenic mutations have been described to be clustered. [Results] We found 33 patients and 5 patients with mutations in AR-HSP and AD-HSP genes, respectively. SPG11 (12.1 %), SPG28 (4.3 %), SPG46 (3.4 %), and SPG15 (2.6 %) were frequent subtypes of AR-HSP. In addition, 8 patients with AR neurodegenerative disease such as spinocerebellar ataxias and leukoencephalopathies and 3 patients with autosomal dominant neurodegenerative diseases were found. [Discussion] Exome analysis revealed pathogenic mutations in 42 % of patients suspected to have AR inheritance. Although exome sequencing is highly efficient for the molecular diagnosis of HSP, assessment of pathogenicity of missense mutations is still a challenging issue. Mutations were found in 23 genes, indicating clinical heterogeneity of these patients. The causative genes for the rest of the patients remain to be further elucidated.

2917S

New candidate genes associated with autosomal dominant partial epilepsy with auditory features identified by whole exome sequencing. *F.R. Torres[1,3], P.A. Oliveira[1,3], R. Secolin[1,3], C.V. Soler[1,3], M.G. Borges[1,3], A.C. Coan[2,3], M.E. Morita[2,3], B.S. Caravalho[1,3], F. Cendes[2,3], I. Lopes-Cendes[1,3].* 1) Department of Medical Genetics, Unicamp, Campinas, SP, Brazil; 2) Department of Neurology, Unicamp, Campinas, SP, Brazil; 3) The Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, SP, Brazil.

Background: Epilepsy is a common chronic neurological disorder that affects approximately 1% of the population worldwide. Autosomal dominant partial epilepsy with auditory features (ADPEAF) is a clinically well characterized syndrome. Deleterious mutations in *LGI1* are present in about 50% of families segregating ADPEAF. Objective: To identify genes associated with ADPEAF in two large families with no mutations identified in *LGI1*. Methods: Whole exome sequencing was performed in nine individuals from two ADPEAF families (F- 1 and F- 2), including seven patients and two unaffected parents. Exome was targeted with Nextera Rapid Capture Expanded Exome kit (Illumina®) and sequenced in a high-performance Hiseq Illumina® 2500 sequencing machine (Illumina®) to obtain >50X average coverage per sample. Bioinformatics analysis was performed using the GATK software package. Sequences were aligned using the BWA algorithm. Variant calling and in silico functional prediction of mutations was performed using VariantAnnotator and SnpEff tools. We prioritized non-synonymous, frameshift, splicing, and indel variants according to novelty, quality score, and putative pathogenicity. Results: We found a total of 1,007,575 and 1,006,376 variants in F-1 and F-2 families, respectively. After bioinformatics analysis, we identified 54 variants in F-1 and 83 in F-2, which are shared by all patients and are absent in unaffected individuals. Among these we observed mutations in eight genes that are potentially associated with ADPEAF. Conclusions: Putative roles related to formation of axon connections, protein-protein interaction, neuronal differentiation and ionic channels make the eight candidate genes identified relevant for ADPEAF.

2918M

Hoyeraal Hreidarsson syndrome, a severe variant of dyskeratosis congenita, caused by biallelic mutations in TPP1. *B.J. Ballew[1], H. Kocak[2], K. Bisht[2], J.F. Boland[3], B.D. Hicks[3], L. Burdett[3], A. Vogt[3], A.A. Hutchinson[3], N. Giri[1], B.P. Alter[1], M. Yeager[3], C. Keegan[2], J. Nandakumar[2], S.A. Savage[1], NCI DCEG Cancer Sequencing Working Group.* 1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), Rockville, MD, USA; 2) University of Michigan, Ann Arbor, MI, USA; 3) Cancer Genomics Research Laboratory, Leidos Biomedical Research, NCI-Frederick, Rockville, MD, USA.

Dyskeratosis congenita (DC) is an inherited telomere biology disorder resulting in heightened risk of bone marrow failure, cancer, pulmonary fibrosis, liver disease, and other complications. Leukocyte telomeres <1st percentile for age are diagnostic. Causative germline mutations in any of 9 telomere genes (*DKC1*, *TERC*, *TERT*, *TINF2*, *NOP10*, *NHP2*, *WRAP53*, *CTC1*, or *RTEL1*) are present in ~70% of DC. Gene discovery is complicated by incomplete penetrance, variable expressivity, and genetic anticipation, but is important to confirm diagnosis and for family planning. DC genes also provide insight into potential cancer susceptibility loci: polymorphisms in 6 DC genes are associated with cancer risk in the general population. We performed whole exome sequencing on mutation-negative DC families in our IRB-approved cohort study. Based on data from the NHLBI Exome Sequencing Project, 1000 Genomes, and an in-house database of 1400 exomes, we identified rare variants. We further filtered variants by family structure, examining all plausible inheritance models. We then combined our naïve analysis with a candidate-gene approach, highlighting variants in any of ~3000 genes associated with DC-related phenotypes. Finally, we analyzed variants' impact using in silico techniques. This led to the discovery of novel mutations in *TPP1* (*ACD*, MIM 609377), part of the telomeric shelterin complex, in a DC family. The proband of family NCI-275 is a 5 year-old boy with Hoyeraal Hreidarsson syndrome, a severe variant of DC that includes cerebellar hypoplasia, immunodeficiency and intrauterine growth retardation. The proband, his older sister, and father have telomeres <1st percentile; all three harbor a single-amino acid deletion in the TEL patch of TPP1. The proband also has a missense TPP1 mutation in the TIN2 binding domain; his mother is a silent carrier of this mutation. Functional analyses demonstrated that the deletion severely compromises both telomerase processivity and recruitment to telomeres. The missense mutation does not directly impact telomerase processivity or recruitment; however, it reduces TPP1-TIN2 binding, which may affect recruitment of TPP1 to telomeres. Modeling of the deletion using the crystal structure of TPP1-OB as a template suggests a structural mechanism for the defects. These data demonstrate that the compound heterozygous inheritance of these two alleles, each with defects in different aspects of TPP1 function, results in clinically severe DC.

2919T

Targeted next generation sequencing in DNA diagnostics for familial cancer. *A.H. van der Hout, Y.J. Vos, B. Raddatz, K. van Dijk-Bos, A. Knopperts, L.K. Leegte, J. ter Beest, H. Westers, R. Sinke, R. Sijmons.* Genetics, University Medical Center Groningen, Groningen, Netherlands.

Purpose Phenotypes of many hereditary tumour syndromes are known to overlap. In addition, different genes may underlie one syndrome. Therefore the number of genes to examine in a particular clinical case can be relatively large. Next generation sequencing (NGS) makes it is possible to analyze large numbers of genes in parallel at relatively low cost. We studied the expediency of a targeted NGS gene panel approach in DNA diagnostics for a group of 100 patients with either early-onset breast cancer or colorectal cancer in whom a pathogenic mutation in the *BRCA1/2* genes was already excluded by Sanger sequencing, or their tumours showed no signs of Lynch syndrome, respectively. Methods We developed and validated a sequencing panel based on Agilent Sure Select Target Enrichment® for mutation scanning in 71 genes known to be associated with tumour syndromes.. The samples were sequenced using 151 base pair paired-end reads on an Illumina MiSeq® sequencer and analyzed using Softgenetics' NextGENe® and Cartagenia's Benchlab NGS® software. The 71 genes are divided in three virtual, non-overlapping gene subpanels, based on the levels of preventive options and strength of risk information, in which for genes in subpanel 1 the most information is available about preventive options and specific tumour risks. In pre-test genetic counselling the genes to be tested were discussed as the 3 subpanels rather than individually. Patients can choose for any of the 3 subpanels to be tested and have the results returned to them. In addition, all 71 genes are tested in all patients for research purposes after de-identifying the patients. Results and Conclusion Approximately 65% of the patients chose to have results returned for all subpanels, 30% for subpanels 1 and 2 and 5% for subpanel 1 only. In 5 cases a pathogenic mutation was detected, 4 times a heterozygous mutation in *CHEK2* and one mutation in *RAD51D*. Heterozygosity for a pathogenic mutation in *MUTYH* was detected in 3 additional cases. In 13 cases one or two mutations classified as 'likely pathogenic' were detected in 13 different genes. Co-segregation and tumor analysis in the families may provide more information about pathogenicity of these mutations. In most cases the patient's phenotype does not match with the currently known tumour spectrum associated with the mutated gene. Analysis of large gene panels may broaden our knowledge of tumour spectra associated with mutations in certain genes.

2920S

Mutation screening of retinal dystrophy patients by targeted capture from tagged pooled DNAs and next generation sequencing. *M.E. Elasrag[1], C.M. Watson[3], D.A Parry[2], J.E Morgan[2], C.V Logan[1], I.M Carr[2], E. Sheridan[2], R. Charlton[3], C.A Johnson[1], G. Taylor[3], C. Toomes[1], M. McKibbin[1,4], C.F Inglehearn[1], M. Ali[1].* 1) Section of Ophthalmology & Neuroscience, Leeds Institute of Biomedical & Clinical Sciences, University of Leeds, UK; 2) Section of Genetics, Leeds Institute of Biomedical & Clinical Sciences, University of Leeds, UK; 3) Yorkshire Regional Genetics Service, St. James's University Hospital, Leeds, UK; 4) Department of Ophthalmology, St. James's University Hospital, Leeds, UK.

Retinal dystrophies are genetically heterogeneous, resulting from mutations in over 200 genes. Prior to the development of massively parallel sequencing, comprehensive genetic screening was unobtainable for most patients. Identifying the causative genetic mutation facilitates genetic counselling, carrier testing and prenatal/pre-implantation diagnosis, and often leads to a clearer prognosis. In addition, in a proportion of cases, when the mutation is known treatment can be optimised and patients are eligible for enrolment into clinical trials for gene-specific therapies. The genomic DNA of twenty patients each from a different family was sheared, tagged and pooled in batches of four samples, prior to targeted capture and next generation sequencing. The enrichment reagent was designed against genes listed on the RetNet database. Sequence data were aligned to the human genome, variants were filtered to identify potential pathogenic mutations, which were confirmed by Sanger sequencing. Molecular analysis identified likely pathogenic mutations in 12 cases, many of them known and/or confirmed by segregation. These included previously described mutations in ABCA4 (c.6088C>T,p.R2030*; c.5882G>A,p.G1961E), BBS2 (c.1895G>C,p.R632P), GUCY2D (c.2512C>T,p.R838C), PROM1 (c.1117C>T,p.R373C), RDH12 (c.601T>C,p.C201R; c.506G>A,p.R169Q), RPGRIP1 (c.3565C>T,p.R1189*) and SPATA7 (c.253C>T,p.R85*) and new mutations in ABCA4 (c.3328+1 G>C), CRB1 (c.2832_2842+23del), RP2 (c.884-1G>T) and USH2A (c.12874A>G,p.N4292D). We report here that tagging and pooling DNA prior to targeted capture of known retinal dystrophy genes identified mutations in 60% of cases. This relatively high success rate may reflect enrichment for consanguineous cases in the local Yorkshire population, and the use of multiplex families. Nevertheless this is a promising high throughput approach to retinal dystrophy diagnostics.

2921M

Homozygosity mapping and exome sequencing of a Faroese family with albinism. *K. Grønskov[1], K. Vilhelmsen[2], T. Rosenberg[3, 4].* 1) Clinical Genetic Clinic, Kennedy Center, Rigshospitalet, Glostrup, Denmark; 2) , Dept. of Ophthalmology, National Hospital of The Faroe Islands, FO-100 Torshavn, The Faroe Islands; 3) The National Eye Clinic for the Visually Impaired, Kennedy Center, Glostrup Hospital, Glostrup, Denmark; 4) Gordon Norrie Center for Genetic Eye Diseases, Copenhagen, Denmark.

Albinism is a genetically heterogeneous disorder characterized by a number of specific deficits in the visual system such as nystagmus, low vision, and crossed asymmetry of the visual pathway in association with a variable hypopigmentation phenotype. The lack or reduction of pigment might affect the eyes, skin and hair (oculocutaneous albinism, OCA [MIM 203100]), or the eyes only (ocular albinism, OA [MIM 300500]). In a lightly pigmented population such as the Faroese, it can be difficult to distinguish between OCA and OA. Mutations in six genes (*TYR* [MIM 606933], *OCA2* [MIM 611409], *TYRP1* [MIM 115501], *SLC45A2* [MIM 600202], *SLC24A5* [MIM 609802] and *C10orf11* [MIM 614537]) are known to cause autosomal recessive OCA and furthermore, one locus (OCA5 [MIM 615312]) has been identified. A founder mutation in *C10orf11* was found in six out of nine Faroese families with OCA/OA. Mutations in one gene, *GPR143* [MIM 300808], are known to cause X-linked OA. More genes are expected to be associated with OCA/OA. A consanguineous family from The Faroe Islands with two children affected with OCA/OA and two unaffected children were investigated. Mutation analysis of *TYR*, *OCA2*, *SLC45A2* and *C10orf11* in affected individuals showed no pathogenic changes. Homozygosity mapping showed 12 homozygous regions with approximately 100 genes. Exome sequencing will be performed.

2922T

Identifying novel genes that cause Rett syndrome by trio-based exome sequencing of MECP2-negative patients. *S.A. Sajan[1,2], S.N. Jhangiani[3], D.M. Munzy[3], R.A. Gibbs[3,4], J.R. Lupski[3,4,5], D. Glaze[1], W.E. Kaufmann[6], S.A. Skinner[7], M. Friez[7], A.K. Percy[8], J.L. Neul[1,2,4].* 1) Section of Child Neurology and Developmental Neuroscience, Department of Pediatrics, Baylor College of Medicine, Houston, TX; 2) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Department of Pediatrics, Baylor College of Medicine and Texas Children's Hospital, Houston, TX; 6) Department of Neurology, Boston Children's Hospital, Boston, MA; 7) Greenwood Genetic Center, Greenwood, SC; 8) Department of Pediatrics, University of Alabama at Birmingham, Birmingham, AL.

Rett syndrome (RTT [MIM#312750]) is a neurodevelopmental disorder which in the majority of cases is caused by mutations in Methyl CpG binding protein 2 (*MECP2*) on the X chromosome. The disorder affects about 1 out of 10,000 live female births and is typified by apparently early normal development in the first 6-18 months of life followed by psychomotor regression involving loss of speech and hand use, the development of gait problems and characteristic repetitive hand stereotypies. While some clinical subtypes of RTT, referred to as atypical RTT, have been shown to be caused by mutations in *CDKL5* and *FOXG1*, classical/typical RTT still remains an essentially monogenic disorder with de novo *MECP2* mutations accounting for 95-97% of such cases. We sought to identify the molecular basis of some of the 3-5% of typical RTT cases that are *MECP2*-negative by trio-based exome sequencing. We used the Illumina platform to sequence the exomes of 10 trios with RTT probands who did not harbor mutations in *MECP2* as determined by Sanger sequencing of coding exons as well as by copy number variation analysis on a clinical array. Data analysis using the Genome Analysis Toolkit (GATK) revealed 8 heterozygous de novo protein altering mutations in 8 different genes in 7 patients, with one patient having de novo mutations in two different genes. These two genes were *MDMX* and *ZNF536* with missense and frameshift insertion mutations, respectively. Missense mutations were also identified and confirmed in *GABRB2*, *IMPDH2*, and *SAFB2* and were predicted to be damaging by at least 2 out of 6 mutation prediction algorithms. Two non-frameshift deletions were found in *GRIN2B* and *IQSEC2*, and a splice site mutation was found in *TCF4*. Other mutations in the two nervous system-expressed ion channel genes *GRIN2B* and *GABRB2*, as well as in *TCF4* and *IQSEC2*, have previously been reported to result in developmental delay and intellectual disability. Our results suggest that *MECP2*-negative RTT is heterogeneous and caused by mutations in genes that also result in other types of neurodevelopmental disorders. We recommend including these genes in a list of high-priority genes when exome or targeted sequencing is performed on *MECP2*-negative RTT patients.

2923S

Baratela Scott Syndrome is a recessive skeletal dysplasia syndrome caused by disruption of the XYLT1 gene. *K. Sol-Church[1], M. Kircher[2], D. Stabley[1], K. Gripp[1], M. Bober[1].* 1) Nemours AI duPont Hospital for Children, Wilmington, DE., USA; 2) Department of Genome Sciences, Seattle WA.

Baratela Scott Syndrome (OMIM 300881) is characterized by skeletal dysplasia, facial features, and developmental delay. Skeletal findings include patellar dislocation, short tubular bones, mild metaphyseal changes, brachymetacarpalia with stub thumbs, short femoral necks, shallow acetabular roofs, and platyspondyly. Facial features include: a flattened midface with broad nasal bridge, cleft palate or bifid uvula and synophrys. Though some of the cognitive delays are typically masked by a warm and engaging personality, all of the patients demonstrated preschool onset of a cognitive developmental delay with a shortened attention span. We previously described seven male patients from six different families with Baratela Scott Syndrome (BSS), and performed Whole Exome Sequencing on the patients and available unaffected family members. We identified in a single individual a homozygous c.1290 -1G>A splice variant in the Xylosyltranferase (XYLT1) gene located on chromosome 16p. Both parents were heterozygous for this splice variant. This mutation caused splicing out of exon 6 and premature termination. In 3 patients from 2 unrelated families we discovered a large 3 Mb deletion covering 16p13-11 to 16p12.3 which includes the XYLT1 gene. Because of poor capture performance for exon 1 of the XYLT1 gene, we used Sanger sequencing and identified other variants in exon 1 of XYLT1 in two additional patients: One patient carried a c.319 G>T. Gly107Ter mutation while the other carried a 26 bp deletion (c.281-306). We conclude that Baratela Scott Syndrome is a recessive disorder caused by disruption of the XYLT1 gene.

2924M

Whole exome sequencing a consanguineous family in search for a novel genetic cause of Charcot-Marie-Tooth (CMT) disease. *S. Tey[1], N. Shahrizaila[2], K.J. Goh[2], A.P. Drew[3], M.L. Kennerson[3,4], A. Ahmad Annuar[1].* 1) Department of Biomedical Science, Faculty of Medicine, University of Malaya, Malaysia; 2) Department of Medicine, Faculty of Medicine, University of Malaya, Malaysia; 3) Northcott Neuroscience Laboratory, ANZAC Research Institute, Sydney, Australia; 4) Molecular Medicine Laboratory, Concord Hospital, Sydney, Australia.

Charcot-Marie-Tooth (CMT) neuropathy is a clinically and genetically heterogeneous group of disorders affecting the motor and sensory neurons with an estimated prevalence of 1 in 2500 people [1]. Over eighty genes have been identified for CMT and related peripheral neuropathies however there are many cases yet to be resolved [2]. In this study, we have taken a whole exome sequencing (WES) approach to investigate the genetic aetiology of CMT in two affected brothers from a consanguineous marriage. The onset of symptoms was in their teens and clinical examination revealed bilateral pes cavus with weakness of dorsiflexion resulting in foot drop. The proband had associated brisk reflexes whereas in his brother, the reflexes were normal/reduced but there was evidence of mild intellectual impairment. Nerve conduction studies indicated a demyelinating form of CMT. The segregation of the disease in the family suggests either an autosomal recessive or X-linked inheritance. WES was performed on two affected brothers, the parents and an unaffected brother. Mutations in known genes associated with autosomal recessive and X-linked CMT were found to be negative after querying the WES data. This suggests that the family may present a novel genetic cause of CMT. By analysing the WES data of the family members, 152 candidate variants were identified in the affected brothers and are not present in the unaffected family members. To narrow down the number of variants, we analysed our WES data using MERLIN [3] and PLINK [4] for linkage analysis and homozygosity mapping. We were able to reduce the number of candidate variants to only five. Currently we are investigating the potential involvement of these candidate variants in disease pathology. References: 1. Skre, H. Genetic and clinical aspects of Charcot-Marie-Tooth's disease. Clin Genet 6, 98-118 (1974). 2. Timmerman, V., Strickland, A. & Züchner, S. Genetics of Charcot-Marie-Tooth (CMT) Disease within the Frame of the Human Genome Project Success. Genes 5, 13-32 (2014). 3. Smith, K. et al. Reducing the exome search space for Mendelian diseases using genetic linkage analysis of exome genotypes. Genome Biology 12, R85 (2011). 4. Purcell, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81, 559-575 (2007).

2925T

Deciphering the genetic basis of idiopathic short stature. *C.T. Thiel[1], N.H. Hauer[1], K. Kessler[1], U. Uebe[1], A.B. Ekici[1], H. Sticht[2], H.-G. Doerr[3], A. Reis[1].* 1) Institute of Human Gentics, Friedrich-Alexander-University of Erlangen-Nuremberg, Erlangen, Germany; 2) Institute of Biochemistry Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany; 3) Department of Pediatrics and Adolescent Medicine Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany.

Shortness of stature is one of the most common pediatric concerns and has an incidence of 3 % in the general population. In 80 % of patients with growth deficit the etiology remains elusive in the absence of morphological details. Uncovering the genetic basis of short stature is therefore not only important for clinical diagnosis, prognosis and genetic counseling of affected individuals and their families, but is also a prerequisite for future development of therapeutic approaches. While Genome-wide association studies identified hundreds of common single nucleotide polymorphisms and copy number variants (CNVs) contributing to the height variation in the healthy population, we confirmed a frequent disease - rare variant hypothesis by the identification of pathogenic CNVs in 10% of patients with short stature. Our results implied a heterogeneity with more than 200 genes involved in short stature by power analysis. To address this hypothesis we thoroughly build a study group of more than 500 families with idiopathic short stature and performed whole exome sequencing (WES) in 60 trios. Thoroughly filtering and validation of the identified variants resulted in the identification of potential pathogenic variants in genes involved in epigenetic modification, cell cycle regulation, ubiquitination and protein synthesis. We found compound heterozygous inherited and de novo variants to be associated with short stature. In addition, including CNV analysis from WES further identified compound heterozygous variants. Thus, we were able to identify novel candidate genes for short stature. Analysis of further individuals with short stature will lead to a more elaborate and detailed view on mechanisms involved in growth regulation.

2926S

Coffin-Siris syndrome is a BAF complex disorder. *Y. Tsurusaki, M. Nakashima, H. Saitsu, N. Miyake, N. Matsumoto.* Yokohama City University Graduate School of Medicine, Yokohama, Japan.

Coffin-Siris syndrome (CSS; MIM 135900) is a congenital disorder characterized by intellectual disability, growth deficiency, microcephaly, coarse facial features, and hypoplastic or absent fifth fingernails and/or toenails. The majority of affected individuals represent sporadic cases, which is compatible with an autosomal dominant inheritance mechanism. The genetic cause for this syndrome has not been elucidated. We previously reported that five genes are mutated in CSS, all of which encode subunits of the Brahma-associated factor (BAF) (also known in yeast as the SWI/SNF) ATP-dependent chromatin-remodeling complex: *SMARCB1*, *SMARCA4*, *SMARCE1*, *ARID1A*, and *ARID1B*. In this study, we examined 49 newly recruited CSS-suspected patients, and re-examined three patients who did not show any mutations (using high-resolution melting analysis) in the previous study, by whole-exome sequencing or targeted resequencing. We found that *SMARCB1*, *SMARCA4*, or *ARID1B* were mutated in 20 patients. By examining available parental samples, we ascertained that 17 occurred *de novo*. All mutations in *SMARCB1* and *SMARCA4* were non-truncating (missense or in-frame deletion) whereas those in *ARID1B* were all truncating (nonsense or frameshift deletion/insertion) in this study as in our previous study. Our data further support that CSS is a BAF complex disorder. Aknowledgments: Drs. N Okamoto, H Ohashi, S Mizuno, N Matsumoto, Y Makita, M Fukuda, B Isidor, J Perrier, S Aggarwal, AB Dalal, A Al-Kindy, J Liebelt, and D Mowat are highly appreciated for contributing to this work.

2927M

Whole-Exome Sequencing and Linkage Analysis Reveal a Novel Genetic Locus for Autosomal Dominant Pattern Dystrophy of the Retinal Pigment Epithelium. *A. Vincent[1,2,3], N. Forster[2], J.T. Maynes[3,4], T.A. Paton[2,5], G. Billingsley[2], N. Roslin[2,5], A. Ali[2], J. Sutherland[1], T. Wright[1], C. Westall[1,3], A.D. Paterson[2,3,5], C.R. Marshall[2,5], . FORGE Canada Consortium[6], E. Heon[1,2,3].* 1) Department of Ophthalmology, Hospital for Sick Children, Toronto, Canada; 2) Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Canada; 3) University of Toronto, Toronto, Canada; 4) Department of Anaesthesiology, Hospital for Sick Children, Toronto, Canada; 5) The Centre for Applied Genomics, Hospital for Sick Children, Toronto, Canada; 6) FORGE Canada Consortium: Finding of Rare Disease Genes in Canada.

Purpose: To identify the genetic cause of autosomal dominant pattern dystrophy (PD) of the retinal pigment epithelium (RPE) in two families. **Methods and Results:** Eight members of a two generation family with PD (Family 1, five affected) were genotyped using Illumina's Infinium Linkage-24 chip. Multipoint genome wide linkage analysis identified 7 regions of potential linkage; microsatellite markers were used to refine these regions in an extended version of Family 1 (n = 12, six affected) and a maximum LOD score of 2.09 was observed across regions of four chromosomes. Exome sequencing on two affected family members identified 15 shared non-synonymous coding sequence variants within the linked regions; candidate genes were prioritized and further analyzed. Sanger sequencing confirmed a novel heterozygous missense variant (E79K) in *OTX2* that segregated with the disease phenotype. A second family (Family 2) with PD harbored the same missense variant in *OTX2*. A shared haplotype of 20.62 cM and 1,211 single nucleotide polymorphisms between the two families (HumanCoreExome chip) encompassing the *OTX2* gene suggest a shared ancestry. All cases from both families had decrease in distance vision and myopia. All except one affected case across the two families (aged 13 years) had a "pattern" at the macula. In vivo structural retinal imaging showed discrete areas of RPE-photoreceptor separation at the macula in all cases across both families; photoreceptor outer segments appeared disrupted in these regions. Electroretinogram testing showed generalized photoreceptor degeneration in three cases. Developmental anomalies observed include optic dysplasia (n = 4), microcornea (n = 1) and Rathke's cleft cyst (n = 1). **Conclusion:** This is the first report implicating *OTX2* to underlie PD phenotype. This phenotype resembles conditional mice models that show slow photoreceptor degeneration secondary to loss of *Otx2* function in the adult RPE. This allelic *OTX2* variant appears to spare pituitary hormonal abnormalities.

2928T

Cardiomyopathy Pathology of a GSDIIIa Patient Revealed by Whole Genome Sequencing. *Q. Zhao, S. Song, N. Zheng, F. Li, J. Huang, S. Hu.* State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

Cardiac involvements have been recognized in glycogen storage disease type III (GSD III, OMIM 232400) and the severity varies among patients, however, up to now little is known about the pathology. The endeavor to link the genotypes of AGL (the causal gene of GSD III) to cardiac phenotypes didn't find any significant associations. In this case study, we sequenced the whole genome of a GSD IIIa patient with severe heart failure on Illumina HiSeq 2500 system. Generally, cardiac involvement is apparent only after age 30 and the common phenotypes are ventricular hypertrophy. Our case was very special as she showed ventricle dilation and rapid progression to end stage heart failure at her early 20s. On physical examination the case was assessed in NYHA class IV. Echocardiography showed the enlargement of the heart and wall thickening accompanied with the left ventricular diastolic and systolic dysfunction. She received intensive anti heart failure therapy but got very limited improvements and then underwent heart transplantation after discussion with hepatology experts and approved by the Medical Ethics Committee in Fuwai Hospital. In her DNA we identified two causal mutations (c.1735+1G>T and c.G3299A) in AGL genes. Besides we also found the case harbored four damaging mutations in two genes (3 in TTN and 1 in SGCD) implicated in cardiomyopathies. Taken the clinical symptoms and DNA analysis together we conclude that the severe cardiomyopathy in our case was at least partly attributed to gene mutations.

2929S

PNPLA6 mutations in Laurence-Moon Syndrome (LMS) illustrate its distinct genetic etiology from Bardet-Biedl syndrome (BBS) and suggest its classification as part of a newly described neurodegenerative spectrum. *H. Dollfus[1,2], M. Prasad[1], E. Schaefer[1], D. Bonneau[3], C. Mutter[4], C. Stoetzel[1].* 1) Laboratoire de Génétique Médicale, INSERM U1112, Strasbourg, France; 2) Centre de référence pour les Affections Rares en Génétique Ophtalmologique (CARGO), Hôpitaux Universitaires de Strasbourg, France; 3) Service de génétique, CHU d'Angers, Angers, France; 4) Centre d'Investigations Cliniques (CIC) Hôpitaux Universitaires de Strasbourg, France.

There has been a long-standing confusion concerning Bardet-Biedl Syndrome (BBS) and Laurence-Moon Syndrome (LMS) as, historically, their names were linked together because of overlapping features. We have been studying a large family, with 5 affected siblings, reported as having LMS (Chalvon-Demersay at al., 1993) because of early onset retinal dystrophy, obesity, hypogonadism and spastic paraplegia (the patients did not have polydactyly or kidney impairment). In-depth clinical exploration identified cerebellar vermis atrophy in two patients that were wheel chair bound with spastic paresis. Moreover, we show that the retinal phenotype is remarkably distinct compared to BBS retinal dystrophy. In order to identify the underlying molecular alteration we performed, in 2010, a next-generation sequencing exome scan, but failed to find any pathogenic mutations with the initial commercial analysis pipeline. In 2014, an in-house analysis pipeline revealed a novel (p.[Arg1031Glnfs*38]; c. [3088_3091insAGCC]) mutation in the PLNLA6 gene. Homozygous mutations in PNPLA6 were reported at the same time in Boucher-Neuhäuser and Gordon Holmes syndromes (Synofzik et al, 2013). Due to poor coverage of the gene by exome sequencing, Sanger sequencing was used to identify a second mutation, a novel missense mutation (p.[Gly726Arg]; c.[2176G>C]) in a residue that is highly conserved through C. elegans. Both mutations segregated in the family. We suggest that Laurence-Moon syndrome enters the Boucher-Neuhäuser and Gordon Holmes spectrum in a broad group of neurodegenerative diseases and should not be considered as a part of the BBS and ciliopathy spectrum.

2930M

Microdeletion in the PITX2 Upstream Region in a Family with Axenfeld-Rieger Syndrome. *M. Walter, T. Footz.* Dept Med Gen, Univ Alberta, Edmonton, AB, Canada.

Purpose: Axenfeld-Rieger Syndrome (ARS) is a genetically-heterogeneous autosomal dominant disorder that affects ocular development and puts patients at an elevated risk of acquiring glaucoma, a neurodegenerative blinding disorder. DNA variations that cause ARS have been discovered in two transcription factor-encoding genes, PITX2 and FOXC1, but only account for about half of the reported cases of ARS.

Methods: To investigate the missing heritability of the syndrome, we commissioned whole-exome sequencing (WES) for three members of a family with autosomal dominant inheritance of ARS who lack mutation of the coding regions of PITX2 and FOXC1. DNA variants detected by WES were analysed by PCR-based targeted sequencing of additional family members. Quantitative genomic PCR was used to examine gene dosage.

Results: No non-synonymous coding region mutations that fully segregated with the disease were discovered. However, the correct segregation of alleles of five DNA variants at the PITX2 locus (a missense single nucleotide polymorphism (SNP), three intronic SNPs and a nearby intergenic microsatellite marker), led us to examine PITX2 gene copy number in the family. As a result, we discovered half-dosage of a subset of conserved regulatory elements directly upstream of the PITX2 transcript, but normal dosage of PITX2 exons.

Conclusions: Our results suggest that these ARS patients harbour a deletion of no more than 360 kilobases, the smallest known deletion that is restricted to the PITX2 upstream region. Since deletion of the entire upstream region is sufficient to cause ARS, we conclude that this small upstream segment contains PITX2 regulatory elements essential for normal ocular development and function.

2931T

Mutations in DOCK7 in individuals with epileptic encephalopathy and cortical blindness. *F.F. Hamdan[1], I. Perrault[2, 3], M. Rio[4], J-M. Capo-Chichi[1], N. Boddaert[5], J-C. Décarie[6], B. Maranda[7], R. Nabbout[8], M. Sylvain[9], A. Lortie[1], P. Roux[10], E. Rossignol[1], X. Gérard[2, 3], G. Barcia[8], P. Berquin[11], A. Munnich[4], G. Rouleau[12], J. Kaplan[2,3], J. Michaud[1], J-M. Rozet[2,3].* 1) CHU Sainte-Justine Research Center, MONTREAL, Canada; 2) INSERM UMR 1163, Laboratory of Genetics in Ophthalmology, 75015 Paris, France; 3) Paris Descartes - Sorbonne Paris Cité University, Imagine Institute, 75015 Paris France; 4) Department of Genetics, Hôpital Necker Enfants Malades, 75015 Paris, France; 5) Department of Pediatric Radiology, Hôpital Necker-Enfants Malades, APHP, Paris, Descartes University, 75015 Paris, France; 6) Department of Medical Imaging, Sainte-Justine Hospital, Montreal, Canada, H3T 1C5; 7) Division of Genetics, Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, Canada, J1H 5N4; 8) Department of Pediatric Neurology, Centre de Reference Epilepsies Rares, Hôpital Necker-Enfants Malades, APHP, Paris, Descartes University, 75015 Paris, France; 9) Division of Neurology, Centre Hospitalier Universitaire de Québec, Québec, Canada, G1V 4G2; 10) Institute for Research in Immunology and Cancer, University of Montreal, Montreal, Canada, H3C 3J7; 11) Department of Pediatric Neurology, CHU Amiens, 80054 Amiens Cedex, France; 12) Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada, H3A 2B4.

Epileptic encephalopathies are increasingly thought to be of genetic origin, although the exact etiology remains uncertain in many cases. We describe here three girls from two non-consanguineous families showing a clinical entity characterized by dysmorphic features, early-onset intractable epilepsy, intellectual disability, and cortical blindness. Brain imaging also showed specific changes in individuals from each family, including an abnormally marked pontobulbar sulcus as well as abnormal signals (T2 hyperintensities) and atrophy in the occipital lobe. Exome sequencing performed in the first family did not reveal any gene with rare homozygous variants shared by both affected siblings. It did, however, show one gene, DOCK7, with two rare heterozygous variants (c.2510delA (p.Asp837Alafs*48) and c.3709C>T (p.Arg1237*)) found in both affected sisters. Exome sequencing performed in the proband of the second family also showed the presence of two rare heterozygous variants in DOCK7 (c.983C>G (p.Ser328*) and c.6232G>T (p.Glu2078*)). Sanger sequencing confirmed that all three individuals are compound heterozygotes for these truncating mutations in DOCK7. These mutations have not been observed in public SNP databases and are predicted to abolish critical domains for DOCK7 function. DOCK7 codes for a Rac guanine nucleotide exchange factor that has been implicated in the genesis and polarization of newborn pyramidal neurons as well as in the morphological differentiation of GABAergic interneurons in the developing cortex. All together, these observations suggest that loss of DOCK7 function causes a previously unrecognized syndrome by affecting multiple neuronal processes.

2932S

Nonsense mutation in coiled coil domain containing 151 gene (*CCDC151*) causes Primary ciliary dyskinesia. *M. Erzurumluoglu[1], M.M. Alsaadi[2], S. Rodriguez[1], T.R. Gaunt[1], A.C. Al-Rikabi[4], P.A.I. Guthrie[1], K.K. Alharbi[3], H.Z. Omar[2], M. Mubarak[2], I.N.M. Day[1].* 1) Bristol Genetic Epidemiology Laboratories, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; 2) College of Medicine, King Saud University, Riyadh, Kingdom of Saudi Arabia; 3) College of Applied Medical Sciences, King Saud University, Riyadh, Kingdom of Saudi Arabia; 4) Department of Pathology, King Saud University, Riyadh, Kingdom of Saudi Arabia.

Primary ciliary dyskinesia (PCD, OMIM: #244400) is an autosomal recessive disorder characterised by impaired ciliary function and/or structure which leads to subsequent clinical phenotypes such as chronic sino-pulmonary disease. PCD is also a genetically heterogeneous disorder with many single gene mutations (with >20 genes identified so far) leading to similar clinical phenotypes. Here we present a novel PCD causal gene, coiled coil domain containing 151 (*CCDC151*), which has been shown to be essential in motile cilia of many animals and vertebrates but its nullizygous effects in humans was not observed previously. We observed a novel nonsense mutation in the *CCDC151* gene (c.925C>A, p.E309X) in a clinically diagnosed PCD patient from a consanguineous family of Arabic ancestry. We then directly assayed the variant using PCR-RFLP in 238 randomly selected individuals to find the prevalence of the variant in the local population. The variant was absent in all individuals indicating that the mutation is not a founder mutation and may have occurred relatively recently. *CCDC151* will be an addition to the already identified PCD causal genes (with six of them being coiled-coil domain containing genes: *CCDC39*, *CCDC40*, *CCDC65*, *CCDC103*, *CCDC114* and *CCDC164*) and facilitate our understanding of this complex disorder. Our finding also shows that given prior knowledge from model organisms, even a single whole-exome sequence (with high read depth, e.g. 50X) can be sufficient to discover a novel causal gene (i.e. once a strong candidate is pinpointed, the variant status can be confirmed in the other family members by amplifying the region of interest).

2933M

Combined exome and targeted gene NGS gene panel identifies mutations in *CCDC151* as a cause of Primary Ciliary Dyskinesia. *A. Onoufriadis[1,2,17], R. Hjeij[3,17], CM. Watson[4,5,17], N. Klena[6,17], M. Kurkowiak[3,7,8], NT. Loges[3], GW. Dougherty[3], CP. Diggle[5], G. Gabriel[6], KL. Lemke[6], Y. Li[6], P. Pennekamp[3], T. Menchen[3], JK. Martin[9], C. Werner[3], T. Burgoyne[10], C. Westermann[11], A. Rutman[12], IM. Carr[5], C. O'Callaghan[12,13], E. Moya[14], EMK. Chung[15], . UK10K[16], E. Sheridan[5], KG. Nielsen[9], CW. Lo[6], H. Omran[3], HM. Mitchison[2].* 1) Medical and Molecular Genetics, King's College London, London, United Kingdom; 2) Genetics and Genomic Medicine Programme, University College London (UCL) Institute of Child Health, London WC1N 1EH, UK; 3) Department of General Pediatrics, University Children's Hospital Muenster, 48149 Muenster, Germany; 4) Yorkshire Regional Genetics Service, St. James's University Hospital, Leeds, LS9 7TF, UK; 5) Leeds Institute of Biomedical and Clinical Sciences, University of Leeds, St. James's University Hospital Leeds LS9 7TF, UK; 6) Department of Developmental Biology, University of Pittsburgh, Pittsburgh, PA 15201, USA; 7) Department of Molecular and Clinical Genetics, Institute of Human Genetics, Polish Academy of Sciences, Strzeszynska 32, 60-479 Poznan, Poland; 8) International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland; 9) Danish PCD Centre and Pediatrics Pulmonary Service, Department of Pediatrics and Adolescent Medicine, Copenhagen University Hospital, Rigshospitalet, Denmark; 10) UCL Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL, UK; 11) Gerhard-Domagk-Institut for Pathology, University Children's Hospital Muenster, 48149 Muenster, Germany; 12) Centre for PCD Diagnosis and Research, Department of Infection, Immunity and Inflammation, RKCSB, University of Leicester, Leicester, LE2 7LX, UK; 13) Respiratory, Critical Care & Anaesthesia, Institute of Child Health, University College London & Great Ormond Street Children's Hospital, 30 Guilford Street, London, WC1N 1EH, UK; 14) Bradford Royal Infirmary, Bradford, West Yorkshire BD9 6R, UK; 15) General and Adolescent Paediatrics Section, Population, Policy and Practice Programme, University College London (UCL) Institute of Child Health, London WC1N 1EH, UK; 16) uk10k.org; 17) These authors contributed equally to the work.

Primary ciliary dyskinesia (PCD) is a heterogenous genetic disorder of ciliary/flagellar dysmotility characterized by chronic upper and lower respiratory infections and defects in laterality. The disease is recessively inherited and affects one per 15,000-30,000 births. So far, mutations in 27 genes have been identified to cause PCD including *DNAH5, DNAH11, DNAI1, DNAI2* and *DNAL1* which encode subunits of the axonemal outer dynein arm (ODA) components, *CCDC114* which encodes an ODA docking complex component and *ARMC4* which is essential for proper targeting and anchoring of ODAs. Resequencing of a panel of candidate ciliopathy genes was applied to an affected Bedouin-Arabic individual from a consanguineous marriage. Therefore, we focused on homozygous non-synonymous or splice-site substitutions or indels, that were novel or present in the 1000 Genomes Project with a frequency <0.01. This analysis revealed a homozygous protein truncating variant in *CCDC151* (c.925G>T; p.Glu308*). In parallel, exome sequencing combined with autozygosity mapping was applied to an affected offspring from a UK-based consanguineous Pakistani-origin family. This strategy highlighted a large autozygous region in chromosome 19p13 harbouring an additional homozygous protein-truncating variant in *CCDC151* (c.1256C>T; pSer419*). Sanger sequencing of the *CCDC151* coding exons and their associated splice-sites in a cohort of 150 cases affected with PCD resulting from ODA defects identified an individual carrying the c.925G>T nonsense variant. Segregation analysis of the c.925G>T and c.1256C>T substitutions in all available members of the pedigrees confirmed recessive inheritance of both variants. Transmission electron microscopy of respiratory cilia cross-sections from individuals carrying *CCDC151* mutations showed loss of ODA. Immunofluorescence analysis showed that *CCDC151* encodes an axonemal coiled coil protein, mutations in which abolish assembly of CCDC151 into respiratory cilia, and furthermore cause a failure in axonemal assembly of the ODA component DNAH5 and ODA-DC associated proteins CCDC114 and ARMC4. In summary, these data suggest that *CCDC151* mutations cause PCD by disruption of the ODA docking complex formation.

2934T

Defective core protein IFT81 as a rare cause of a ciliopathy with neurological involvement. *I. Perrault[1, 2], J. Halbritter[3], J. Porath[3], X. Gérard[1,2], H. Yung Gee[3], D. Braun[3], H. Fathy[4], S. Thomas[7], S. Saunier[5], V. Cormier-Daire[6], T. Attié- Bitach[7], N. Boddaert[8], M. Taschner[9], M. Schueler[3], E. Lorentzen[9], R. Lifton[10], E. Otto[11], P. Bastin[12], J. Kaplan[1, 2], F. Hildebrandt[3, 13], J-M. Rozet[1, 2].* 1) Genetics in ophthalmology, IMAGINE, INSERM UMR 1163, PARIS, France; 2) Paris Descartes - Sorbonne Paris Cité University, Imagine Institute, 75015 Paris France; 3) Division of Nephrology, Department of Medicine, Boston Children's Hospital, Harvard MedicalSchool, Boston, MA 02115, USA; 4) Pediatric Nephrology Unit, University of Alexandria, Alexandria, Egypt; 5) INSERM UMR 1163, Molecular bases of hereditary kidney diseases: nephronopthisis and hypodysplasia, 75015 Paris, France; 6) INSERM UMR 1163, Molecular and Physiopathological bases of osteochondrodysplasia, 75015 Paris, France; 7) INSERM UMR 1163, Embryology and genetics of human malformation, 75015 Paris, France; 8) Department of Pediatric Radiology, Hôpital Necker-Enfants Malades, APHP, Paris, DescartesUniversity, 75015 Paris, France; 9) Department of Structural Cell Biology, Max Planck Institute of Biochemistry, Martinsried, Germany; 10) Department of Genetics and Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT 06510, USA; 11) Departments of Pediatrics, University of Michigan, Ann Arbor, MI 48109, USA; 12) Trypanosome Cell Biology Unit, Institut Pasteur and CNRS, URA 2581, 75015, Paris, France; 13) Howard Hughes Medical Institute, Chevy Chase, MD, USA.

Bidirectional intraflagellar transport (IFT) consists of two major protein complexes, IFT-A and IFT-B. In contrast to IFT-B, all components of IFT-A have recently been linked to human ciliopathies when defective. By screening of 1,056 individuals with nephronophthisis-related ciliopathies for mutations in all genes encoding IFT-B, we identified one family with a homozygous obligatory splice site mutation in IFT81. Independently, by applying targeted ciliome resequencing of 572 individuals with early-onset severe retinal dystrophies or multisystemic ciliopathies, we found another family with a homozygous frameshift resulting in a loss-of-stop mutation with extension of the deduced protein by 10 amino acids. Both affected individuals presented with typical ciliopathy features, such as nephronophthisis, retinal dystrophy, cerebellar atrophy, and mild skeletal malformations. Cultured fibroblasts from one of the affected individuals showed no difference to control fibroblasts in regards of IFT81 localization to primary cilia or binding to IFT25. However, there was a statistically significant decrease in ciliated cell abundance compared to control fibroblasts. In addition, expression of GLI1 and PTCH1 appeared markedly increased, suggesting deranged sonic hedgehog signaling. Unlike the peripheral IFT-B members IFT172, and IFT80, IFT81 is only the second core protein found defective in humans. Our data further elucidate the role of IFT-B in human disease and show that defects in the IFT-B core are an exceedingly rare finding, probably due to its indispensable role for ciliary assembly in development.

2935S

Genetic study of patients with Joubert Syndrome and Related Disorders. *T. Vilboux[1], D. Yildirimli[1], M. Vemulapalli[2], A.R. Cullinane[1], M.C.V Malicdan[1], Y.M. Chang[1], J. Bryant[1], C. Sinclair[1], R. Fischer[1], B. Maskeri[2], A.C. Young[2], J. Mullikin[2], M. Huizing[1], W.A. Gahl[1,3,4], M. Gunay-Aygun[1,3], NISC Comparative Sequencing Program, NHGRI, NIH.* 1) MGB, NHGRI, NIH, Bethesda, MD; 2) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Rockville MD 20852, USA; 3) Office of the Clinical Director, NHGRI, National Institutes of Health, Bethesda, MD USA; 4) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, Maryland, USA.

Joubert Syndrome and Related Disorders (JSRD) is a clinically and genetically heterogenous group of ciliopathies defined based on a distinctive brain malformation (molar tooth sign) on brain imaging. Most JSRD patients display hypotonia, developmental delay, abnormal eye movements, and an abnormal respiratory pattern in infancy. Variable features include fibrocystic kidney disease, congenital hepatic fibrosis, retinal degeneration, retinal colobomas, and polydactyly. To date, twenty-two genes have been identified; they account for approximately 50% of JSRD. Over the past 10 years, we have evaluated over 100 JSRD patients at the NIH Clinical Center, under the NHGRI protocol "Clinical and Molecular Investigations into Ciliopathies" (www.clinicaltrials.gov, trial NCT00068224). The enrollment criterion was a clinical diagnosis of JSRD made based on brain MRI report describing the pathognomanic finding of "molar tooth sign" on brain MRI. In about fifty patients in whom we found no mutation in known JSRD genes using standard methods, we performed whole exome sequencing; to facilitate the analysis, family members were included. To date, the analysis has revealed that 2/3 of these patients have mutations in known ciliopathy genes. In the remaining 1/3, we have identified 4 novel genes; only 16% of families now lack a molecular basis for their JSRD. Identification of the underlying genetic causes of JSRD patients and further description of the full clinical spectrum of the related phenotypes will provide the groundwork for more focused studies. We are now assessing cilia integrity and function at the cellular level as well as at the level of the whole organism, using animal models such as zebrafish and C. *elegans*. This study will help to define genotype-phenotype correlations, improve diagnosis and prognosis, and stimulate development of improved supportive treatments as well as specific, novel and targeted therapies.

2936M

The Role of a Rare Variants in Genetic Predisposition to Statin-Induced Myopathy. *V. Stranecky[1], K. Hodanova[1], H. Hartmannova[1], L. Piherova[1], A. Pristoupilova[1], M. Neroldova[2], M. Vrablik[3], M. Jirsa[2], S. Kmoch[1].* 1) First Medical Faculty, Charles University of Prague, Prague, Czech Republic; 2) Laboratory of Experimental Hepatology, Institute for Clinical and Experimental Medicine, Prague, Czech Republic; 3) Third Medical Department, First Faculty of Medicine Charles University and General Faculty Hospital in Prague, Czech Republic.

Statin-induced myopathy is an important cause of statin intolerance and the most common cause of statin discontinuation. Statins exhibit a class-wide side effect of muscle toxicity and weakness, which has led regulators to impose both dosage limitations and a recall. This study focuses on the genetic factors associated with increased statin muscle concentrations, especially the genes encoding an influx transporter (SLCO1B1), efflux transporters (ABCB1 and ABCG2) and others, still unknown genes. We have analyzed gDNA of extreme cases of statin-induced myopathy. Patients had to fulfill the criteria: age over 18, clinically and biochemically determined myopathy, elevation of creatinin kinase and histological confirmation of muscle disease, no intake of inhibitors of OATP1B1 membrane transporters. We have performed exom sequencing, whole genome genotyping, analyzed CNV's and searched for large deletions. Here we present preliminary results of 11 patients. We found LOH (loss of heterozygosity) in *SLO1B* loci in 4 patients and a novel frameshift mutation c.209_210insATTGA (p.D70fs) in *SLCO1B3* in one individual. Heterozygous variant c.521T>C (p.V174A) dbSNP rs4149056 in *SLCO1B1* associated with station myopathy (Link et al, NEJM 2008), was found in 5 patients. In two cases this variation was combined with LOH in the *SLCO1B3* region. The incidence of this variant in our study is 5/22 alleles, which exceeds population frequency (MAF in 1000Genomes 0.123). No potential pathogenic variants in genes *SLCO1B1* and *SLCO1B3* was found in 3 of 11 patients. Furthermore we found null mutations in 823 genes. The most relevant for the study are *CYP3A43* (1 patient), *CYP3A5* (1 patient) a *AMPD1* (2 patients). Frequent occurrence of null mutations was found in *CYP* (*CYP2F1* 4x, *CYP4B1* 3x) gene family and transporters (*SLC7A13* 3x, *ABCB10* 2x). We found also null mutations in *APOB1* (1 patient) and *LPL* (1 patient) which contribute to hyperlipoproteinemia. Theoretical consequence of our project lies in detailed knowledge of genetic predisposition to toxicity of statins. Our results will help to predict pharmacodynamics and kinetics of statins and other drugs. Individual approach based on the assessment of the patient's genotype will improve the safety of the therapy.

2937T

Targeted resequencing identifies *PTCH1* as a major contributor to ocular developmental anomalies. *N. Chassaing[1,2,21], E.E. Davis[3,4,21], A. Causse[2,5], V. David[6,7], A. Desmaison[2], A.R. Niederriter[3], S. Lamarre[8,9], C. Vincent-Delorme[10], L. Pasquier[11], C. Coubes[12], D. Lacombe[13,14], M. Rossi[15], J-L. Dufier[16], H. Dollfus[17], J. Kaplan[18], N. Katsanis[3,4], H.C. Etchevers[2,19], S. Faguer[20], P. Calvas[1,2].* 1) Department of Medical Genetics, CHU Toulouse, France; 2) EA-4555, Toulouse III University, Toulouse, France; 3) Center for Human Disease Modeling, Duke University Medical Center, Durham, North Carolina, USA; 4) Department of Pediatrics and Department of Cell Biology, Duke University Medical Center, Durham, North Carolina, USA; 5) Service d'Ophtalmologie, Hôpital Purpan, CHU Toulouse, 31059 Toulouse, France; 6) Institut de Génétique et Développement, CNRS UMR6290, Université de Rennes 1, IFR140 GFAS, Faculté de Médecine, 35043 Rennes, France; 7) Laboratoire de Génétique Moléculaire, CHU Pontchaillou, 35043 Rennes Cedex, France; 8) INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, Toulouse, France; 9) Plateforme Biopuces de la Génopole de Toulouse Midi Pyrénées, INSA/DGBA 135, Toulouse, France; 10) Service de Génétique Médicale, Hôpital Jeanne de Flandre, 59037 Lille, France; 11) Service de Génétique clinique, Hôpital Sud, 35200 Rennes, France; 12) Service de Génétique Médicale, Hôpital Arnaud de Villeneuve, 34295 Montpellier, France; 13) Service de Génétique Médicale, Hôpital Pellegrin, 33076 Bordeaux Cedex, France; 14) Université Bordeaux Segalen, Laboratoire MRGM, 33076 Bordeaux, France; 15) Service de Cytogénétique Constitutionnelle, Hospices Civils de Lyon, Groupement Hospitalier Est, 69677 Bron, France; 16) Service d'Ophtalmologie, Hôpital Necker enfants Malades, 75015 Paris,; 17) Service de Génétique Médicale, Hôpitaux Universitaires de Strasbourg, 67091 Strasbourg, France; 18) INSERM U781 & Department of Genetics, Paris Descartes University, 75015 Paris, France; 19) INSERM, U910, Université de la Méditerranée Faculté de Médecine, 13385, Marseille, France; 20) INSERM unit 1048, I2MC, Team 12, 31432 Toulouse, France; 21) These authors contributed equally to this work.

Ocular developmental anomalies (ODA) such as Anophthalmia/Microphthalmia (AM) or anterior segment dysgenesis (ASD) have an estimated combined incidence of 3.7 in 10,000 births. Mutations in *SOX2* are the most frequent contributors to severe ODA, yet account for a minority of the genetic drivers. To identify novel ODA loci, we conducted targeted high-throughput sequencing of 407 candidate genes in an initial cohort of 22 sporadic ODA patients. Patched 1 (*PTCH1*), an inhibitor of sonic hedgehog (SHH) signaling, harbored an enrichment of rare heterozygous variants in comparison to either controls, or to the other candidate genes (four missense and one frameshift), and targeted resequencing of *PTCH1* in a second cohort of 48 ODA patients identified two additional rare nonsynonymous changes. Consistent with a role of *PTCH1* in ODA, functional analysis in a zebrafish *in vivo* complementation model showed that all six patient missense mutations affect SHH signaling. These results underline the SHH signaling pathway role during human ocular development, and demonstrate that PTCH1 contributes mutations to as much as 10% of ODA.

2938S

Deciphering the endothelin pathway in auriculocondylar syndrome and isolated question mark ears. *C. Gordon[1], F. Petit[2], P. Kroisel[3], L. Jakobsen[4], R. Zechi-Ceide[5], M. Oufadem[1], C. Bole-Feysot[1], P. Nitschke[1], A. Munnich[1], S. Lyonnet[1], M. Holder-Espinasse[6], J. Amiel[1].* 1) Institut Imagine, INSERM U-1163, Paris, France; 2) Hopital Jeanne de Flandre, Centre Hospitalier Regional Universitaire de Lille, Lille, France; 3) Institute of Human Genetics, Medical University of Graz, Graz, Austria; 4) Copenhagen University Hospital, Herlev, Denmark; 5) Hospital for Rehabilitation of Craniofacial Anomalies, University of Sao Paulo, Bauru, Brazil; 6) Department of Clinical Genetics, Guy's Hospital, London, UK.

Auriculocondylar syndrome (ACS) is a rare craniofacial disorder involving micrognathia, condyle hypoplasia and question mark ear (QME). QME, which involves a defect in the fusion of the lobe and helix, can occur as an isolated anomaly. Recently, mutations in PLCB4 (encoding a phospholipase C enzyme) and GNAI3 (encoding a G alpha protein) were identified in ACS. Both genes are predicted to function in the endothelin 1 (EDN1)-endothelin receptor type A (EDNRA) signalling pathway during the development of the pharyngeal arches. Following exclusion of PLCB4 and GNAI3 in a series of ACS and QME patients, we performed exome sequencing in four unsolved families. We identified a mutation in EDN1 in three cases, and a mutation in a gene known to regulate intracellular calcium release downstream of phospholipase C activity, in the fourth. Another EDN1 mutation was subsequently identified by direct sequencing. Two of the four EDN1 mutations occurred in patients affected with ACS, born to consanguineous, healthy parents; these patients harboured homozygous missense mutations in EDN1, each predicted to interfere with enzymatic cleavage of the EDN1 proprotein. The other two EDN1 mutations occurred in patients with dominantly-inherited isolated QME; they harboured heterozygous EDN1 mutations - a premature stop in one case and a missense mutation affecting a highly conserved residue of the mature EDN1 peptide in the other. The nature of the mutations and the different modes of inheritance suggest that heterozygous loss of function mutations in EDN1 cause isolated QME and that homozygous hypomorphic mutations cause ACS. These are the first reported mutations of EDN1 in humans. Our studies have also identified a mutation in a regulator of intracellular calcium release not previously implicated in the EDN1 pathway or in craniofacial development. Our results suggest that ACS and QME are a phenotypic continuum resulting from impaired EDN1-EDNRA signaling.

2939M

Mutations in *COG2* Encoding a Subunit of the Conserved Oligomeric Golgi Complex Cause a Congenital Disorder of Glycosylation. *H. Kodera[1], N. Ando[2], I. Yuasa[3], Y. Wada[4], Y. Tsurusaki[1], M. Nakashima[1], N. Miyake[1], S. Saitoh[2], N. Matsumoto[1], H. Saitsu[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Kanagawa, Japan; 2) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Aichi, Japan; 3) Division of Legal Medicine, Faculty of Medicine, Tottori University, Tottori, Japan; 4) Department of Molecular Medicine, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan.

The conserved oligomeric Golgi (COG) complex is involved in intra-Golgi retrograde trafficking, and mutations in six of its eight subunits have been reported in congenital disorders of glycosylation (CDG). Here we report a patient showing severe acquired microcephaly, psychomotor retardation, seizures, liver dysfunction, hypocupremia, and hypoceruloplasminemia. Analysis of his serum glycoproteins revealed defects in both sialylation and galactosylation of glycan termini. Trio-based whole-exome sequencing identified two heterozygous mutations in *COG2*: a *de novo* frameshift mutation [c.701dup (p.Tyr234*)] and a missense mutation [c.1900T>G (p.Trp634Gly)]. Sequencing of cloned reverse-transcription polymerase chain reaction products revealed that both mutations were located on separate alleles, as expected, and that the mutant transcript harboring the frameshift mutation underwent degradation. The c.1900T>G (p.Trp634Gly) mutation is located in a domain highly conserved among vertebrates and was absent from both the public database and our control exomes. Protein expression of COG2, along with COG3 and COG4, was decreased in fibroblasts from the patient. Our data strongly suggest that these compound heterozygous mutations in *COG2* are causative of CDG.

2940T

De novo *SOX11* mutations cause Coffin-Siris syndrome. *N. Matsumoto[1], Y. Tsurusaki[1], E. Koshimizu[1], H. Ohashi[2], S. Phadke[3], N. Miyake[1].* 1) Yokohama City Univ Grad Sch Med, Yokohama, Japan; 2) Division of Medical Genetics, Saitama Children's Medical Center, Iwatsuki, Japan; 3) Department of Medical Genetics, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Uttar Pradesh, India.

Coffin-Siris syndrome (CSS) is a congenital disorder characterized by growth deficiency, intellectual disability, microcephaly, characteristic facial features and hypoplastic nails of the fifth fingers and/or toes. We previously identified mutations in five genes encoding subunits of the BAF complex, in 55% of CSS patients. To further address genetic causation,Here we performed whole exome sequencing in additional CSS patients, identifying de novo SOX11 mutations in two CSS patients with mild intellectual disabilitya mild end of CSS phenotype. sox11a/b knockdown in zebrafish causes brain abnormalities, potentially explaining the brain phenotype of CSS. SOX11 is the downstream transcriptional factor of the PAX6-BAF complex, highlighting the importance of the BAF complex and SOX11 transcriptional network in brain development. Aknowledgment: Drs. Kou I, Shiina M, Suzuki T, Okamoto N, Imamura S, Ymashita M, Watanabe S, Yoshiura K-i, Kodera H, Miyatake S, Nakashima M, Saitsu H, Ogata K are highly appreciated for their contribution to this work.

2941S

Copy Number Variations detection for Congenital Absence of bilateral ACL and PCL ligaments. *Y. Liu[1], R. Golhar[1], ME. March[1], Y. Guo[1], Y. Li[1], ME. March[1], J. Li[2], J. Zhang[2], X. Xu[2], MA. Deardorff[1], B. Keating[1], H. Hakonarson[1].* 1) Center for Applied Genomics, The children's hospital of philadelphia, Philadelphia, PA; 2) BGI-Shenzhen, Shenzhen, China 518083.

Absence of the anterior (ACL) or posterior cruciate ligament (PCL) are rare congenital malformations that results in knee joint instability with a prevalence of 0.017 per 1,000 live births and can be associated with other lower limb abnormalities such as ACL agnesia and absence of menisci. While few cases of absence of ACL/PCL are reported in literature, a number of large familial case series of related conditions such as ACL agnesia suggest a potential underlying Mendelian etiology. We report a Copy Number Variation (CNV) deletion of exons in the gene CEP57L identified in a mother and daughter of the same family based on exome sequencing. CEP57L1 (centrosomal protein 57kDa-like 1) is a protein-coding gene that annotated as microtubule binding and molecular_function based on Gene Ontology (GO). Corresponding protein centrosomal protein may be required for microtubule attachment to centrosomes indicates the gene involved with mitosis processes. The deletion is validated through qPCR. The specificity of the CNV deletion has been check through ~700 exome sequencing data including 274 healthy individuals, and Database of Genomic Variations (DGV), a database contains large numbers of annotated CNVs from previous scientific reports. The CNV deletion we reported is a de novo variation that associated with ACL/PCL.

2942M

New standards in OMIM for gene-phenotype relationships in the era of whole genome/exome sequencing and a new way to follow disease gene discovery through MIMmatch. *J.S. Amberger, C.A. Bocchini, N. Sobreira, A.F. Scott, A. Hamosh.* Institute of Genetic Medicine, Johns Hopkins Univ Sch Med, Baltimore, MD, USA.

For almost 50 years, OMIM has been cataloging phenotype-gene relationships. Over time, variants have been ascertained by a variety of technologies and the certainty of these relationships is highly variable. The use of whole-genome/exome sequencing has enabled rapid growth in the publication of causative variants, but still does not confirm the certainty of the relationships. Effective January 2013, OMIM has adopted the following criteria to establish a gene-phenotype relationship: (1) the existence of multiple, unrelated individuals with pathogenic variants in the same gene; (2) the variants segregate with the phenotype in multiplex families; and/or (3) the variants occur de novo in a statistically significant number of individuals. Functional data and/or animal models support the causality but are not required. A "qualified" gene-phenotype relationship is established based on the following: (1) only one multiplex family is reported to have variants in a single gene and the variants segregate with the phenotype in the family, and (2) there is supportive functional data such as in vitro enzyme activity, a comparable phenotype in a model organism experiment, or an animal model; in this case the gene-phenotype relationship is qualified by noting that the variant has been identified in only "1 family". In rare instances, a similar gene-phenotype relationship may be established on the basis of a single patient if there is robust supporting phenotype and functional data. The morbid map listing of a qualified phenotype is preceded by a "?". When other information is entered into OMIM that substantiates the phenotype-gene relationship, the qualifier is removed. No relationship is established if a report identifies only 1 patient or family and provides no supporting functional data; however, an allelic variant may be created as a "variant of unknown significance" with a discussion of a possible relationship with the phenotype. The same criteria is being applied retrospectively to all existing OMIM gene-phenotype relationships that are based on a single allelic variant. To help users follow OMIM entries, we have implemented a new service, MIMmatch. MIMmatch members can designate entries or Phenotypic Series to follow and receive email alerts when these are updated; find other researchers who share their interest in certain entries; and/or receive a daily update on new gene-phenotype relationships established in OMIM.

2943T

Using an Augmented Exome to Improve Diagnostic Yield: Case Studies in Retinal Disorders. *S. Garcia[1], J. Tirch[1], M.J. Clark[1], S. Strom[2], A. Martinez[3], G. Chandratillake[1], J. Harris[1], A. Patwardhan[1], S. Chervitz[1], M. Li[1], M. Pratt[1], G. Bartha[1], S. Luo[1], R. Chen[1], J. West[1], M.B. Gorin[3].* 1) Personalis, Inc., Menlo Park, CA; 2) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA; 3) Department of Ophthalmology and Jules Stein Eye Institute, UCLA, Los Angeles, CA.

Identifying the genetic etiology for retinal disorders is challenging because of allelic, phenotypic, and locus heterogeneity, as well as environmental toxicities resulting in phenocopies. Patients often endure long diagnostic odysseys involving many single gene and/or gene panel tests, with a genetic etiology remaining undetermined in a large percentage of cases (30-50%). Whole exome sequencing (WES) is a highly appealing alternative to panels which require frequent revision as new causative genes are discovered; however, incomplete coverage of relevant genes means standard WES is also non-ideal. To address these limitations, we developed an augmented exome (ACE Exome), which improves sensitivity to detect variants by enhancing coverage over genes of biomedical relevance. We conducted ACE Exome sequencing for members of eleven families with undiagnosed retinal disorders and used a novel automated system to rank variants by integrating family history and phenotypic information with the exome data. For ten of these families, we successfully identified the likely genetic basis of their retinal disorder. Several of these diagnoses would have been either missed completely or timely and expensive to pick-up via sequential gene testing protocols due to the involvement of genes that are thought to cause disease only rarely, e.g. variants in CRX, which is associated with only 1% of cases of retinitis pigmentosa (RP), were detected in three of our families. Three diagnoses involved genes not present on gene panels available in the US at the time of testing: one family has a novel homozygous variant in LEPREL1, a gene that has only been associated with retinal disorders in a single family in the literature; the second family has a variant in a new gene, previously considered a candidate gene for RP; and a third family was found with variation in NMNAT1, a recently described gene added to panels only very recently. Finally, one family has an X-linked genetic etiology that would have been missed by standard exome sequencing due to poor gene coverage of RPGR. These successes demonstrate the efficacy of enhanced exome sequencing to diagnose the genetic cause of retinal disorders.

2944S

Identifying genetic determinants of poor cochlear implantation outcomes using massively parallel DNA sequencing. *Y.H. Lin[1,2], C.C. Wu[2,3], T.C. Liu[2], W.S. Yang[3,4,5,6], C.J. Hsu[2,7], P.L Chen[1,3,5,6], C.M. Wu[8].* 1) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 2) Department of Otolaryngology, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 4) Graduate Institute of Clinical Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 5) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 6) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan; 7) Department of Otolaryngology, Taichung Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation; 8) Department of Otolaryngology, Chang Gung Memorial Hospital, Chang Gung University, Taoyuan, Taiwan.

Cochlear implantation, although extremely expensive, is currently the treatment of choice for children with severe to profound hearing impairment. However, the outcomes with cochlear implants (CIs) vary significantly among recipients. The purpose of the present study is to identify the genetic determinants of poor CI outcomes. Twelve children with poor CI outcomes (the "cases") and 30 "matched controls" with good CI outcomes were subjected to comprehensive genetic analyses using massively parallel sequencing, which targeted 129 known deafness genes. Audiological features, imaging findings, and auditory/speech performance with CIs were then correlated to the genetic diagnoses. We identified genetic mutations in 7 (58%) of the 12 cases; 4 cases had bi-allelic *PCDH15* mutations and 3 cases had bi-allelic *DFNB59* mutations. Mutations in the *WFS1*, *GJB3*, *ESRRB*, *LRTOMT*, *MYO3A*, and *POU3F4* genes were detected in 7 (23%) of the 30 matched controls. The allele frequencies of *PCDH15* and *DFNB59* mutations were significantly higher in the cases than in the matched controls (both $P < 0.001$). In the 7 CI recipients with *DFNB59* or *PCDH15* mutations, otoacoustic emissions were absent in both ears, and imaging findings were normal in all 7 implanted ears. *DFNB59* or *PCDH15* mutations are major genetic determinants of poor CI performance, probably because of their role in the pathology of spiral ganglion neurons and/or brainstem auditory nuclei. Because children with *DFNB59* or *PCDH15* mutations show clinical features indistinguishable from those of other typical pediatric CI recipients, genetic examination is indicated in all CI candidates prior to operation.

2945M

Exome sequencing reveals *TPO* mutations in Pseudo-Pendred syndrome. *A. DENOMME-PICHON[1], E. COLIN[1,2], S. MARLIN[3], L. JONARD[3], S. ROULEAU[4], S. KURY[5], S. DUMONT[5], D. GOUDENEGE[1], R. COUTANT[4], D. BONNEAU[1,2].* 1) Department of Biochemistry and Genetics, CHU Angers, Angers, France; 2) UMR CNRS 6214 INSERM 1083, Angers, France; 3) Department of Genetics, CHU Necker-Enfants-Malades, Paris, France; 4) Department of Pediatrics, CHU Angers, Angers, France; 5) Department of Genetics, CHU Nantes, Nantes, France.

Pseudo-Pendred syndrome (PPDS) is defined by the association of sensorineural deafness, hypothyroidism due to iodide organification defect, absence of inner ear malformation and absence of mutation in *SLC26A4* (MIM 605646), the gene responsible for classical PDS (MIM 274600).

In order to determine the cause of PPDS, we performed whole exome sequencing (WES) in a family with two children affected with hypothyroidism, developmental delay, positive perchlorate test and absence of inner ear malformation on CT-scan. Parents were healthy and non-consanguineous and direct sequencing of *SLC26A4* was normal in both patients.

WES was performed in both patients and their father. Variants were ranked by segregation, allele frequency, protein change and degree of conservation to assess likelihood of causation. Both patients were found to be compound heterozygous for missense mutations (c.T1357G,p.Y453D and c.G699C,p.W233C) in *TPO* (MIM 606765) and the father was heterozygous for the Y453D mutation.

In silico prediction tools determined a deleterious mutational impact of both variants on protein function. Sanger sequencing confirmed the mutation segregation and found that the mother was heterozygous W233C. The latter mutation was novel and both were absent in the control population.

TPO encodes thyroperoxidase which catalyzes key reactions in thyroid hormone synthesis and mutations in *TPO* are responsible for thyroid dyshormonogenesis 2A. Mutations in *TPO* have been reported in 4 patients with hypothyroidism and deafness in a series of Israeli patients with iodide organification defect (Tenenbaum-Rakover, 2007) but their phenotype was not described as PPDS. These cases together with the present report suggest that mutations in *TPO* can be responsible for pseudo-PDS.

2946T

Alleles of the reported deafness genes are major contributors to the etiology of moderate to severe hearing loss in Pakistani population. *A. Imtiaz[1,2], R. Bashir[2], G. Mubtaja[2], A. Maqsood[2], I. Bukhari[2], A. Rehman[1], R.J. Morell[1], T.B. Friedman[1], S. Naz[2].* 1) NIDCD, National institute of Health, Bethesda, MD; 2) School of Biological Sciences, University of the Punjab, Lahore, Pakistan.

Most of the genes that cause recessively inherited deafness are associated with profound, prelingual hearing loss, which occurs approximately 1/1,000 live births. The genetic causes of moderate, progressive, post-lingual hearing loss are not well understood, despite the fact that this type of hearing loss is much more prevalent in the general population. We studied the molecular genetic bases of recessively inherited, stable or progressive, moderate to severe hearing loss in 44 consanguineous Pakistani families. We used a stepwise experimental approach, starting with homozygosity mapping at reported DFNB loci, followed by targeted sequencing of the known deafness genes, and finally whole-exome sequencing (WES) to identify mutations in novel genes associated with this phenotype. 11 known and 17 novel mutations in the known deafness genes cosegregate with moderate to severe hearing loss in 68% of these families. The majority of these reported mutant alleles are thought to cause only profound deafness. Nevertheless, our finding of their association with less severe phenotype may suggest the presence of environmental or genetic modifiers (suppressors) that reduce phenotypic severity. *SLC26A4, TMC1* and *OTOF* are three of the major contributors in this cohort. In 14 of the families we did not find causative variants in known deafness genes. Our findings indicate that the same genes and alleles known to cause profound deafness are also major contributors to the genetic basis of moderate to severe hearing loss.

2947S

Mutations in a metabolic kinase gene lead to autosomal dominant retinitis pigmentosa. *F. Wang[1,2], Y. Wang[3], B. Zhang[1,2], H. Li[4], L. Zhao[1,5], K. Wang[1,2], M. Xu[1,2], Y. Li[1,2], F. Wu[6], C. Wen[6], P. Bernstein[7], H. Wang[1,2], R. Sui[4], K. Zhang[6,8], R. Chen[1,2,5,9,10].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 3) Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 510060, China; 4) Department of Ophthalmology, Peking Union Medical College Hospital, Peking Union Medical College, Beijing 100730, China; 5) Structural and Computational Biology and Molecular Biophysics Graduate Program, Houston, TX 77030, USA; 6) Shiley Eye Center and Institute for Genomic Medicine, University of California San Diego, La Jolla, CA 92093, USA; 7) Department of Ophthalmology and Visual Sciences, Moran Eye Center, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA; 8) Veterans Administration Healthcare System, San Diego, CA 92161, USA; 9) The Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA; 10) Program in Developmental Biology, Baylor College of Medicine, Houston, TX 77030, USA.

Retinitis pigmentosa (RP) is a genetically heterogeneous disease with over 60 causative genes known to date. Nevertheless, approximately 40% of RP cases remain genetically unsolved, suggesting many novel disease-causing genes are yet to be identified. Here, we identified a glucose metabolism-related kinase gene as a novel causative gene for autosomal dominant RP (adRP). This is the first report that associates the glucose metabolic pathway with human retinal degenerative disease, suggesting a potential new disease mechanism. A large adRP family with negative result from known-retinal-disease-gene screening was recruited. Linkage analysis identified a minimal disease region of 8 Mb with a peak parametric LOD score of 3.500. Further whole-exome sequencing identified a heterozygous missense mutation that segregated with the disease phenotype in the family. Screening over 200 unsolved RP patients identified two unrelated simplex RP cases carrying an identical mutation which is only 11 amino acids away from the initial mutation identified in the adRP family, suggesting that both alleles may impact the same functional domain. Biochemical assays showed that the mutation does not affect kinase enzymatic activity or the protein stability, suggesting that the mutation may impact other uncharacterized function or result in a gain-of-function.

2948M

Expansion of the fibrosing poikiloderma phenotype caused by FAM111B to include cytopenia and pancreatic dysfunction. *A. Seo[1], T. Walsh[2], M.K. Lee[2], M-C. King[1,2], A. Shimamura[3].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Division of Medical Genetics, University of Washington, Seattle, WA; 3) Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA.

Poikiloderma, a dermatological disorder including telangiectasia, uneven pigmentation, and epidermal atrophy, can be inherited recessively as part of Rothmund-Thomson syndrome (RTS [MIM 268400]), Kindler Syndrome (MIM 173650), hereditary sclerosing poikiloderma of Weary [MIM 173700], and dyskeratosis congenita (DKCX [MIM 305000]). Recently, heterozygous missense mutations in FAM111B (MIM 615584) were shown to be responsible for a newly defined autosomal dominant condition including poikiloderma, tendon contraction, myopathy, pulmonary fibrosis, heat intolerance, hypohidrosis, eczematous lesions, and alopecia/sparse hair (POIKTMP [MIM 615704]). The critical FAM111B mutations all lie in the C-terminal trypsin-like cysteine/serine peptidase domain. Little is known about the function of FAM111B, and it remained unclear whether variants in other regions of FAM111B would result in the same or different clinical phenotype. We evaluated a large family of Northern European ancestry with a history of poikiloderma, eczema, heat intolerance, hypohidrosis, sparse hair, and elevated liver enzymes, apparently inherited in autosomal dominant fashion. The two youngest affected siblings also exhibited intermittent cytopenias, marrow hypocellularity, and exocrine pancreatic dysfunction, which are the characteristic features of Shwachman-Diamond Syndrome (SDS). A grandfather had a history of interstitial pulmonary disease, and several individuals had signs of tendon contracture. SBDS, the causative gene for SDS, was wildtype in the affected siblings. Whole exome sequencing of these two siblings, the affected father, and the unaffected mother revealed a heterozygous in-frame deletion FAM111B c.1261_1263delAAG (p.Lys421del) in the affected individuals. Genotyping the family revealed the mutation to be present in all 7 affected relatives genotyped and absent in all 3 unaffected relatives genotyped. The mutation does not appear on any public databases or among our in-house control exomes. In FAM111B, lysine at residue 421 lies outside the trypsin-like cysteine/serine peptidase domain, which may explain phenotypic differences between this family and those harboring mutations in the peptidase domain. Our study expands the phenotype caused by variants in FAM111B to include exocrine pancreatic dysfunction, cytopenias, and marrow failure. Our findings suggest FAM111B may play an important role in regulating fundamental processes of cellular maintenance, growth, and proliferation.

2949T

An atypical Bloom syndrome identified by exome sequencing in a ten year old girl with microcephaly. *C. Dupont[1], Y. Vial[2,3], I. Baatout[1], B. Diarra[1], S. Jacquier[3], B. Benzacken[1,3,4], A.C. Tabet[1], H. Cave[2,5], A. Baruchel[6], A. Verloes[7], M. Ouachee[6], S. Drunat[2,3].* 1) Unite de Cytogenetique, Departement de Genetique, APHP-Hopital Robert Debre, CHU Paris, Paris, France; 2) Unite de Biologie Moleculaire et Biochimie, Departement de Genetique, Hopital Robert Debre-AP-HP, CHU Paris, Paris, France; 3) Unite Inserm UMR1141- Universite Paris 7, France; 4) CHU Paris, Hopital Jean Verdier, Unite de Cytogenetique, Bondy, France, Universite Paris 13, Sorbonne Paris Cite; 5) Unite Inserm UMR1131- Universite Paris 7, France; 6) Service d'hemato-immunologie, Hopital Robert Debre-AP-HP, CHU Paris, Paris, France; 7) Unite de Genetique Clinique, Departement de Genetique, Hopital Robert Debre-AP-HP, CHU Paris, Paris, France.

Introduction: Bloom's syndrome (BSyn) is a very rare autosomal recessive chromosomal breakage disorder mainly characterized by pre-and post-natal growth retardation, cancer predisposition, facial sun-sensitivity and increased susceptibility to infections. BSyn is caused by mutations in the *RECQL3* gene, a DNA helicase, and is more common in the Ashkenazic jewish population where a founder mutation was identified. We report the unexpected diagnosis of Bloom syndrome in a Caribbean girl referred for microcephaly. Case report: The patient is a 10 year-old girl which presented with growth retardation of prenatal onset, severe microcephaly (-5SD) without intellectual disability, moderate thrombopenia, immune deficiency, and small café-au-lait spots on the trunk and limbs. Material and methods: Trio-based exome sequencing was performed in order to find new candidate genes for microcephaly in this patient. The data analysis searched for homozygous mutations found in the patient and not in her parents or compound heterozygous mutation inherited from one parent. Sister chromatid exchanges were BrDU induced on patient lymphocytes according to standard protocols and compared to an age-matched control to confirm the molecular diagnosis. Results: Exome sequencing identified compound heterozygous mutation in the *RECQL3* gene: NM_000057.2:c.c.2650_2651del (p.Lys884Alafs*5) and NM_000057.2:c.2809C>T (p.Gln937*) which both introduce a premature stop codon. The mutations and their de novo occurrence were confirmed by Sanger sequencing. BrDu-cytogenetic analysis showed a higher percentage of sister chromatid exchanges (five to eight times more exchanges) in patient's lymphocytes compared to the control. Discussion: In the reported case, the molecular diagnosis led to a retrospective clinical diagnosis in a patient with a phenotypic presentation not suggestive of Bloom syndrome. This result further illustrates that exome sequencing may be a comprehensive tool to detect genetic mutations, allowing for the identification of unexpected genetic causes in atypical clinical phenotypes.

2950S

Achieving genetic diagnosis in deaf patients with non-confirmative *GJB2* genotypes using massively parallel DNA sequencing. *C.J. Hsu[1,2], Y.H Lin[2,3], C.C. Wu[2,4], P.L. Chen[3,4,5].* 1) Department of Otolaryngology, Taichung Tzu Chi Hospital, Taichung, Taiwan; 2) Department of Otolaryngology, National Taiwan University Hospital, Taipei, Taiwan; 3) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 4) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 5) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan.

Recessive mutations in the *GJB2* gene are the most common genetic cause of idiopathic sensorineural hearing impairment (SNHI). However, in about 10% to 50% of patients, only 1 recessive *GJB2* mutation can be detected, constituting a diagnostic dilemma. In these patients, we hypothesize that there might be undetected mutations in: (1) the non-coding region of *GJB2* (such as untranslated exons, introns, promoter or enhancer) leading to compound heterozygous inheritance; (2) other gap junction genes leading to digenic inheritance; or (3) other modulating deafness genes. Accordingly, we designed a massively parallel sequencing (MPS) panel which target the whole *GJB2* gene and the exons of 129 deafness genes (including all gap junction genes), and applied this panel to 12 patients with only 1 recessive *GJB2* mutation detected by conventional Sanger sequencing of the *GJB2* coding region. Causative mutations in other deafness genes were identified in 4 patients, including *SLC26A4* mutations in 2 patients (genotypes: p.K369X/p.T410M and p.P8T/p.P8T, respectively), *KCNQ4* mutation in 1 patient (genotype: p.F182L/wt), and *MYO15A* mutations in 1 patient (genotype: p.W58C/c.8596-1G>C). A splicing site mutation (IVS1+1G>A) in *GJB2* was detected in 1 patient, contributing to compound heterozygosity of *GJB2* mutations. Of note, we also found ~67% mosaicism of *GJB2* c.235delC in another patient which was difficult to be identified by Sanger sequencing alone. In total, our MPS panel achieved genetic diagnosis in 6 (50%) of 12 patients with only 1 recessive *GJB2* mutation detected by conventional Sanger sequencing, including 4 with SNHI attributable to mutations in other deafness genes and 2 with SNHI attributable to compound heterozygosity or mosaicism of *GJB2* mutations. Our results demonstrated the utility of MPS in achieving genetic diagnosis in deaf patients with non-confirmative *GJB2* genotypes on conventional genetic examinations.

2951M

Megacystis microcolon intestinal hypoperistalsis syndrome and related phenotypes - five new cases with two supporting the autosomal recessive inheritance hypothesis. *C.A. Moreno[1], KC. Silveira[1], TF. Araujo[1], K. Metze[2], DR. Bertola[3], LC.Jr Llerena[4], DP. Cavalcanti[1].* 1) Departamento de Genética Médica, Faculdade de Ciências Médicas, Universidade Estadual de Campinas (Unicamp). Rua Tessália Vieira de Camargo, 126 - Cidade Universitária Zeferino Vaz. 13083-887 - Campinas, SP, Brasil; 2) Departamento de Anatomia-Patológica, Faculdade de Ciências Médicas, Universidade Estadual de Campinas (Unicamp). Rua Tessália Vieira de Camargo, 126 - Cidade Universitária Zeferino Vaz. 13083-887 - Campinas, SP, Brasil; 3) Unidade de Genética, Instituto da Criança, Faculdade de Medicina, Universidade de São Paulo. Av. Dr. Enéas Carvalho de Aguiar, 647, Cerqueira César. 05403900 - São Paulo, SP - Brasil; 4) Departamento de Genética Médica, Instituto Fernandes Figueira, Fundação Oswaldo Cruz. Avenida Rui Barbosa, 716, Flamengo. 22250020 - Rio de Janeiro, RJ - Brasil.

Megacystis microcolon intestinal hypoperistalsis syndrome (MMIHS) is a heterogenous condition characterized by marked dilatation of the bladder without structural obstruction, microcolon and decreased or absent intestinal hypoperistalsis. The adverse prognosis is due to complications related mainly to parenteral nutrition and renal failure. Although more than 200 described cases are sporadic, the ocurrence of MMIHS associated with consanguinity and recurrence in siblings with asymptomatic parents gave support for many years to the hypothesis of autosomal recessive (AR) inheritance. Recently, the identification of a de novo mutation in the ACTG2 gene in patients with MMIHS changed this concept and highlighted the possibility of genetic heterogeneity. The ACTG2 gene was also associated with phenotypes related to the spectrum of visceral smooth muscle diseases [familial visceral myopathy and chronic intestinal pseudo-obstruction (CIP)] with autosomal dominant pattern. The objective of this presentation is to show the preliminary investigation of five new Brazilian patients with MMIHS and related phenotypes. So far, we have identified the following patients: three with the classical MMIHS, being two sibs with consanguineous parents [F= 1/16], one presenting CIP, and the last one with megacystis without a defined etiology. The investigation strategy is firstly to sequence the ACTG2 by the Sanger method and, if negative to study these cases by the whole-exome sequencing. The ACTG2 sequencing of three patients is ongoing (MMIHS, CIP and megacystis without etiology). In conclusion, we present five new cases of the MMIHS and related phenotypes with a strong evidence of AR inheritance in one family (two sibs with F= 1/16). The up coming results should confirm the genetic heterogenity of the MMIHS spectrum and can contribute to improve the knowledge about its genetic basis.

2952T

Prevalence of *EIF2AK4* gene mutations in patients with a clinical diagnosis of pulmonary arterial hypertension. *K.L. Sumner[1], K. Damjanovich[1], B. Smith[2], I. Nakayama[3], Y. Ha[1], E. Paul[1], A. Morris[1], L.M. Brown[4,5], P. Bayrak-Toydemir[1,6], C.G. Elliott[4,5], D.H. Best[1,6].* 1) ARUP Institute for Clinical & Experimental Pathology, ARUP Laboratories, Salt Lake City, UT; 2) Division of Pulmonary and Critical Care Medicine, Maine Medical Center, Portland, ME; 3) Good Samaritan Hospital, Puyallup, WA; 4) Dept. of Medicine, Intermountain Medical Center and the University of Utah, Murray, UT; 5) Division of Pulmonary Medicine, University of Utah School of Medicine, Salt Lake City, UT; 6) Dept. of Pathology, University of Utah School of Medicine, Salt Lake City, UT.

Background: Pulmonary arterial hypertension (PAH) is a rare disorder that may be heritable (HPAH), idiopathic (IPAH), or associated with either drug-toxin exposures or other medical conditions. HPAH is caused by autosomal dominant (with reduced penetrance) mutations in *BMPR2*, and less frequently by mutations in *ACVRL1*, *ENG*, *CAV1*, or *KCNK3*. Recently, autosomal-recessive mutations in *EIF2AK4* have been associated with pulmonary capillary hemangiomatosis and pulmonary veno-occlusive disease. However, it is not clear whether *EIF2AK4* mutations also occur in patients diagnosed with PAH without pathologic examination of lung tissue.

Materials and Methods: Sixty patients with a clinical diagnosis of PAH (7 HPAH, 48 IPAH, 5 unknown) were evaluated for mutations in *BMPR2* by Sanger sequencing and deletion/duplication analysis. *BMPR2* mutation negative patients (n=47) were then tested for mutations in the other known PAH-associated genes (*ACVRL1*, *ENG*, *CAV1*, and *KCNK3*) by Sanger sequencing. The *ACVLR1* and *ENG* genes were also tested for large deletions and duplications. All remaining mutation negative patients (n=46) were then evaluated for mutations in *EIF2AK4* by Sanger sequencing.

Results: Gene mutations were identified in a total of 14 patients. As expected, mutations in *BMPR2* represented the bulk of mutation positive patients (n=13). Mutations were also identified in *ENG* (n=1) and variants of unknown significance were found in *ACVRL1* (n=1) and *KCNK3* (n=1). No mutations were identified in *CAV1* or *EIF2AK4*.

Conclusion: *EIF2AK4* mutations do not appear to be a common cause of disease in patients with a clinical diagnosis of PAH.

2953S

Comprehensive mutation analysis using Ion PGM in 95 patients with neonatal intrahepatic cholestasis. *T. Togawa[1], T. Sugiura[1], K. Ito[1], T. Endo[1], A. Kikuchi[2], N. Ichinoi[2], S. Kure[2], S. Saitoh[1].* 1) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Department of Pediatrics, Tohoku University School of Medicine, Sendai, Japan.

Next generation sequencing (NGS) technology has revolutionized genomic and genetic research. Neonatal intrahepatic cholestasis is a heterogeneous disorder and caused by mutations in a number of genes, making genetic diagnosis challenging. We developed a diagnostic custom panel of causative 19 genes for neonatal intrahepatic cholestasis using the Ion AmpliSeq™ (LifeTechnologies) and Ion Torrent PGM technology. gDNA samples were obtained from 95 patients divided into two groups. Group A; 50 patients with neonatal or infantile cholestasis, in whom conventional mutation screening was not successful. Group B; 45 patients suspected to have NICCD, and no variants in *SLC25A13* was found by a melting-curve analysis with adjacent hybridization probes (HybProbe, Roche) detecting 95% of mutations in Japanese patients with citrin deficiency. NGS workflow based on a custom AmpliSeq™ panel was designed for sequencing the 14 prevalent known causative genes and 5 unknown genes on Ion PGM™ Sequencer. A panel of genes included *JAG1*, *NOTCH2*, *SLC25A13*, *ATP8B1*, *ABCB11*, *ABCB4*, *AKR1D1*, *HSD3B7*, *CYP7B1*, *TJP2*, *BAAT*, *EPHX1*, *ABCC2*, *VPS33B*, *SLC10A1*, *LST1*, *ABCB1*, *SLCO1A2,* and *SLC4A2*. Number of amplicons, total targeted bases and exons were 551, 53.426kb, 355, respectively. Our AmpliSeq™ custom cholestasis panel allowed us to cover 98.4% of the targeted sequences. NGS was performed on Ion PGM™ with Ion 316™ or 318™ sequence chips. Average coverage of depth in the target region was 654, and 97.1% of sequenced region was covered over 100 coverage. Variant call and annotation were performed with CLC Genomics Workbench version 7.0 and IonReporter™ 4.0. Variants were filtered out by common dbSNP138 and in-house filtering data. Accordingly, we detected disease-causing mutations in 13 patients in group A; four *JAG1* mutations, 1 *NOTCH2*, 2 *ATP8B1*, 2 *ABCB11*, 3 *SLC25A13*, and 1 *ABCC2*. In group B, six patients were identified in 6 causative genes; one *JAG1* mutation, 1 *NOTCH2*, 1 *ABCB11*, 1 *SLC25A13*, 1 *ABCC2*, and 1 *TJP2*. This study shows clinical usefulness of comprehensive mutation analysis by NGS for neonatal intrahepatic cholestasis with unknown genetic etiology.

2954M

An update on the CMTX3 locus: using whole genome sequencing to search for the elusive mutation. *M.H. Brewer[1,2], J. Qi[1,3], R. Chaudhry[1,2], A. Drew[1], G. Nicholson[1,2,4], M. Kennerson[1,2,4].* 1) Northcott Neuroscience Laboratory, ANZAC Research Institute, Concord, NSW, Australia; 2) Sydney Medical School, University of Sydney, Camperdown, NSW, Australia; 3) Department of Pathology, University of Sydney, Camperdown, NSW, Australia; 4) Molecular Medicine Laboratory, Concord General Repatriation Hospital, Concord, NSW, Australia.

Charcot-Marie-Tooth (CMT) disease is the collective name given to a group of inherited peripheral neuropathies featuring both sensory and motor involvement. The disease is genetically heterogeneous, with over 1000 mutations identified in over 80 genes, and can be inherited in an autosomal dominant, autosomal recessive or X-linked fashion. Six X-linked CMT loci (CMTX1-6) have been reported, of which the gene mutations for four loci have been identified (GJB1, AIFM1, PRPS1 and PDK3). CMTX3 was first mapped to the q-arm of chromosome X in 1991 [1] and yet this locus remains genetically unsolved. The overarching aim of this project is to identify the genetic cause of CMTX3. Genetic linkage studies performed on two large families confirmed and refined the locus to a 5.7 Mb region at Xq26.3-q27.3 [2,3]. Interestingly, both families carry the same disease haplotype suggesting they share the same disease mutation inherited from a common founder [3]. Mutation scanning the coding exons, splice-sites and untranslated regions of all annotated genes within the disease locus did not reveal a causative mutation. Therefore we hypothesize that CMTX3 is caused by a mutation disrupting a non-coding regulatory element. To thoroughly investigate the CMTX3 locus for regulatory variants we have performed whole genome sequencing (WGS) on six individuals - two affected males and one related healthy male from each family. Analysis of the known CMT genes did not reveal any mutation segregating with the disease phenotype, confirming a known gene does not cause CMT in these two families. Under the assumption that these two families share a common founder disease mutation, WGS data was filtered for: a) novel variants; b) that map within the CMTX3 locus and; c) were found in all affected individuals across both families and absent from both controls. This identified 21 candidate variants (16 SNPs and five indels). Variants were either intronic or intergenic. Importantly, no novel coding variants were identified supporting our previous mutation scanning results. These variants will be assessed for pathogenicity by: a) testing for segregation with the disease phenotype in both CMTX3 families and; b) screening these variants in over 600 neurologically normal controls. References: 1. Ionasescu et al. (1991) Am J Hum Genet; 2. Huttner et al. (2006) Neurology; 3. Brewer et al. (2008) Neurogenetics.

2955T

PITUITARY HORMONE DEFICIENCY: HUNT FOR NOVEL CAUSATIVE GENES AND GENETIC CONTRIBUTIONS TO VARIABLE PENETRANCE AND EXPRESSIVITY. *Q. Fang[1], I.J.P. Arnhold[3], A.F. Benedetti[3], T. Brue[4], L.R.S. Carvalho[3], F. Castinetti[4], F.A. Correa[3], N. Foyouzi[2], M.M. Franca[3], J.Z. Li[1], Q. Ma[1], B.A. Ozel[1], B.B. Mendonça[3], M. Moreira[3], A.P. Otto[3], R. Reynaud[4], A. Sadeghi-Nejad[5], S.A. Camper[1].* 1) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 2) Department of Obstetrics and Gynecology, University of Michigan Medical School, Ann Arbor, MI; 3) University of São Paulo, Clinical Hospital, São Paulo, Brazil; 4) Aix-Marseille University, Center for Research in Neurobiology and Neurophysiology of Marseille (CRN2M) - CNRS UMR7286, Timone Hospital, Marseille, France; 5) Division of Pediatric Endocrinology, Tufts University School of Medicine, Boston, MA.

Pituitary hormone deficiency is genetically heterogeneous and fatal if hypoglycemia and adrenal insufficiency are untreated. The known causal genes explain a minority of the cases. Predicting patient outcomes is difficult because some genes exhibit dominant inheritance with incomplete penetrance, and most mutant genes have variable clinical presentations. Mutations in the transcription factors OTX2 and HESX1 cause defects in craniofacial development involving the eyes, pituitary gland, or both. To identify novel causal genes and to search for genes that influence penetrance we sequenced the exomes of 22 individuals with hypopituitarism. We found two cases of mutations in transcription factor HESX1. A familial case of pituitary aplasia with neonatal crisis is a novel compound heterozygote for HESX1, R159W/R160H, which disrupts the DNA sequence recognition helix. We found a previously reported p.I26T lesion that impairs HESX1 repressor function in a consanguineous case of multiple pituitary hormone deficiency. We discovered a homozygous, recurrent p.L144H mutation in the growth hormone releasing hormone receptor (GHRHR) gene in a consanguineous case of isolated growth hormone deficiency (IGHD). A proband with hypopituitarism, polydactyly, and normal eyes was ascertained in a large pedigree, has a novel heterozygous variant in OTX2 (p.H230L). This histidine is conserved across all vertebrate species, and the leucine variant acts as a dominant negative in both heterologous (293T) and homologous GnRH neuronal cell lines (GT1-7) transfected with either a consensus OTX2 binding site or a known target, the gonadotropin releasing hormone receptor promoter. The proband's mother is a p.H230L carrier with short stature, and several unaffected individuals in the pedigree are heterozygous for this variant. The presence of polydactyly in the father and the proband suggests the possibility of digenic disease. We found a few heterozygous variants shared by the father and the proband that are rare, likely damaging, and are associated with polydactyly or pituitary enriched expression. These genes are candidates for enhancing the effects of the OTX2 variant on hormone production and polydactyly. This exome sequencing pilot revealed that approximately 20% of probands have mutations in known genes, and nominates novel genes and pathways that affect normal pituitary function.

2956S

A Combined Exome sequencing and RNA-Seq Strategy Reveals a Novel Mutation in DOCK8 that Results in Immunodeficiency and Radiosensitivity. *S. Khan[1], M. Kuruvillas[1,2], B. Wakeland[1], C. Liang[1], K. Vishwanathan[1], R. Gatti[3], T. Torgersen[4], N. van Oers[1], E. Wakeland[1], M. Teresa de la Morena[1,2].* 1) UT Southwestern Medical Center, Dallas, TX; 2) Children's Medical Center, Dallas, TX; 3) University of California, Los Angeles, CA; 4) Seattle Children's Research Institute, Seattle, WA.

Primary immunodeficiency diseases (PIDs) are rare in the human population, with the causal genetic mutations difficult to identify. To date, more than 180 genes have been linked to diverse PIDs, with the vast majority due to monogenic defects. Recent strategies to identify the genetic causes of PIDs include exome and whole genome sequencing. We performed RNA sequencing and exome sequencing on RNA and DNA from two siblings with prominent manisfestations of warts and cryptosporidium sclerosing cholangitis, profound T cell lymphopenia and radiosensitivity. Comparative analyses including an unaffected sibling and parents using exome sequencing revealed a novel homozygous 2 bp insertion in the splice site region in intron (26-27) of dedicator of cytokinesis 8 (DOCK8) gene in the two affected siblings. RNA-SEQ analysis of their whole blood showed a significant reduction in mRNA levels of DOCK8 gene as well as a cluster of aberrant transcripts within the splice junction. The consequence was a significant loss in DOCK8 protein expression. Both affected siblings had an increased sensitivity to ionizing radiation. While mutations in the DOCK8 gene have been reported as causal to immune dysfunction, this is the first report that indicates a role in radiosensitivity . Also this is the first report identifying a causal mutation in an intronic site by combining RNA sequencing with exome sequencing.

2957M

Preliminary analysis of 14 Brazilian patients with Thyroid Dysgenesis using Whole Exome Sequencing. *M.M.L. Kizys[1], G.K. Furuzawa[1], M. Mitne-Neto[2], L.T. Cerdeira[2], M.C.C. Melo[1], M.G. Cardoso[3], S. Nesi-França[4], M.R. Dias-da-Silva[1], R.M.B. Maciel[1].* 1) Medicine, Universidade Federal de Sao Paulo, Sao Paulo, Sao Paulo, Brazil; 2) Fleury Group, Sao Paulo, Sao Paulo, Brazil; 3) Biochemistry, Universidade Federal de Sao Paulo, Sao Paulo, Sao Paulo, Brazil; 4) Pediatrics, Universidade Federal de Parana, Curitiba, Parana, Brazil.

Thyroid dysgenesis (TD [MIM 218700]) is the main cause of congenital hypothyroidism. It is believed that TD result from alterations in specification, proliferation, migration and survival of thyroid cell precursors. The objective of this work was to perform Whole Exome Sequencing (WES) in patients with TD. Total DNA was isolated from 3 thyroid hemiagenesis, 3 hypoplasia, 3 ectopic, 5 agenesis cases and 14 controls for WES. Exons were captured using SeqCap EZ Human Exome v3 kit (Roche NimbleGen, Madison, WI, USA) and sequenced in Illumina HiSeq2000. Mapping and variant calling were performed using the CLC Genomic Workbench 6.5. Sequencing coverage was greater than 60x. The average number of mapped reads was 67 million per sample with about 7 billion aligned bases. The first step of the analysis focused on variants in thyroid embryogenesis related genes. The other variants were classified according to the gene function were they occurred based on Gene Ontology website (http://www.geneontology.org/), including their paralogs and their receptors. The average of variants found on these genes was 846 per patient. The majority of variants were placed in non-coding regions (introns, 5' and 3' UTR). We prioritized the non-synonymous exonic variants and those characterized by loss or gain of stop codon, deletion or insertion of codons, frameshift, loss or gain of splicing sites and variants with minor allele frequency (MAF) less than or equal to 0.01. Specifically for agenesis, a symptomatic condition, we filtered the variants which frequency in the 1000Genome population database (http://www.1000genomes.org/) was lower than 0.01%. Using those filters, we found 48 variants for agenesis, 47 for ectopic gland, 45 for hypoplasia and 29 for hemiagenesis. The analysis after comparing with 14 control samples resulted in 29 variants (26 in heterozygosis) for agenesis, 26 for ectopic (25 in heterozygosis), 23 (21 in heterozygosis) for hypoplasia and 14 (13 in heterozygosis) for hemiagenesis groups. The majority of variants were observed in transcription factors coding genes, allowing us to identify new candidate genes for TD development. Herein, we first report the preliminary analysis using WES and suggest a rather complex genomic mechanism in TD development. Financial support: FAPESP.

2958T

Whole-exome sequencing as a diagnostic tool reveals *POC1A* mutation in Primordial Dwarfism. *A. Koparir[1], H. Ulucan[1], O.F. Karatas[1,2], A. Gezdirici[1], B. Yuceturk[3], B. Yuksel[3], A.O. Bayrak[4], O.F. Gerdan[4], E. Fenercioglu[1], M. Seven[1], A. Yuksel[1], M. Ozen[1,5].* 1) Medical Genetics Department, Istanbul University, Istanbul, Turkey; 2) Molecular Biology and Genetics Department, Erzurum Technical University, Erzurum, Turkey; 3) Genetic Engineering and Biotechnology Institute, TUBITAK Marmara Research Center, Kocaeli, Turkey; 4) Information Technologies Institute, TUBITAK Marmara Research Center, Kocaeli, Turkey; 5) Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX, 77030, USA.

Primordial dwarfism is a group of genetically heterogeneous disorders and the term is used for describing prenatal onset growth failure, which persists postnatally. Short stature, onychodysplasia, facial dysmorphism and hypotrichosis (SOFT) syndrome (MIM #614813), a recently described disorder, is characterized by prenatal onset growth retardation, brachydactyly, onychodysplasia, postpubertal onset hypotrichosis, normal psychomotor development, facial dysmorphisms, short long bones, femoral and vertebral body ossification delay. The only associated gene with this syndrome, *POC1A*, encodes POC1A protein which localizes to centrioles and spindle poles. In the current study, we report two patients from a family with primordial dwarfism. The clinical pictures of our patients resembled 3M syndrome particularly, because of short stature, facial dysmorphism and tall vertebral bodies. We utilized WES in our primordial dwarfism patients to screen all primordial dwarfism related genes and to define putative novel candidate genes. A novel homozygous missense mutation was detected in *POC1A* and was confirmed by using Sanger sequencing. Parents have also been demonstrated to be heterozygous carriers for this mutation. This result shows that the clinical pictures of the patients are more consistent with SOFT syndrome rather than 3M syndrome. There are two more reports associated with SOFT syndrome, therefore, here we report the third family and a novel mutation. Moreover, to confirm the pathogenicity of the detected mutation, punch biopsies from the patients and a control individual were used to culture fibroblasts and demonstrate the effects of the identified mutation through a series of functional studies.

2959S

Mutation of *CLPB*, a human homologue of bacterial ClpB /yeast Hsp104 mitochondrial molecular chaperone, causes a novel form of autosomal recessive 3-methylglutaconic aciduria. *C. Saunders[1,2], F. Wibrand[3], P. Bross[4], K. Ravn[3], L. Smith[1], I. Thiffault[1], A. Atherton[5], N. Miller[1], E. Farrow[1], S. Kingsmore[1,2], E. Østergaard[2].* 1) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO; 2) Department of Pathology and Laboratory Medicine, Children's Mercy Hospital, Kansas City, MO; 3) Department of Clinical Genetics, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark; 4) Research Unit for Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark; 5) Department of Pediatrics, Section of Genetics,Children's Mercy Hospital, Kansas City, MO.

Methylglutaconic aciduria comprises a heterogeneous group of five different syndromes of inborn errors of metabolism characterized by mitochondrial dysfunction (3-MGA-uria: MIM#250950). The metabolic landmark is urinary excretion of 3-methylglutaconic acid and its derivative, 3-methylglutaric acid. Types I-III and V are caused by mutations in four different genes, however, the genetic cause of 3-MGA-uria type IV has not been established. Here we describe a novel, lethal 3-MGA-uria characterized by onset within the first nine months with severe hypotonia, myoclonus, bilateral cataracts, and lactic acidosis, associated with variants in the *CLPB* gene encoding the ClpB caseinolytic peptidase B homolog. Mutations were identified in five patients, including four Greenlandic individuals from three families not known to be related, and one from North America. Homozygosity mapping in the Greenlandic patients identified a single 4.5 Mb region of interest on chromosome 11, encompassing 62 genes. Sequencing of putative mitochondrial genes in this region yielded the same homozygous *CLPB* variant, c.803C>T (p.Thr268Met), in all four patients. The variant changes a highly conserved threonine to a methionine, located in the ankyrin-repeat domain. It is predicted to be pathogenic by SIFT, PolyPhen2 and MutationTaster. Screening of Greenland controls showed a carrier frequency of 3.3%, similar to other severe recessive diseases in this population. Trio exome sequencing of the North American family revealed two nonsense variants in *CLPB*, p.Lys321* and p.Arg417*. Western blot analysis showed absence of CLPB protein in fibroblasts from the Greenland patients and liver from the North American patient. *CLPB* encodes a mitochondrial chaperone important for mitochondrial protein biogenesis, and is required to maintain organelle function under stress conditions. CLPB has mainly been studied in bacterial ClpB /yeast Hsp104, and belongs to the Clp/Hsp104 family (Hsp100 family), a group of sequence-related AAA+ proteins containing two consensus ATP-binding sites. Our data indicate that recessive variants in *CLPB* result in a specific phenotype including a new form of 3-MGA-uria. Further studies in these patients may provide new insights into mammalian mitochondrial dysfunction through protein aggregation in neurodegenerative disorders.

2960M

Leveraging Population Structure to Improve Causal Variant Identification in Exome Sequencing Studies of Mendelian Diseases. *R. Brown[1], H. Lee[2], A. Eskin[3], G. Kichaev[1], K. Lohmueller[1,4], B. Reversade[5], S. Nelson[2,3], B. Pasaniuc[1,2,3].* 1) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA; 2) Department of Pathology and Laboratory Medicine, Geffen School of Medicine, UCLA, Los Angeles, CA; 3) Department of Human Genetics, Geffen School of Medicine, UCLA, Los Angeles, CA; 4) Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, CA; 5) Institue of Medical Biology, A*STAR, Singapore.

Recent breakthroughs in whole exome sequencing technologies have made possible the identification of many causal variants of Mendelian (monogenic) diseases. Although extremely powerful when related individuals (e.g. family trios) are simultaneously sequenced, exome sequencing of unrelated individuals with Mendelian traits is often unsuccessful due to the large number of variants that need to be followed-up for validation. The standard approach for reducing the number of putatively causal variants is to filter out all common variants above a pre-specified frequency threshold (e.g. 1%), followed by prioritization of the remaining variants according to their predicted functionality. In this work, we present new methods that leverage population structure (e.g. the variability of allele frequencies across populations) to improve the variant filtering step. We show that the existing filtering approaches are susceptible to high false negative rates (i.e. probability of missing a true causal variant) when reference panel sizes are limited. We introduce a framework that takes into account the finite size of the reference panels while correctly accounting for the ancestry of the sequenced individuals. In simulations starting from 1000 Genomes data, we show that our methods reduce the number of variants to be followed-up from an average of 408.1 variants per exome when ancestry is ignored to 254.8 when ancestry is taken into account (a 38% reduction). Most importantly, our proposed approaches are well calibrated with respect to the false negative rate. We validate our findings using exome-sequencing data from 22 unrelated individuals with Mendelian diseases for which the true causal variants are known. In this real data our method is able to reduce the number of variants for follow-up by 42% without filtering the true causal variants.

2961T

Truncating Mutation in CIB2 causes DFNB48 and not USH1J. *K.T. Booth[1], K. Kahrizi[2], A.C. Simpson[1], A.E. Shearer[1], C.M. Sloan[1], H. Najmabadi[2], H. Azaiez[1], R.J.H. Smith[1].* 1) Otolaryngology - Head & Neck Surgery, University of Iowa, Iowa City, IA; 2) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

Deafness is the most frequent sensory disorder worldwide affecting one of every five hundred newborns. Currently there are over 90 genes causally related to hearing loss and of those 55 are associated with autosomal recessive non syndromic hearing loss (ARNSHL) making deafness extremely heterogenous. Here we report a consanguineous family of Iranian origin segregating ARNSHL. Audiologic examination of affected individuals revealed bilateral profound hearing loss across all frequencies. Ophthalmological examination was normal. To identify the genetic cause of deafness in this family, we concurrently performed linkage analysis to identify segments of homozygosity-by-descent and whole exome sequencing (WES) in one affected individual. The deafness-causing gene in this family mapped to the DFNB48/USH1J locus and WES data revealed a novel homozygous large deletion encompassing exon 2 of CIB2 gene. Variants in CIB2 are known to cause ARNSHL. The deletion we identified leads to a frameshift and premature stop codon at amino acid 24; c.52_86del: p.Asp18Alafs*7, yielding a null allele. We used long-range PCR and Sanger sequencing to map the breakpoints to a 3.1 kb interval that spans CIB2 exon 2 and includes several repetitive elements (LINE, SINE...) on either side of the deletion. This type of sequence homology favors genomic rearrangement by non-allelic homologous recombination. CIB2 encodes the calcium- and integrin-binding protein 2 (CIB2), which is expressed in both the retina and inner ear, where it localizes to the tips of stereocilia. In the inner ear CIB2 interacts with MYO7A and WHRN, making it part of the USH1 interactome. To date only four missense mutations have been described in CIB2. Two of them represent a major cause of ARNSHL in the Pakistani population. This study is the first to report of a copy number variation associated with CIB2-related deafness and further uncovers the genetic spectrum of deafness in the Iranian population.

2962S

A Perturbed Transcriptome Underlies Cornelia de Lange Syndrome and Related Phenotypes. *B. Yuan[1], D. Pehlivan[1], E. Karaca[1], N. Patel[2], T. Gambin[1], C. Gonzaga-Jauregui[1], V.R. Sutton[1], G. Yesil[3], S.T. Bozdogan[4], T. Tos[5], E. Koparir[6], C.R. Beck[1], S. Gu[1], H. Aslan[7], O.O. Yuregir[7], K. Rubeaan[8], D. Nakeeb[8], M. Alshammari[9], Y. Bayram[1], M.M. Atik[1], H. Aydin[10], B. Geckinli[10], S. Jhangiani[11], D.M. Muzny[11], E. Boerwinkle[11,12], B. Tuysuz[13], F.S. Alkuraya[2,9], R.A. Gibbs[1,11], J.R. Lupski[1,14,15], Baylor-Hopkins Center for Mendelian Genomics.* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 3) Department of Medical Genetics, Bezmialem University, Istanbul, Turkey; 4) Department of Medical Genetics, Mersin University, Mersin Turkey; 5) Department of Medical Genetics, Sami Ulus Children's Hospital, Ankara, Turkey; 6) Department of Medical Genetics, Cerrahpasa Medical School of Istanbul University, Istanbul, Turkey; 7) Department of Medical Genetics, Adana Numune Hospital, Adana, Turkey; 8) University Diabetes Center, College of Medicine, King Saud University, Riyadh, Saudi Arabia; 9) Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia; 10) Center of Genetics Diagnosis, Zeynep Kamil Maternity and Children's Training and Research Hospital, Istanbul, Turkey; 11) Human Genome Sequencing Center, Baylor College of Medicine, Houston TX, USA; 12) University of Texas Health Science Center at Houston, Houston, Texas, United States of America; 13) Department of Pediatrics, Division of Medical Genetics, Cerrahpasa School of Medicine, Istanbul, Turkey; 14) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 15) Texas Children's Hospital, Houston, TX, USA.

Cornelia de Lange syndrome (CdLS) is a genetically heterogeneous disorder manifesting extensive phenotypic variability. To date, mutations in *NIPBL*, *SMC1A*, *SMC3*, *RAD21* and *HDAC8*, which encode subunits or regulators of the cohesin complex, are found in about 65% of patients. Wiedemann-Steiner syndrome (WDSTS), caused by mutations in *KMT2A*, shares phenotypic features with CdLS. We utilized genomic approaches to determine additional molecular etiologies for CdLS-like phenotypes and investigate molecular underpinnings of shared clinical features. Whole exome sequencing (WES) of two male siblings clinically diagnosed with WDSTS revealed a hemizygous, predicted-deleterious, missense mutation in *SMC1A*. Extensive clinical evaluation and WES of a Turkish cohort of 30 patients clinically diagnosed with CdLS revealed a *de novo* heterozygous nonsense *KMT2A* mutation in one patient without characteristic WDSTS features. Moreover, a *de novo* heterozygous frameshift mutation in *SMC3* was identified in a patient with combined CdLS and WDSTS features. Furthermore, in families from two separate world populations segregating an autosomal recessive disorder with CdLS-like features, we identified homozygous mutations in *TAF6*, which encodes a core component in a transcriptional regulation pathway. Our findings suggest CdLS and related phenotypes may result from a "transcriptomopathy" rather than a cohesinopathy: a conclusion supported by recent transcriptomic studies.

2963M

Deep re-sequencing of *CFTR* bearing the common F508del mutation reveals a rare variant associating with variation in lung infection. *B. Vecchio-Pagán[1], M. Lee[1], M.R. Knowles[3], S.M. Blackman[2], G.R. Cutting[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Division of Pediatric Endocrinology, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Cystic Fibrosis/Pulmonary Research and Treatment Center, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Common and rare intragenic variation has the potential to modify protein expression that may correlate with severity of disease phenotypes. Here we describe deep re-sequencing of a well-characterized locus, *CFTR*, in 482 patients homozygous for the most common disease mutation, F508del (rs113993960). Variants in *CFTR* cause the autosomal recessive disorder, cystic fibrosis (MIM #219700), in which the primary cause of mortality is lung disease complicated by infection. A custom designed capture of the 215kb region containing *CFTR* in 964 chromosomes harboring the F508del mutation revealed 272 common variants (MAF > 1%), and 307 rare variants. 301 rare SNPs and small INDELs present on 6 or more chromosomes were used to define linkage disequilibrium (LD) blocks and construct unique haplotypes in this population. Two major LD blocks were observed, and could be explained by a recombination event in intron 16. These results further refine previous reports of recombination occurring within intron 22. The most common haplotype in both of these blocks accounts for 63% of all F508del chromosomes, suggesting that this is the ancestral F508del background. The remaining 37% of F508del chromosomes are combinations of minor haplotypes from both blocks, indicating a wealth of genotypic diversity with the potential for modifying disease severity. Testing of 579 total variants (~32 independent tests due to LD) with cystic fibrosis disease-specific traits revealed association between the rs35453239 A allele and later age at first *Pseudomonas aeruginosa* infection (p=9.16e-5, Bonferroni p=2.93e-3). The A allele of rs35453239 was present on 3 haplotype backgrounds (frequencies: 1.6%, 1.0% 0.7%); two of which were associated with later infection (p=9.16e-5 and p=5.62e-6) and each differs from the third haplotype at a SNP within intron 21 (chr7:117259340A>G; MAF: 0.016). These findings indicate that rs35453239 modifies age at infection only in the presence of the major allele at chr7:117259340. Under this condition, each rs35453239 A allele associates with ~2.32 yrs later age at first lung infection. Of note, both variants are located near putative transcription factor binding sites (ENCODE), suggesting a transcriptional regulation mechanism of effect. These results demonstrate the value of re-sequencing a well-characterized disease locus in an affected population to elucidate origins of phenotypic variability.

2964T

Whole Exome Sequencing in Autosomal Recessive Non-syndromic Deafness: 4 years experience. *G. Bademci[1], O. Diaz-Horta[1], I. Menendez[1], D. Duman[2], J. Foster II[1], F.B. Cengiz[2], N. Mahdieh[3], M. Bonyadi[4], F. Huesca Hernandez[5], M. Arenas Sordo[5], J. Dominguez-Aburto[5], E. Hernandez-Zamora[5], P. Montenegro[6], R. Paredes[6], G. Moreta[6], R. Vinueza[6], F. Villegas[6], R. Alyazidi[7], S. Mendoza Benitez[8], S. Guo[1], S.H. Blanton[1], M. Tekin[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Ankara University School of Medicine, Division of Genetics, Department of Pediatrics, Ankara, Turkey; 3) Deputy of Research and Technology, Ministry of Health and Medical Education, Tehran, Iran; 4) Faculty of Natural Sciences, Center of Excellence for Biodiversity, University of Tabriz, Tabriz, Iran; 5) Genetic Service. National Institute of Rehabilitation. Mexico D.F, Mexico; 6) Molecular Genetic Lab, FF.AA. Hospital. Quito, Ecuador; 7) Department of Pediatrics, King Abdulaziz University Hospital, Jeddah, Saudi Arabia; 8) Audiology Department. Cuernavaca General Hospital, Morelos, Mexico.

Identification of pathogenic mutations underlying autosomal recessive non-syndromic deafness (ARNSD) is difficult, since causative mutations in 55 genes have been reported. After excluding mutations in the most common gene, GJB2, we performed whole-exome sequencing (WES) in 109 unrelated multiplex families with ARNSD. Families in our sample set originated from Turkey, Iran, Saudi Arabia, Mexico, Ecuador, and Puerto Rico. Agilent SureSelect Human All Exon 50 Mb kits and an Illumina Hiseq2000 instrument were used. GATK (Genome Analysis Tool Kit) was used for detection of Single Nucleotide Variants (SNVs) and insertion/deletions (INDELs); CoNIFER (Copy Number Inference From Exome Reads) and XHMM (eXome-Hidden Markov Model) were used for Copy Number Variation (CNV) detection. Autozygous regions as extended runs of homozygous genotypes in the studied samples were sought via WES data with AgileGenotyper and AgileVariantMapper. Candidate variants were confirmed with Sanger sequencing and evaluated for co-segregation with deafness in all available family members. We identified causative variants in 65% of the families in our sample set. The 10 most common genes with either an SNV or an INDEL are MYO15A, SLC26A4, TMC1, TMPRSS3, ILDR1, LOXHD1, MYO7A, OTOF, PCDH15, and USH1C. Homozygous CNVs were detected in OTOA and STRC. Analysis of the WES data in the families remaining negative for causative variants in known deafness genes is used to identify novel deafness genes. In conclusion, comprehensive analysis of SNVs, INDELs and CNVs via WES allows us to identify causative variants in both known and novel genes thus improving our ability to explain the underlying etiology in more families.

2965S

Resolving clinical diagnoses for syndromic cleft lip and palate phenotypes using whole-exome sequencing. *A. Collins[1], L. Arias[2], J. Martinez[2], R. Upstill-Goddard[1], R. Pengelly[1], J. Gibson[1], S. Ennis[1], I. Briceno[2].* 1) Human Genetics, Univ Southampton, Southampton, United Kingdom; 2) Department of Biomedical Sciences, Medical School, Universidad de La Sabana, Bogota, Colombia.

Individuals from three families ascertained in Bogota, Colombia, showing unusual syndromic phenotypes which included cleft lip and/or palate were exome sequenced. In each case exome sequencing revealed the underlying causal variations confirming or establishing diagnoses. The findings include rare and novel variants which provide insights into genotype and phenotype relationships for the conditions. These include the molecular diagnosis of an individual with Nager syndrome and another family exhibiting an atypical Incontinentia Pigmenti phenotype for which we identified a previously reported missense mutation in the IKBKG gene. In this family this rare mutation is associated with a less severe phenotype in which affected male relatives, atypically, survive to full term. The third family exhibited unusual and variably penetrant phenotypic features suggesting the Pierre Robin Sequence (PRS). Affected individuals were found to share a novel mutation in the IRF6 gene which has not previously been associated with PRS. Exome sequencing is a powerful and increasingly cost-effective route to establishing molecular diagnoses for conditions showing phenotypic and/or genetic heterogeneity. It is particularly effective for conditions featuring cleft lip and/or palate for which a substantial proportion of underlying causal genes have not yet been identified.

2966M

Recurrent mutations cause Ablepharon-macrostomia syndrome and Barber-Say syndrome. *T. Davis[1], S. Marchegiani[1], F. Tessadori[2], C. Markello[1], H. Huang[3], D. Schanze[4], A. Hoischen[5], G. van Haaften[2], V. Maduro[1], F. Brancati[6], B. de Vries[5], C. Boerkoel[1], S. Lin[3], W. Gahl[1,7], M. Zenker[4], T. Markello[1], AMS/BSS Consortium.* 1) Undiagnosed Diseases Program, Office of the Director, National Institutes of Health, Bethesda, MD, USA; 2) Hubrecht Institute-KNAW, University Medical Center Utrecht, Utrecht, Netherlands; 3) Department of Molecular, Cell and Developmental Biology, University of California-Los Angeles, Los Angeles, CA, USA; 4) Medizinische Fakultat und Uniklinikum Magdeburg, Institute of Human Genetics, Magdeburg, Germany; 5) Radboud University Nijmegen Medical Centre, Department of Human Genetics, Nijmegen, Netherlands; 6) University of G. d'Annunzio Chieti and Pescara, Department of Medical, Oral and Biotechnological Sciences, Chieti, Italy; 7) Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD, USA.

Ablepharon-macrostomia syndrome (AMS, MIM 200110) and Barber-Say syndrome (BSS, MIM 209885) are rare congenital disorders characterized by a similar array of craniofacial and dermal features. In order to identify the genetic basis of these disorders, we performed exome and candidate gene sequencing. Seven independent families were found to have a recurrent missense mutation causing AMS, and nine other independent BSS families had a different recurrent missense mutation of the same amino acid. In all cases where unaffected parental DNA was available, we found that these mutations arose de novo. By detailed genetic and molecular analyses, we show that these mutations are neomorphic in nature.

2967T

Mutations in *MAB21L2* result in ocular coloboma, microcornea, and cataracts. *B. Deml[1,2], A. Kariminejad[3], R.H.R. Borujerdi[4], S. Muheisen[1], L. Reis[1], E.V. Semina[1,2].* 1) Department of Pediatrics and Children's Research Institute, Medical College of Wisconsin and Children's Hospital, Milwaukee, WI; 2) Cell Biology, Neurobiology and Anatomy, Medical College of Wisconsin, Milwaukee, WI; 3) Kariminejad-Najmabadi Pathology and Genetics Center, Tehran, Iran; 4) Qom Welfare Organization, Qom, Iran.

Ocular coloboma is a congenital defect defined as the incomplete closure of the optic fissure and is often associated with additional ocular and systemic defects. Numerous genetic mutations have been identified in coloboma, but many cases still remain to be explained. Whole exome sequencing was performed to identify the causative mutation in a three-generation family with coloboma, microcornea, cataracts and skeletal dysplasia. Sequencing of affected first cousins identified a shared heterozygous allele in *MAB21L2* [MIM 604357], c.151C>G, p.(Arg51Gly), as the most likely causative mutation. Confirmation and segregation analyses were performed by Sanger sequencing and revealed that the c.151C>G, p.(Arg51Gly) allele was present in all five affected individuals and absent in the unaffected grandparents. The mutation was predicted to be damaging by SIFT, Polyphen2, Mutation Taster and MutationAssessor, and demonstrated high conservation scores for the affected nucleotide. Function of MAB21L proteins is largely unknown. *In silico* analysis using I-TASSER predicted the MAB21L2 structure to most closely resemble human cyclic GMP-AMP synthase (cGAS), a protein involved in innate immunity and binding of cytosolic double stranded DNA. According to this model, the arginine at position 51 is predicted to be located only 4 amino acids downstream of the possible DNA binding helix and thus may play a role in this possible interaction. *In situ* hybridization in zebrafish identified *mab21l2* transcripts in the presumptive eye field and midbrain at 18-hours post fertilization with continued strong expression in the retina, lens, brain, spinal cord, pharyngeal arches and developing fins at later developmental stages. To further evaluate the role of *mab21l2* in development, the gene was disrupted using TALEN genome editing technology targeting the region of the human mutation. Zebrafish carrying either homozygous or compound heterozygous frameshift mutations demonstrated microphthalmia with small or absent lenses in 100% of embryos and ocular coloboma in 76%. Zebrafish homozygous for an in-frame deletion, p.(Arg51_Phe52del), developed ocular coloboma in 100% of embryos but had overall normal lens appearance and eye size. These findings support *MAB21L2* as a novel factor involved in human ocular disease. Further, this study demonstrates the power of TALEN technology for the rapid development of animal models to study candidate genes found in human whole exome sequencing studies.

2968S

Homozygous *LRRC10* Mutation in Sporadic Pediatric Dilated Cardiomyopathy. *P.A. Long[1], J.M. Evans[2], B.T. Larsen[3], A.C. Grimes[4], Y. Lee[5], R.C. Balijepalli[4], T.M. Olson[1,6].* 1) Cardiovascular Genetics Research Laboratory, Mayo Clinic, Rochester, MN; 2) Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN; 3) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 4) Cellular and Molecular Arrhythmia Research Program, Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI; 5) Department of Cell and Regenerative Biology, Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI; 6) Division of Pediatric Cardiology, Mayo Clinic, Rochester, MN.

Idiopathic dilated cardiomyopathy (DCM) is a heritable, genetically heterogeneous disorder characterized by progressive heart failure. DCM typically exhibits an autosomal dominant mode of inheritance and remains clinically silent until adulthood. We sought to identify the genetic basis of severe, non-syndromic DCM diagnosed in a two-month-old girl who underwent cardiac transplantation two months later. Her non-consanguineous parents displayed no phenotypic evidence of DCM on screening echocardiography. Array comparative genomic hybridization detected no clinically reportable chromosomal aberrations. Whole exome sequencing of the family trio revealed approximately 40,000 variants per individual in the 50 megabase capture region. Ingenuity® Variant Analysis™ was utilized to filter for rare, functionally significant variants and model homozygous recessive, compound heterozygous, *de novo*, and uniparental disomy modes of inheritance, resulting in only 3 candidate variants. Homozygosity for c.584T>C, p.Ile195Thr in the leucine rich repeat containing 10 gene (*LRRC10* [MIM 610846]) was identified in the proband, the only variant predicted to be damaging by SIFT and PolyPhen2. Heterozygosity for I195T occurs at a frequency of 0.44% in the population. However, no instances of compound heterozygosity or homozygosity of missense, splice site, or frameshift variants in *LRRC10* were identified in 7,595 exomes in Exome Variant Server and 1000 Genomes. While the molecular function of LRRC10 remains unknown, it was shown previously to have cardiac-specific expression in zebrafish, mice, and humans, with markedly increased expression at birth that is maintained through adulthood in mice. Knockout of *Lrrc10* in mice was shown to cause prenatal systolic dysfunction and postnatal DCM. Preliminary immunohistochemical studies of explanted heart tissue revealed LRRC10-I195T expression in cardiomyocytes. The I195 residue is located in the 7th leucine rich repeat at the 2nd conserved Leu within the highly conserved segment LxxLxLxxNxL. We speculate that substituting this conserved hydrophobic isoleucine with a polar threonine may disrupt proper protein folding and/or interactions with critical cardiac protein(s). In conclusion, we have identified *LRRC10* as a novel gene associated with autosomal recessive, pediatric DCM. Functional studies are underway to clarify the mechanism by which genetic perturbation of this cardiac-specific gene leads to early-onset heart failure.

2969M

Mutation in ANKFY1 as a Cause of Charcot-Marie-Tooth Neuropathy. *M. Park[1], B. Choi[2], S. Choe[3], H. Woo[1], H. Jo[1], K. Chung[4], S. Koo[1].* 1) Division of Intractable Diseases, National Institute of Health, Chungcheongbuk-do, South Korea; 2) Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea; 3) Department of Microbiology, School of Medicine, Wonkwang University, Iksan, South Korea; 4) Department of Biological Science, Kongju National University, Gongju, South Korea.

Charcot-Marie-Tooth disease (CMT) is the most common inherited neuromuscular disorder involving motor and sensory peripheral neuropathies. We investigated the disease-causing mutation in a Korean CMT2 family with not linked to major genes, and examined the pathogenic effect of the mutation. We performed the whole-exome sequencing for two patients. After filtering out with known variants, we focused on linkage analysis, sharing between two individuals and co-segregating within the family. We identified the causative mutation in the ANKFY1 gene which has never been reported to be caused or associated with motor and sensory peripheral neuropathies yet. The identified mutation is located in the ankyrin-repeat domain. It is known to be a cause of neurological disorder when the repeat domain is disrupted by mutations. Knockdown of ANKFY1 in zebrafish resulted in frequently disconnected motor axons and disorganized neuromuscular junctions, supporting the pathogenic role of mutated ANKFY1 in peripheral neuropathy. Therefore, our data strongly suggest that ANKFY1 mutation is exposed as a novel cause of CMT2. Furthermore, this study may improve understanding of ankyrin-repeat domain-associated axonal CMT neuropathies and their debated pathologic mechanism.

2970T

Characterization of mutation negative autosomal dominant polycystic kidney disease families using whole exome sequencing. *B.M. Paul, K. Hopp, J.L. Sundsbak, C.M. Heyer, V.E. Torres, P.C. Harris.* Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN.

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is a common (frequency 1: 1000), adult onset nephropathy accounting for 4-10% of patients requiring dialysis or renal transplant worldwide, with an annual US cost of >$1 billion. Mutations to either *PKD1* or *PKD2* cause most ADPKD. However, the description of ADPKD families unlinked to either gene, and no mutations detected in ~8% of ADPKD families, suggests further genetic heterogeneity. We identified ~130 no-mutation detected (NMD) families out of ~1500 screened. Interestingly, they are characterized by milder kidney disease and more frequently have negative family history than is typical in ADPKD. Initially, we analyzed two families by whole exome sequencing (WES), P75 (3 affected, 2 unaffected) and M560 (3 affected). Library preparation and capture was done using the Agilent SureSelectXT Human All Exon V5+UTRs on the Agilent Bravo workstation and sequencing employed on Illumina HiSeq2000. Alignment was performed using Bowtie/BWA and GATK used for variant calling according to best standard practices. Data mining was performed using the Golden Helix SNP and Variation suite (SVS). Initial filtering was set by read depth (10X), genotype quality (≥20) and assignment of autosomal dominant inheritance according to the affected status of the individuals. The ESP6500 and 1000 Genomes databases were used to remove variants with a MAF of ≥ 0.1% and dbNSFP to characterize coding variants. Independent VAAST analysis was also employed. We found a *PKD1* nonsense mutation in M560, previously missed by Sanger sequencing. In P75, 15 candidate variants were detected by both SVS and VAAST and six by VAAST alone. Eleven variants were removed due to misalignment or because of poor consensus score in dbNSFP. Five variants were annotated as damaging by four prediction tools in dbNSFP and five others by two tools. We are prioritizing these ten candidates by their likely localization and involvement in processes and pathways associated with ADPKD. As a next step, we are analyzing six other multiplex NMD families by WES. Finding variants in the same protein, or related pathways/functions, in different families will help identify high-level candidates. Analysis of these candidates will be extended to simplex NMD families using Agilent HaloPlex enrichment. Our large and well characterized cohort of NMD ADPKD families and modern genomic screening methods are allowing rigorous testing for further genetic heterogeneity in ADPKD.

2971S

Whole Exome Sequencing to Uncover Causative Genes in Families with Inherited Autonomic Dysfunction. *J.E. Posey[1], T. Gambin[1], S.N. Jhangiani[2], D.M. Muzny[2], R. Martinez[3], J.E. Lankford[3], M.T. Numan[4], W. Wiszniewski[1], R.A. Gibbs[1,2], I.J. Butler[3], J.R. Lupski[1,2,5,6], Baylor Hopkins Center for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 3) Division of Child and Adolescent Neurology, Department of Pediatrics, The University of Texas Health Science Center, Houston, TX 77030, USA; 4) Division of Cardiology, Department of Pediatrics, The University of Texas Health Science Center, Houston, TX 77030, USA; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 6) Texas Children's Hospital, Houston, TX 77030, USA.

Autonomic dysfunction can be a crippling medical condition, and is characterized by a variety of symptoms including dizziness, palpitations, syncope, temperature dysregulation, increased pain sensitivity, and gastrointestinal distress. These clinical manifestations may be associated with several well-defined genetic syndromes, including Hereditary Sensory and Autonomic Neuropathy, Ehlers-Danlos Syndrome or Familial Dysautonomia [OMIM 223900]. Familial Dysautonomia, also known as Riley-Day Syndrome, is an autosomal recessive type 3 Hereditary Sensory and Autonomic Neuropathy in which patients suffer from cardiovascular instability, gastrointestinal dysfunction, and altered sensitivity to pain and temperature. We identified 14 families in which autonomic dysfunction is inherited in an autosomal dominant manner. Extensive clinical characterization has demonstrated a very specific phenotype that includes postural orthostatic tachycardia syndrome (POTS), atrial septal defect (ASD), headaches, and gastrointestinal distress. These patients do not have decreased or absent lingual fungiform papillae, a hallmark feature of Familial Dysautonomia, and they do not report Ashkenazi Jewish heritage. Given the distinct phenotype, ethnicity, and mode of inheritance seen in these families, we suspect that their autonomic dysfunction is caused by mutations in novel gene(s) playing a role in development and/or functioning of the nervous system. We have performed whole exome sequencing and cSNP analysis on affected members of these families. We are using a combination of cohort analysis, candidate gene analysis, copy number analysis, and individual variant review to identify potentially pathologic variants. Thus far, 17 potentially pathogenic variants have been identified, and we are actively recruiting additional affected family members to further investigate these variants. We are also continuing to enroll new families demonstrating mendelian inheritance of autonomic dysfunction.

2972M

Exome sequencing identifies a recurrent de novo mutation in *ZSWIM6* that causes Acromelic Frontonasal Dysostosis. *J. Smith[1], A. Hing[2,3,4], C. Clarke[4], N. Johnson[2], F. Perez[7], S. Park[4], J. Horst[5], B. Mecham[6], L. Maves[2,3], D. Nickerson[1], M. Cunningham[2,3,4], University of Washington Center for Mendelian Genomics.* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA; 3) Department of Pediatrics, University of Washington, Seattle, WA; 4) Craniofacial Center, Seattle Children's Hospital, WA; 5) Department of Biochemistry and Biophysics, UCSF, San Francisco, CA; 6) Trialomics Inc, Seattle, WA; 7) Department of Radiology, UW Medicine, Seattle, WA.

Acromelic Frontonasal Dysostosis (AFND MIM 603671) is a rare genetic disorder characterized by distinct craniofacial, brain and limb malformations including frontonasal dysplasia. We performed exome sequencing on one trio and two unrelated probands, and identified an identical missense variant (p. Arg1163Trp) in a highly conserved protein domain of ZSWIM6. Sanger validation of the three trios confirmed its de novo status in the three probands. A fourth case was Sanger sequenced and the identical variant was observed. *ZSWIM6* encodes a 133.5 kDa protein containing a zinc finger SWIM domain. Proteins with SWIM domains have been identified in a diverse number of species from bacteria to eukaryotes and are predicted to have DNA binding and protein-protein interacting properties, but little has been elucidated about their function in vivo. In an effort to understand the potential impact of the p.Arg1163Trp mutation, we conducted detailed structural in silico modeling of amino acid substitutions in ZSWIM6. Interestingly, in situ hybridization of early zebrafish embryos at 24 hours post fertilization (hpf) demonstrates telencephalic expression with onset of midbrain, hindbrain and retinal expression at 48 hpf. Immunohistochemistry of later stage mouse embryos demonstrates widespread expression in the derivatives of all three germ layers. Furthermore, gene expression analysis using qRT-PCR on osteoblast and fibroblast cell lines from two of the probands revealed evidence of Hedgehog pathway activation. Our analysis suggests the mutation in *ZSWIM6* leads to rare craniofacial, brain and limb malformations through disruption of Hedgehog signaling pathway.

2973T

The Gene Discovery Core: four years of experience in determining the genetic basis of orphan diseases. *M.C. Towne[1], C.A. Brownstein[1,2], D.M. Margulies[1,2], A.H. Beggs[1,2], P.B. Agrawal[1,2].* 1) The Manton Center for Orphan Disease Research, Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA.

The Gene Discovery Core (GDC) of The Manton Center for Orphan Disease Research at Boston Children's Hospital has been enrolling individuals and families with rare and unknown disorders in an IRB-approved protocol for the last four years. Goals of the GDC include: determination of the genetic and molecular bases of orphan diseases using various genomic technologies, evaluation of the functional effects of gene mutations using various models, and return of the results to participants in a research setting with independent clinical confirmation. Enrollment is open to any family with a rare or unknown diagnosis that has a suspected, yet unidentified, genetic etiology despite extensive clinical testing. Following identification of a new family with a rare condition, we often ascertain and enroll additional families with similar phenotypes to develop a cohort to create a critical mass for further study. To date, 550 families have participated, of which 57% represent single instances of unique and undiagnosed disorders. Genetic analysis has been initiated for 263 families: 234 of which received whole exomic/genomic sequencing and 30 of whom had targeted gene sequencing of candidate genes. Analysis has been completed for 139 families. Of those with completed analysis, 39% of the families were found to have a mutation in a known gene that fits their phenotype, and these results were returned to the participants after confirmation in a CLIA-certified laboratory. For 20 families, variants were identified in novel genes without prior reports of association with human disease, and suspected to be pathogenic based on *in silico* analysis, gene expression data and pathway involvement. Following review of the literature, we are performing various follow-up studies on these cases, including analysis of mouse and zebrafish models, as well as patient-derived cell or tissue samples, to evaluate causality. Whenever possible, functional assays are developed and tailored for each gene in question. We have successfully linked the phenotype to these novel genes in three of those 20 cases, and we are actively working on the remaining families.

2974S

Exome sequencing identifies the cause of a novel multiple pterygium syndrome and expands the spectrum of phenotypes caused by variants in *MYH3*. *A.E. Beck[1, 2], J.X. Chong[1], C.T. Marvin[1], M.J. McMillin[1], H.I.S. Gildersleeve[1], K.M. Shively[1], W. Best[1], J.C. Carey[3], J. Shendure[4], D.A. Nickerson[4], M.J. Bamshad[1,2,4], University of Washington Center for Mendelian Genomics.* 1) Dept Pediatrics, Div Genetic Med, Univ Washington, Seattle, WA; 2) Seattle Children's Hospital, Seattle, WA; 3) Dept Pediatrics, Div Med Genetics , Univ Utah, Salt Lake City, UT; 4) Dept Genome Sciences, Univ Washington, Seattle, WA.

Mutations in *MYH3*, encoding embryonic myosin, cause Distal Arthrogryposis type 1 (DA1), Sheldon-Hall syndrome (SHS or DA2B), and Freeman-Sheldon syndrome (FSS or DA2A). Each of these conditions is characterized by multiple congenital contractures (MCC) but have distinct clinical characteristics and natural histories that enable them to be distinguished from one another. Individuals with DA type 8 (DA8) likewise have MCC but also have short stature, severe scoliosis and multiple pterygia reminiscent of multiple pterygium syndrome (MPS). Sanger sequencing of candidate genes (*TNNI2*, *TNNT3*, *TPM2*, *CHRNG*, and *RAPSN*) that underlie DA1, DA2B, and MPS in a parent and two affected children with DA8 failed to identify a compelling causal variant. Subsequently exome sequencing revealed a novel 3-base pair duplication in exon 25 of *MYH3*, c.3214_3216dup predicted to lead to an in-frame duplication of a single amino acid residue, p.Asn1072dup. Pathogenic variants in *MYH3* were first identified in individuals with FSS, and subsequent screening of phenotypes with clinical characteristics overlapping FSS led to finding *MYH3* variants that cause SHS and DA1. While the phenotype of DA8 is similar to FSS and SHS, it was distinct enough for *MYH3* not to have been considered a compelling candidate gene. In contrast, exome sequencing (ES) offers the advantage of searching for causal variants with less bias, which can lead to the identification of variants in a single gene that cause new phenotypic features of known conditions (i.e., phenotypic expansion) as well as novel conditions. In large-scale ES efforts to find genes for Mendelian conditions, the identification of genes in which different variants can have very diverse phenotypic effects is becoming increasingly commonplace. We call this property of a gene, specifically the spectrum of phenotypes associated with variants in a given gene, phenotropy in contrast to pleiotropy, which is the spectrum of different effects of a given gene on the body. It also contrasts with allelic heterogeneity, in which variants in a given gene cause variable expression of the same phenotype. We suggest that one of the eventual major impacts of genotype-driven characterization of Mendelian conditions enabled by ES will be a more complete catalogue of phenotropy for virtually every human gene.

2975M

Exome sequencing to identify the genetic bases for lysosomal storage diseases of unknown etiology. *N. Wang[1,2], E. Gedvilaite[1], D. Kumar[3], R. Donnelly[4], J. Xing[1,2], D. Sleat[5,6], P. Lobel[5,6].* 1) Department of Genetics, Rutgers University, Piscataway, NJ; 2) Human Genetics Institute of New Jersey, Rutgers University, Piscataway, NJ; 3) Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ; 4) Molecular Resource Facility, Rutgers - New Jersey Medical School, Newark, NJ; 5) Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, NJ; 6) Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers University, Piscataway, NJ.

Lysosomes are membrane-bound, acidic eukaryotic cellular organelles. As an enzyme container, they play important roles in the degradation of macromolecules. Monogenic mutations resulting in the loss of enzyme activities in the lysosome may lead to severe health problem, such as neurodegeneration, early death, etc. These conditions are categorized as lysosome storage diseases (LSDs). The diagnosis of LSDs is typically straightforward, but in some cases, mutations that result in atypical clinical presentation or defects in previously undescribed lysosomal disease genes may complicate the identification of the underlying genetic defect. Here, we performed whole exome sequencing on 14 suspected LSD cases, with the goal of finding the causal mutations in each case. From the raw sequence data, we first identified DNA variants in each individual using three variant discovery pipelines: the Genome Analysis Toolkit, LifeScope and CLC Genomics Workbench. We then used the Variant Annotation Analysis Selection Tool (VAAST) to prioritize disease-causing mutations in 848 candidate LSD genes. As a probabilistic disease gene finder, VAAST integrates allele frequency, amino acid substitution severity and conservation information into a composite likelihood framework. Different from hard filtering methods, VAAST preserves all the candidates by listing them according to their disease-causing potential. So far a number of candidate variants have been identified, and we are performing downstream mutation validation and proteomic analyses to investigate the potential connection between our candidate variants and LSDs. Our project makes use of bioinformatics analyses to decode enormous exome sequencing data, narrowing down candidate lists and largely increasing the efficiency of downstream proteomic studies. Our results will shed light on the genetic basis of LSDs.

2976T

Clinically significant copy number variants (CNVs) in a cohort of retinal dystrophy probands inferred from whole exome sequence data. *G. Arno[1], S. Hull[1,2], P. Sergouniotis[1,2], E. Lenassi[1,2], A.R. Webster[1,2], V. Plagnol[3], A.T. Moore[1,2].* 1) Inherited Eye Diseases, ORBIT, UCL Institute of Ophthalmology, London, United Kingdom; 2) Moorfields Eye Hospital, London, United Kingdom; 3) University College London Genetics Institute, London, United Kingdom.

Introduction High throughput exome sequencing has enabled the detection of many, novel rare genes for Mendelian disorders. However, there is limited discussion of the inference of clinically meaningful CNVs from the same whole exome sequencing data. We sought to determine the frequency of detectable CNVs in a series of retinal dystrophy patients who had undergone exome sequencing. **Methods** Patients comprised 101 probands affected with a retinal dystrophy. Exome sequencing was carried out and reads were aligned to the hg19 human reference sequence. CNV analysis was carried out with ExomeDepth(1), based on read depth data. Exome data displayed using the Integrated Genomics Viewer (IGV - Broad Institute) was inspected directly over candidate genes to detect reads having an unusual orientation. **Results** 54/101 patients had likely disease causing homozygous or compound heterozygous mutations in known retinal dystrophy genes. Of these, 5 CNVs were detected by ExomeDepth. A homozygous deletion of exons 2-7 of *TRPM1* in a patient with congenital stationary night blindness (CSNB) and a heterozygous duplication of exons 2-9 of *TRPM1* in trans with a novel missense change (p.Y56C) was found in a second patient with CSNB. A heterozygous 42 exon deletion of *GPR98* was found in trans with a frameshift single nucleotide deletion in an Usher syndrome patient and a heterozygous single exon deletion of *GPR98* was found in trans with a frameshift single nucleotide insertion in a second Usher syndrome patient. Finally, a heterozygous 6 exon duplication was found in trans with a novel missense change (p.V1215A) in *PNPLA6* in a patient with an appropriate phenotype. One duplication was identified using IGV to inspect read orientation of candidate genes. A heterozygous single exon duplication characterised by clusters of everted read pairs flanking exon 4 of *NMNAT1* was found. This was in trans with the common missense mutation p.E257K. **Conclusions** CNV analysis based on read depth and close scrutiny of probable disease causing genes in this cohort revealed deletions or duplications in 6 patients. It is essential to incorporate CNV calling methods such as these in exome analysis since they form a significant proportion of clinically significant mutations, in this cohort 7/108 disease-causing alleles. **Reference** 1- Plagnol V, et al. Bioinformatics. 2012 1;28(21):2747-54.

2977S

Pathogenic mutation of coagulation factor X deficiency may prevent atypical hemolytic uremic syndrome. *F. Bu[1,2], N. Borsa[2], W. Tollefson[3], M. Schnieders[3], H. Azaiez[2], K. Wang[4], C. Thomas[2,5], C. Nester[2,5], R. Smith[1,2,5].* 1) Interdepartmental PhD Program in Genetics, University of Iowa, Iowa City, IA; 2) Molecular Otolaryngology and Renal Research Laboratories, University of Iowa, Iowa City, IA; 3) Department of Biochemistry, University of Iowa, Iowa City, IA, USA; 4) Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, IA; 5) Rare Renal Disease Clinic, Departments of Pediatrics and Internal Medicine, Carver College of Medicine, University of Iowa, Iowa City, IA,.

Introduction: Atypical hemolytic uremic syndrome (aHUS) is an ultra-rare renal disease caused by uncontrolled activation of the alternative pathway of complement at the cell surface. Approximately 22% of patients carry pathogenic mutations in *CFH*, the penetrance of which ranges from 12.5% to 100% in familial studies. We hypothesized that genetic modifiers exist that impact the penetrance of aHUS. **Patients & Methods:** Five families carrying a well-recognized aHUS mutation - c.3644G>A, p.Arg1215Gln - in *CFH* were studied. Using targeted genomic enrichment and massively parallel sequencing, we screened subjects for variants in coding exons of all genes in the complement and coagulation cascades based on the recognized cross-talk between these two pathways. Data were analyzed using locally implemented Galaxy software. Pathogenic variants were annotated based on computational predictions and the literature. The functional impact of select variants was confirmed using in vitro assays. **Results:** We identified a reported factor VII deficiency variant ('Padua' variant; *F7* c.1091G>A, p.Arg364Gln) and a reported factor X deficiency variant (*F10* c.424G>A, p.Glu142Lys) in two families that may impact the penetrance of aHUS. The factor X variant segregated in a three-generation family in which two children under 5 years of age developed aHUS. Neither child carried the *F10* p.Glu142Lys mutation. All older family members, none of whom developed aHUS, segregated BOTH the *F10* p.Glu142Lys variant and the *CFH* p.Arg1215Gln variant. Protein structure simulations show the mutated positively charged Lys142 prefers to hydrogen bond with Ser146 instead of Cys129 (the hydrogen-bonding partner of Glu142), destabilizing an important intra-light-chain interaction between a two-stranded beta-sheet and a small alpha-helical secondary structure element. A conformation change is consistent with in vitro studies showing altered secretion of factor X Lys142. In addition, activity assays demonstrate a 30% reduction in activity of the factor X Lys142 variant. **Summary:** A deficiency mutation in coagulation factor X was identified in healthy carriers of known aHUS *CFH* mutation. The coagulation factor mutation causes hypo-coagulation that is predicted to counter-balance the aHUS phenotype. The consequence may explain some instances of incomplete penetrance. In addition, identification of 'protective' genetic modifiers may offer new therapeutic targets to treat this life-threating disease.

2978M

Genetic testing with targeted exon enrichment and massively parallel sequencing for 272 Chinese cases with hearing loss. *j. cheng[1], Y. Lu[1], Z. Jiang[2], C. Li[2], X. Zhang[1], H. Duan[1], D. Han[1], H. Yuan[1].* 1) Inst Otolaryngology, Chinese PLA Gen Hosp, beijing, China; 2) Center for Genetic & Genomic Analysis, Genesky Biotechnologies Inc., Shanghai, China.

Hereditary non-syndromic hearing loss (NSHL) is extremely heterogeneous. The large reservoir of known deafness genes precluded comprehensive genetic testing and population-scale sequencing. Until recently, the combination of targeted genomic capture and massive parallel sequencing (MPS) has become a promising tool for detecting novel and known mutations involved in hereditary hearing loss. In this study, we aimed to establish Chinese population-level frequencies of reported deafness-causing variants in known genes related hearing loss and to determine the genetic defects for 272 Chinese families and 170 congenital deafness cases. We performed a targeted enrichment that enables the capture of 484 genes included both known regions of the human genome associated with hearing loss and some candidate genes in some important pathways related to hearing. With a single reaction for high throughout mutation identification by massively parallel sequencing, the patient samples were barcoded individually and multiplexed. Mutations in 21 known DFNA genes are identified in 37 families with progressing sensorineural hearing loss, and 22 known DFNB genes are identified in 44 cases with congenital deafness. For the potential candidates, the confirmation by Sanger sequencing and co-segregation analysis has been done. Besides the frequency of mutations in known gene was dissolved, novel genes and mutations are described. Discovery of these new genes will continue to help define compelling mechanisms for deafness.

2979T

Prospecting genetic disorders in a highly inbred region of Brazil: two novel genes for AR intellectual deficiency. *T. Figueiredo[1,2,3], U. Souto[2], A. Pessoa[4], P. Ribeiro[5], A. Brandão[5], M. Zatz[2], F. Kok[2,5], S. Santos[1].* 1) Núcleo de Estudos em Genética e Educação, Universidade Estadual da Paraíba, Campina Grande, Brazil; 2) Centro de Pesquisas sobre o Genoma Humano e Células-Tronco, São Paulo, Brazil; 3) Rede Nordeste de Biotecnologia, RENORBIO, Universidade Federal da Paraíba, João Pessoa, Brazil; 4) Faculdade de Medicina, Universidade de Fortaleza, Fortaleza, Brazil; 5) Faculdade de Medicina, Departamento de Neurologia, Universidade de São Paulo, São Paulo, Brazil.

Intellectual deficiency (ID) is a highly heterogeneous condition affecting 3% of the population worldwide. In a field study conducted in a highly inbred area of Northeastern Brazil, we investigated two large consanguineous families with ID. Genome-Wide Human SNP Array 6.0 (Affymetrix) microarray was used to determine regions of homozygosity-by-descent from 3 affected and 3 normal individuals in each family. Whole exome sequencing (WES) was performed in one affected individual of each family using Nextera Rapid-Capture Exome and Illumina HiSeq2500. Potentially deleterious variants detected in regions of homozigosity-by-descent and not present among 8,000 controls (including 600 Brazilians) were subject to further scrutiny and segregation analysis by Sanger sequencing. **Family A** has 9 affected adults descending from four closely related first-cousin couples affected by severe non-syndromic ID associated to disruptive paranoid behavior. Homozigosity-by-descent analysis disclosed a 20.7 Mb region in 8q12.3-q21.2 (lod score: 3.11). WES identified a homozygous deleterious variant in inositol monophosphatase1 gene (*IMPA1*), consisting of a 5 bp duplication (c.489_493dupGGGCT) leading to frameshift (p.Ser165Trpfs*10). *IMPA1* gene product is responsible for the final step of biotransformation of the second messenger inositol-polyphosphate. Homozygous *Impa1* knockout mice die in utero, but can be rescued by supplementation of myo-inositol. Deficiency of IMPA1 might represent a new inborn error of metabolism, affecting inositol-phosphate biotransformation. **Family B** has 3 adult siblings with severe syndromic ID. Phenotype includes large forehead, prognatism, large cupped ears, mildly downslanted eyes and a very large nose. Additionally, affected male has macropenis. Homozigosity-by-descent analysis found a 4.0 Mb region in 19q13.32-q13.33 (lod score: 3.24). WES disclosed a homozygous variant (c.418C>T, p.Arg140Trp) in mediator complex subunit 25 (*MED25*), predicted as deleterious by Provean and SIFT. MED25 is a component of the Mediator complex, involved in regulation of transcription of nearly all RNA polymerase II-dependent genes. Deleterious mutations in *MED12*, *MED17* and *MED23* have been already associated with ID. These findings demonstrate that the combination of field investigation of large families in highly inbred regions with modern NGS technologies is an effective way to identify new genes and pathways which can be the target of future treatment.

2980S

Molecular Basis of Nieman-Pick A-B and Neimann-Pick C Diseases in the Aegean Region of Turkey: Identification of Three Novel Mutations in SMPD1 Gene. *H. Onay[1], A. Aykut[1], S. Kalkan[2], M. Çoker[2], F. Ozkinay[1,2].* 1) Ege University Faculty of Medicine, Department of Medical Genetics, İzmir, Turkey; 2) Ege University Faculty of Medicine, Department of Pediatrics, İzmir, Turkey.

The Niemann-Pick disease group consists of two different entities: (1) acid sphingomyelinase-deficient forms which are caused by SMPD1 gene mutations and encompasse type A and type B, as well as intermediate forms; (2) Niemann-Pick disease type C is a cellular lipid trafficking disorder caused by NPC1 or NPC2 gene mutations. Niemann-Pick Type A patients present with enlarged livers and spleens, a failure to thrive, and neurological manifestations in the first 6 months of life resuting in death before 3 years of age. Niemann-Pick type B patients have little or no neurological involvement , milder symptoms, and survive into adulthood. Niemann-Pick type C has a broad clinical spectrum ranging from a nonimmune hydrops fetalis to an adult-onset chronic neurodegenerative disease. The aim of this study is to investigate the spectrum of mutations detected in SMPD1, NPC1 and NPC2 genes, in either affected individulas or carriers, analysed in a reference center from the Aegean region of Turkey. Screening for the mutations was performed using the direct sequencing method. There were 42 mutant SMPD1 alleles found with nine different types detected amongst these.The most prevalent mutation was p.L137P with an allelic frequency of 47.6 %. Two mutations, namely p.L137P and p.P189PfsX65, accounted for 69% of the total mutated alleles. Three novel mutations (L161P, W176C, T397M) were detected in the group studied. Among these, L161P was found in 11.9% of the unrelated mutant alleles and was the third most common mutation. One of the patients was compound heterozygous for the novel mutation W176C and previously described I178N mutation. Another patient who had very mild symptoms and low enzyme levels was heterozygous for the novel T397 mutation. Four different mutations were detected in NPC1 gene from among 62 indivudals suspected of having NPC disease. One of the patients was homozygous for c.839delT(p.L280CfsX30) mutation, the other for p.N906Y mutation; and two were heterozygous for G992R and A558T mutations. No NPC2 mutation was detected. The results provided will be useful in genetic counseling and prenatal diagnostic services in our region.

2981M

Frataxin, a Fredrich's ataxia protein is defective in mitochondrial processing peptidase-alpha (*PMPCA*) mutations. *P.B. Agrawal[1], M. Joshi[1], I. Anselm[2], F. Giani[1], M. Towne[1], K. Schmitz-Abe[1], K. Markianos[1], V.G. Sankaran[1].* 1) Department of Medicine, Boston Children's Hospital and Harvard Medical School, Boston, MA; 2) Department of Neurology, Boston Children's Hospital and Harvard Medical School, Boston, MA.

A 4-year-old girl with an unknown mitochondrial disease (MD) had multisystem involvement including cardiomyopathy, hypotonia, respiratory insufficiency, myoclonic jerks and blindness. Her cousin (their fathers are brothers and mothers are first cousins), a boy died at 14 months age due to a similar presentation. A MD was suspected based on clinical and histopathological findings and electron microscopy with abnormally enlarged mitochondria. Respiratory chain enzyme analysis in skin fibroblasts revealed minor reduction in complexes I and II/III activities. Sequencing of the mitochondrial DNA, mitochondrial nuclear gene panel and microarray analysis for copy number variations revealed no abnormality. The family was enrolled in an IRB-approved study and to unveil the molecular basis, whole exome sequencing was performed on the trio. We identified compound heterozygous *PMPCA* (MIM 613036) mutations (c.G1066A; p.G356S and c.G1129A; p.A377T) in the proband. These mutations were also present in the affected cousin and they segregated appropriately in the proband's and cousin's parents. A majority of nuclear and mitochondrial DNA-encoded mitochondrial proteins have N-terminal targeting presequence which is cleaved off after import for optimum function. Mitochondrial processing protein alpha (PMPCA), encoded by *PMPCA*, along with the PMPCB (together called MPP) may play a critical role in this process. This may be especially true for frataxin, a protein altered in Fredrich's ataxia (FA) that needs to be cleaved twice, and studies have implicated MPP in FA. This is the first instance of MPP proteins linked to a human disease. PMPCA is highly conserved with homologs present in yeast and rice. The altered amino acids are highly conserved, with G356 present in yeast and rice, and A377 in drosophila and C. elegans. Polyphen-2, SIFT and Mutationtaster software indicated the mutations to be pathogenic. The amount of PMPCA protein was reduced in the patient compared to age-matched control skin fibroblasts. Further, mature frataxin was markedly reduced in the proband while unprocessed frataxin was increased. We overexpressed wild type *PMPCA* in the patient's fibroblasts using lentiviral transduction and rescued the amounts of processed frataxin thereby confirming that *PMPCA* mutations cause defects in frataxin. Currently, we are using morpholinos to evaluate the effects of PMPCA deficiency in zebrafish models and rescue attempts are underway using wild type and mutant *PMPCA*.

2982T

Genetic diagnosis of mitochondrial disorders in Finland by whole-exome sequencing. *C.J. Carroll[1], P. Isohanni[1,2], V. Brilhante[1], H. Tyynismaa[1,3], E. Palin[1], C. Vasilescu[1], T. Tyni[1], H. Almussa[4], P. Ellonen[4], T. Lönnqvist[2], H. Pihko[2], A. Suomalainen[1,5].* 1) Research Programs Unit, Molecular Neurology, Biomedicum-Helsinki, University of Helsinki, 00290 Helsinki, Finland; 2) Department of Pediatric Neurology, Children's Hospital, Helsinki. University Central Hospital, 00029 Helsinki, Finland; 3) Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki, Finland; 4) Finnish Institute for Molecular Medicine (FIMM), University of Helsinki, Finland; 5) Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland.

Mitochondrial disorders comprise a heterogeneous group of genetic diseases that can manifest in any organ system and have any mode of inheritance. These diseases are caused by mutations in either the mitochondrial genome or in nuclear genes encoding mitochondrial proteins, which makes molecular diagnosis challenging. We report here the results of a whole-exome sequencing approach of a cohort of 68 Finnish patients suspected of having a recessively inherited mitochondrial disease and classified using Bernier criteria. The patient phenotypes included Leigh/Alpers or similar, encephalopathy, cardiomyopathy, infantile lactic acidosis or arPEO. We customized an analysis pipeline, and prioritized rare variants using a unique population variant data resource over 3300 Finnish exomes (Sequencing Initiative Suomi). We identified the genetic background in a total of 29 patients (42%), of whom 20 carried mutations in genes encoding mitochondrial proteins, and nine in non-mitochondrial protein encoding genes, emphasizing the advantageousness of a whole-exome approach instead of targeted sequencing of candidate genes. 24 patients had mutations in known disease genes, of which 15 (22%) presented with known phenotypes previously associated with the mutated gene, thereby enabling a straight forward diagnosis Surprisingly, we did not identify major founder alleles in the Finnish mitochondrial disease population, despite the country being a genetic isolate, but most families had their own specific compound heterozygous gene defects. The few families carrying same defects result in a requirement of extensive functional verification of identified variants, to establish pathogenicity. We conclude that whole-exome sequencing is an effective method for the molecular diagnosis of suspected mitochondrial disorders, with high potential to lead to specific diagnosis early in the diagnostic workup.

2983S

A comprehensive genomic analysis for mitochondrial respiratory chain disorder. *M. Kohda[1], Y. Tokuzawa[2], Y. Kishita[2], Y. Yamashita-Sugahara[2], Y. Mizuno[2], Y. Nakachi[1], Y. Moriyama[2,3], T. Hirata[1], Y. Kanesaki-Yatsuka[2], H. Harashima[4], M. Ajima[5], T. Yamazaki[4], M. Mori[6], K. Murayama[5], H. Kato[1], A. Ohtake[4], Y. Okazaki[1,2].* 1) Div Trans. Res, Res., Center for Genomic Med., Saitama Med Univ, Hidaka, Saitama, Japan; 2) Div. of Func. Genomics & Systems Med., Center for Genomic Med., Saitama Med Univ, Hidaka, Saitama, Japan; 3) Dept. of Anatomy II and Cell Biology, Fujita Health Univ. School of Med; 4) Dept. of Pediatrics, Faculty of Medicine, Saitama Medical University, Saitama, Japan; 5) Dept. of Metabolism, Chiba Children's Hospital, Chiba, Japan; 6) Dept. of Pediatrics, Matsudo City Hospital, Chiba, Japan.

Mitochondria are small organelles and serve as the powerhouse of the living cells because they generate vital energy in their respiratory chain system consisting of protein complexes I to V. Mitochondrial respiratory chain disorder (MRCD) is an intractable disease that develops in childhood. It is a highly frequent inborn errors of metabolism that occurs in at least one out of every 7,000 births. Prominent symptoms develop in such organs as the brain, heart, and muscles, where a great deal of energy is required. In most cases, effective treatment has not been established yet. It is known that various gene abnormalities cause defects of the protein complexes of the respiratory chain, which results in mitochondrial dysfunction. However, identification of the causative genes and the understanding of pathogenic mechanisms of MRCD remain largely unsolved. In this study, we applied long PCR based mtDNA sequencing, high density SNP arrays and whole exome sequencing for comprehensive genomic analysis. Exome sequencing data were filtered using three different criteria: (i) the presence of rare variants in known disease causing genes; (ii) rare variants in genes that code for mitochondrial proteins; (iii) unbiased genome-wide approach with strict filtering strategy. In this study, 144 unrelated individuals were chosen who display juvenile-onset mitochondrial disorders. In 24/144 cases, we identified novel mutations in known disease-causing genes (ACAD9, BOLA3, CO×10 etc). Currently we are validating the rare variants in mitochondria-related candidate genes. In addition, unbiased genome-wide analysis allows us to list up candidates for new disease causing genes thereby extending our understanding of genetic landscape of MRCD. The progress of our precisely controlled strategy will be presented at the meeting.

2984M

Diagnosing mitochondrial disease: Is there an added advantage of whole-exome sequencing? *S.J. Mosca[1], P.M. Gordon[2], L. Dimnik[4], S.T. Nakanishi[2,6], S. Hume[7], D.S. Sinasac[1,5], J.S. Parboosingh[1,2,4], A. Khan[1,2,3].* 1) Department of Medical Genetics, University of Calgary, Alberta, Canada; 2) Alberta Children's Hospital Research Institute, University of Calgary, Alberta, Canada; 3) Metabolic Diseases Clinic, Alberta Children's Hospital, Alberta, Canada; 4) Molecular Diagnostics Lab, Alberta Children's Hospital, Alberta, Canada; 5) Biochemical Genetics Laboratory, Alberta Children's Hospital, Alberta, Canada; 6) Department of Physiology and Pharmacology, University of Calgary, Alberta, Canada; 7) Department of Medical Genetics, University of Alberta, Alberta, Canada.

The diagnosis of mitochondrial disease currently involves multiple invasive procedures, which carry the risk of complications, and in many cases a diagnosis of mitochondrial disease cannot be made or excluded confidently. The Calgary Metabolic Clinic reviewed 292 patients with a clinical suspicion of mitochondrial disease, of which 35% had an eventual diagnosis of mitochondrial disease (MDx), 24% had a diagnosis of non-mitochondrial disease (OtherDx), and 41% had no final diagnosis (NoDx). When muscle-extracted mitochondrial DNA (mtDNA) analysis was performed, a diagnosis was made in 66.7% of cases in the MDx group and none in either of the other groups ($p<0.001$). mtDNA analysis showed the highest specificity compared to muscle histology and enzyme biochemistry. Whole-exome sequencing (WES) was performed on 12 unrelated patients from the NoDx group, who had no mtDNA mutations, and 1 case that had an mtDNA deletion and a suspected causative nuclear mutation. A systematic workflow for the evaluation of variants identified through WES was developed using the Illumina TruSight One gene list as well as a custom mitochondrial disease gene list as a first-line filter in an attempt to identify mutations in known disease-causing genes. Of the 13 patients, the variant prioritization workflow was able to identify the one known cause of mitochondrial disease, and putative disease-causing mutations were identified in five patients (compound heterozygous mutations in TRMU, ATP10A and DGUOK and heterozygous mutations in COL6A1 and NDUFS). Taken together, this data suggests that mtDNA analysis from muscle-extracted DNA produced the highest single diagnostic yield of cases with a mitochondrial disease, and that WES of nuclear DNA provided an answer in an additional 38% of previously unsolved cases of suspected mitochondrial disease with no mtDNA mutations. From this pilot data, we propose that the most efficient model for a genetic diagnosis of mitochondrial disease may be muscle-extracted mtDNA analysis followed by WES if the results are negative. With this approach, we expect to have higher confidence in not only confirming a diagnosis of mitochondrial disease, but also in excluding one and reduce the reliance on invasive procedures.

2985T

Realignment of whole-genome and exome sequencing reads supports novel potassium channel (Kir2.x) isoforms that were formerly identified by Sanger sequencing as polymorphisms of a single channel gene in thyrotoxic periodic paralysis locus. *M.R. Dias da Silva[1], R.M. Paninka[1], I.S. Kunii[1], G.K. Furuzawa[1], M.M.L. Kizys[1], R.M.B. Maciel[1,2], M. Mitne-Neto[2], L.T. Cerdeira[2], S. Pinto[3], H. Rodrigues[3], S. Tufik[3], D.R. Mazzotti[3].* 1) Laboratory of Molecular and Translational Endocrinology, Department of Medicine, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, SP, Brazil; 2) Fleury Group, São Paulo, SP, Brazil; 3) Departament of Psychobiology, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, SP, Brazil.

Thyrotoxic periodic paralysis (TPP [MIM 613239]) is the commonest cause of muscle weakness in adults suffering from hyperthyroidism, characterized by transient hypokalaemia and linked to mutations in the potassium channel *KCNJ18* (Kir2.6 [MIM 613236]) in 33% of patients. While searching for mutations in Kir2.x paralogues we verified that many SNPs have been mistakenly attributable to *KCNJ12* (Kir2.2 [MIM 602323]) due to high homology among family members (95-99% at protein level). The exact positions of *KCNJ17* (Kir2.5) and *KCNJ18* (Kir2.6) remain unclear. Although a BAC containing *KCNJ17* has not been found, both RP11-437N10 (*KCNJ18*) and RP11-728e14 (*KCNJ12*) have been mapped to 17p11.1-2. FISH originally localized *KCNJ12* and *KCNJ17* to 17p11.1. We aimed to identify possible misalignments of whole exome (WES) and whole genome sequencing (WGS) reads in the GRCh37/hg19 reference region of Kir2.2 by performing realignment of those reads to Kir2.2, Kir2.6 and Kir2.5 regions obtained by peripheral blood Sanger cDNA (direct and cloning) sequencing. WGS was performed with paired-end 100bp reads with the Illumina HiSeq2000 at target coverage of 30X. Alignment to the reference genome and variant calling was carried out with a standard BWA/GATK pipeline. We identified a single Kir2.2 and Kir2.5 cDNA, and 3 major Kir2.6 sequences named as RRAI, QHEV and QHAV haplotype based on the amino acid changes at 39, 40, 56 and 249 residue. After inspection of reads aligned to the GRCh37/hg19 genome reference assembly within the pericentromeric region of 17p11.1-2 spanning *KCNJ12* (chr17:21279699-21323179), a notable number of variants were detected, suggesting misalignment of reads due to low coverage of the reference assembly at that region. After realignment to the sequence identified by Sanger, we found 1992 and 613 reads aligned to that in the WES and WGS. Among WES reads, 706 (35.4%), 750 (37.6%) and 536 (26.9%) aligned to Kir2.6, Kir2.5 and Kir2.2, respectively. Similarly, in WGS, 158 (25.8%), 190 (31.0%) and 295 (43.2%) aligned to Kir2.6, Kir2.5 and Kir2.2. Taking together mapping and functional data, our findings support the presence of novel isoforms of Kir channels not adequately mapped to the reference genome. Although pericentromeric duplications are thought to contain heterochromatic DNA and have fewer expressed genes, we have presented several lines of evidence suggesting that Kir2.x paralogues are distinct and may play a role in muscle plasticity as observed in TPP.

2986S

Novel SLC29A3 mutation causing H Syndrome in an Indian Adolescent. *N. Kamath[1], R.D. Shenoy[2], M. Varma[3], R. Khubchandani[4], S. Babay[5], A. Zlotogorski[5,6], V. Molho Pessach[5,6].* 1) Department of Pediatrics, Kasturba Medical College, Manipal University, Mangalore, Karnataka, India; 2) Department of Pediatrics K.S. Hegde Medical Academy, Nitte University, Mangalore, India; 3) Private Rheumatology Clinic, Mumbai, India; 4) Pediatric Rheumatology Clinic, Jaslok Hospital and Research Center, Mumbai, India; 5) Center for Genetic Diseases of the Skin and Hair, Hadassah-Hebrew University Medical Center, Jerusalem, Israel; 6) Department of Dermatology, Hadassah-Hebrew University Medical Center, Jerusalem, Israel.

Purpose: H syndrome (OMIM 6027828) is an autosomal recessive genodermatosis with systemic manifestations due to biallelic mutations in the SLC29A3 gene. The major clinical findings include hyperpigmentation, hypertrichosis, hearing loss, heart anomalies, hepatospleenomegaly, hypogonadism, low height (short stature), hallux valgus/flexion contractures and hyperglycemia/diabetes mellitus. It has recently been recognized as an inherited form of Histiocytosis. Genetic analysis of a case with clinical features consistent with H syndrome was performed in search for the causing mutation. **Methods**: A 16-year-old Indian adolescent, born out of non consanguinous parentage was referred due to short stature, sensorineural hearing loss, insulin dependent diabetes mellitus, delayed puberty and diffuse cutaneous hyperpigmentation. On physical examination she was noted to have height and weight below the third centile in the CDC 2000 growth charts, bilaterally symmetrical indurated sclerodermatous hyperpigmented patches with accompanying hypertrichosis overlying the trunk and lower extremities with sparing of the knees and buttocks, prepubertal sexual maturity rating, hepatosplenomegaly, dilated lateral scleral vessels, arcus senilis, infiltrated cheeks, lateral tibial torsion and hallux valgus. Laboratory evaluation revealed microcytic anemia, highly elevated ESR and hypogonadotropic hypogonadism. DNA was extracted from peripheral blood and analyzed for SLC29A3 mutations by Sanger sequencing. **Results**: We identified a novel misesnse homozygous mutation in the fifth exon of SLC29A3: c.677G>A (G209R). This variant was absent in dbSNP (http://www.ncbi.nlm.nih.gov/SNP), the 1000 Genomes project data (http://www.1000genomes.org/) and the Exome Variant Server (http://evs.gs.washington.edu/EVS). The mutation was predicted to be disease-causing by MutationTaster with a score of 0.99 (http://www.mutationtaster.org/). **Conclusions**: We demonstrated a novel homozygous mutation in the SLC29A3 gene in an adolescent female of Indian origin with clinical features consistent with H syndrome. Our patient emphasizes the pathognomonic cutaneous phenotype as well as the various extra cutaneous features of this autosomal recessive pleomorphic disorder.

2987M

An intergenic 9.4 kb microduplication at chromosome 5p13 as a cause of brachydactyly type A1. *L. Racacho[1], S.M. Nikkel[2], J. MacKenzie[3], C.M. Armour[2], M.E. McCready[4], Y. De Repentigny[5], R. Kothary[5], L.A. Pennacchio[6,7], D.E. Bulman[1,2,8].* 1) Department of Biochemistry, University of Ottawa, Ottawa, ON, Canada; 2) Department of Pediatrics, University of Ottawa, Department of Genetics, CHEO Research Institute, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 3) Department of Medical Genetics, Kingston General Hospital, Queen's University, Kingston, ON, Canada; 4) Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada; 5) Regenerative Medicine Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada; 6) Genomics Division, Lawrence Berkeley National Laboratories, Berkeley, CA, USA; 7) US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA; 8) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada.

Brachydactyly type A1 (BDA1 [MIM 112500]) is an autosomal dominant brachymesophalangeal trait, primarily characterized by a hypoplastic and/or aplastic pattern of the middle phalanges of digits 2-5 in the hands and feet. Missense mutations in either *IHH* or *GDF5* have been associated with BDA1. We previously reported the linkage of a BDA1 locus to chromosome 5p13 (BDA1B [MIM 607004]) in a single large family. The PCR-based sequencing of all the protein coding genes within the 3 Mb critical region did not reveal a mutation. We did not identify any large genomic rearrangements spanning the critical region with either a 3-color FISH or a 500K SNP array. In order to provide a higher sensitivity of mutation detection, we performed a targeted hybridization enrichment of the BDA1B locus on two affected family members followed by high-throughput sequencing. The alignment of the sequence reads from both individuals to the reference genome revealed a novel and shared 9.4 kb intergenic tandem duplication. The microduplication falls within a 'gene desert' consisting of several vertebrate conserved sequence blocks. We were able to demonstrate the enhancer activity of these conserved blocks through the use of an *in-vivo* transgenic reporter assay. Relative quantitative PCR on the patient's fibroblast cDNA showed an up-regulation of distal genes when compared to unaffected controls. Our findings suggest that a *cis*-regulatory mutation is most likely involved in the pathology of BDA1 in these two individuals.

2988T

Novel molecular insights into severe congenital microcephaly through targeted next generation sequencing. *G. Mirzaa[1,2], V. Vasta[1], S. Christian[1], Q. Zhang[3], S. Eun[4], S. Collins[1], S. Hahn[1,2], W. Dobyns[1,2].* 1) Human Genetics, Seattle Children's Research Institute, Seattle, WA; 2) Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, WA; 3) Fred Hutchinson Cancer Research Center; 4) Department of Pediatrics, Korea University School of Medicine, Seoul, Korea.

Microcephaly (MIC) is a developmental brain growth disorder that accounts for a significant fraction of childhood intellectual disability, autism and epilepsy. It occurs in isolation (primary MIC) and as part of a broad range of neurodevelopmental syndromes with or without other cortical malformations. MIC is genetically very heterogeneous and the rate of identification of novel MIC genes has increased dramatically with the advent of next generation sequencing (NGS). The majority of MIC genes are key regulators of critical processes including mitotic spindle assembly and structure, centrosome function, and DNA repair and damage response (DDR) pathways. We screened a cohort of 93 individuals with MIC using a targeted NGS panel that contains 377 known and candidate MIC genes. Most individuals had severe congenital or postnatal MIC (head circumference 3 standard deviations or more below the mean) with or without otherstructural brain malformations, such as polymicrogyria, cerebellar and callosal abnormalities. We identified mutations in 26/93 (28%) of patients. Recessive mutations identified in known MIC genes include ASPM (N=7; 7.5%), CENPJ (N=1), CEP135 (N=1), CASC5 (N=1), WDR62 (N=1), ASNS (N=2), RBBP8 (N=1), ADSL (N=1) and RAB3GAP2 (N=1). Heterozygous de novo mutations were identified in TUBA1A (N=2), DYNC1H1 (N=2), TUBB2B (N=1), DYRK1A (N=1), FOXG1 (N=1), MLL2 (N=1), CASK (N=1), and HCCS (N=1). Our cohort expands a number of phenotypes including HCCS-associated features in males to include MIC, short stature and micropenis without severe eye defects, RBBP8-associated Jawad syndrome in a Pakistani family with MSG but without digital anomalies, adenylosuccincate lyase deficiency (ADSL) in a large sibship with MSG and increased extra-axial space, and Warburg Micro syndrome caused by mutations RAB3GAP to include postnatal MIC, hypotonia, and severe developmental encephalopathy without eye involvement. Our cohort also introduces new reports of rare phenotypes including asparagine synthetase (ASNS) mutations causative of congenital MIC and progressive epileptic encephalopathy with early childhood death, DYNC1H1 mutations causative of bilateral cortical dysplasia (specifically posterior-predominant pachygyria), CEP135 mutations with features of microcephalic osteodysplastic primordial dwarfism, CENPJ with MIC, seizures and growth deficiency, and DYRK1A associated primary microcephaly, epilepsy, severe intellectual disability and growth retardation.

2989S

Molecular genetic characterization of an autosomal recessive Familial Essential Tremor. *D. Monies[1], E. Naim[1], B. Al-Younes[1], M. Al-Breacan[1], M. Al-Saif[1], S. Wakil[1], K. Khabar[2], B. Meyer[1], S. Bohlega[3].* 1) Genetics, RC, King Faisal Specialist Hospital & Research Centre, Riyadh, Saudi Arabia; 2) Biomolecular Medicine, RC, King Faisal Specialist Hospital & Research Centre, Riyadh, Saudi Arabia; 3) Neurosciences, King Faisal Specialist Hospital & Research Centre, Riyadh, Saudi Arabia.

Despite the presence of strong family histories the underlying genetics of familial essential tremor (FET) is not well defined. Although several loci associated with FET have been described, there exists only relatively poor linkage data and no causative mutations having been identified to date. This far the mode of inheritance for FET is best described as autosomal dominant with incomplete penetrance. We describe the molecular genetic characterization of an autosomal recessive familial disorder characterized by tremor of juvenile onset, dystonia and myoclonus with preserved cognitive, cerebellar and peripheral nervous system functions. Mild spasticity appeared with disease progression and some white matter changes were evident upon MRI. Linkage analysis of a consanguineous family with five affected individuals identified a locus on chromosome 17 with a LOD score >4.0. Homozygozity mapping confirmed a single homozygous region shared by all affected individuals only and consistent with the locus identified by linkage analysis. Whole exome sequencing was performed on 1 affected individual with variants being filtered based upon the linkage interval, absence in variant databases at a frequency >0.1%, presence in coding/flanking regions and homozygozity. A single candidate causative variant segregating with disease in the family was identified. We describe further the functional validation of this variant. This is the first description of FET inherited in an autosomal recessive manner and offers opportunities for the further investigation of tremor associated with other disorders including Parkinsons disease.

2990M

Denovo mutations in a novel disease causing gene cause Temple-Baraitser syndrome and non-syndromic epilepsy. *C. Simons[1], L. Rash[1], J. Crawford[1], K. Ru[1], S.M. Grimmond[1,2], D. Miller[2], G. King[1], J. McGaughran[3,4], M. Gabbett[3,4], R.J. Taft[1,5,6].* 1) Institute for Molecular Bioscience, University of Queensland, Jindalee, QLD, Australia; 2) Institute of Cancer Sciences, Translational Research Centre, University of Glasgow, Scotland; 3) Genetic Health Queensland, Royal Brisbane & Women's Hospital, Australia; 4) School of Medicine, The University of Queensland, Australia; 5) School of Medicine and Health Services, Departments of Integrated Systems Biology and of Pediatrics, George Washington University, USA; 6) Illumina, Inc., San Diego, CA USA.

Temple-Baraitser syndrome (TBS) is a multi-system developmental disorder characterized by intellectual disability, epilepsy and hypo- or aplasia of the nails of the thumb and great toe. TBS is rare, with only five cases reported to date, although it is possible that that TBS is largely unrecognized and therefore under-reported. We recruited six unrelated TBS individuals and their parents for analysis by whole exome sequencing. Here we report damaging de novo mutations in a novel disease causing gene that is predominantly expressed in the central nervous system, in six TBS individuals. Functional characterization of two variants revealed that TBS-associated mutations lead to deleterious gain-of-function. Consistent with this result, we found that two TBS mothers with epilepsy, but who were otherwise healthy, were low level (12% and 30%, respectively) mosaic carriers of pathogenic mutations. This suggests that the etiology in some cases of non-syndromic epilepsy may be explained by low level mosaicism for mutations in this gene.

2991T

High diagnostic success rate in a cohort of unresolved leukoencephalopathy patients investigated by whole exome sequencing. *R.J. Taft[1,2,3], G. Helman[4], A. Pizzino[4], A. Vanderver[4], C. Simons[1].* 1) Illumina Inc., San Diego, CA; 2) Institute for Molecular Bioscience, University of Queensland, St Lucia, Australia; 3) School of Medicine and Health Services, The George Washington University, Washington DC, USA; 4) Children's National Medical Center, Washington DC, USA.

Leukoencephalopathies are a heterogeneous group of heritable central-nervous system white matter disorders that are primarily characterized by abnormal signal on brain magnetic resonance imaging. Leukoencephalopathies remain a diagnostic challenge for medical practitioners, with nearly 50% of cases remaining unresolved despite recent advances in both radiological and biochemical testing. We performed whole exome sequencing (WES) on a cohort of 85 patients with unsolved leukoencephalopathies, who were prospectively collected by the Myelin Disorders Bioregistry Project, with non-specific radiologic findings or previously negative diagnostic testing. WES and downstream bioinformatics analysis was performed on trio or greater family groups in all cases, and Sanger sequencing was used to validate and confirm segregation of all candidate pathogenic mutations. Resolution of more than 40% of cases was achieved and included mutations in a wide variety of known and novel disease genes including those associated with mitochondrial cytopathy, epileptic encephalopathy, hypomyelinating leukodystrophies and, unexpectedly, ataxia telangiectasia and dyskeratosis congenita. Overall, fewer than a quarter of the cases solved were due to mutations in genes classically associated with leukoencephalopathies. This study strongly suggests that clinical deployment of agnostic and untargeted high-throughput genomic screening approaches in the leukoencephalopathies will substantially reduce the number of unresolved cases and reduce the burden of the diagnostic odyssey.

2992S

Causal mutations and unique variants identified by exome analysis of Pakistani pedigrees with retinal degeneration. *R. Ayyagari[1], B. Maranhao[1,2], P. Biswas[1], M. Navani[1], J.J. Suk[1], F. Kabir[3], M.A. Naeem[4], J. Chu[1,2], S.N. Khan[4], S. Riazuddin[4], P. Lee[1], S.A. Riazuddin[3,4].* 1) Ophthalmology, University of California, San Diego, La Jolla, CA; 2) Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA; 3) The Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD, 21287, USA; 4) National Centre of Excellence in Molecular Biology, University of the Punjab, Lahore 53700 Pakistan.

Purpose: To define the molecular basis of retinal degeneration in consanguineous Pakistani pedigrees and identify variants unique to the Pakistani population. Methods: A cohort of 26 pedigrees with 269 individuals of Pakistani origin was enrolled. Exomes of selected individuals were captured with Agilent or Nimblegen kits and sequenced on an Illumina HiSeq 2500. Variant call format (VCF) files were generated using standard protocol and subsequently analyzed with exomeSuite to identify candidate variants. Selected variants were examined for segregation with the disease phenotype in their respective pedigrees and their population frequency was investigated in ethnicity matched control chromosomes. All single nucleotide changes displaying no strand bias during sequencing, exhibiting a read depth greater than 10, and mapping quality greater than 30 were further tabulated by pedigree. Results: We identified six novel homozygous mutations in USH2A, CNGB1, FAM161A, CERKL and GUCY2D along with three previously reported causative mutations in RPE65, GUCY2D and LCA5 as the underlying cause of RD in 13 pedigrees. Additionally, we identified more than 2000 single nucleotide variants in the Pakistani cohort that are absent in the HapMap, 1000 Genome and NHLBI ESP6500 databases. Of these, 612 variants present in five or more pedigrees were not reported in the dbSNP database. Conclusions: Here, we identify six novel and three previously reported causative mutations in genes known to be associated with retinal degeneration in 13 of 26 consanguineous pedigrees of Pakistani origin. Additionally, we detected 612 single nucleotide variants that occur in multiple Pakistani pedigrees but are not present in the dbSNP database, thus suggesting that these variants are unique to Pakistani population.

2993M

Identifying the underlying cause of Retinal Degeneration by Exome Sequencing in seven unrelated pedigrees. *P. Biswas[1], B. Maranhao[1], I. Kozak[2], M. Parke[3], K. Branham[4], L. Ganriel[5], J. Heckenlively[4], A. S. Riazuddin[6], J. Duncan[7], R. Ayyagari[1].* 1) Ophthalmology, University of California San Diego, San Diego, CA; 2) Ophthalmology, King Khaled Eye Specialist Hospital, Umm Al Hamam Algharbi, Riyadh, Saudi Arabia; 3) Cornell University, Ithaca, NY, United States; 4) Ophthalmology & Visual Sciences, University of Michigan Kellogg Eye Center, Ann Arbor, MI, United States; 5) Genetics and Ophthalmology, Genelabor , GOIANIA, Brazil; 6) Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, MD, United States; 7) Ophthalmology, University of California San Francisco, San Francisco, CA, United States.

Purpose: To determine the molecular basis of disease in patients with autosomal recessive retinal degeneration. Methods: Ophthalmic evaluation was performed on all affected members and some unaffected relatives. Family history, medical history and blood samples were collected from all the available members and DNA was isolated using standard protocols. Exomes of probands were captured using Nimblegen V3/V2 or Agilent V5+UTR kits, and sequencing was performed on Illumina HiSeq. Reads were mapped against hg19, and analyzed using GATK algorithms. Variants were analyzed by exomeSuite using an autosomal recessive model. Segregation and ethnicity matched control sample analyses were carried out by dideoxy sequencing. Result: Seven unrelated pedigrees with a total of 4-10 individuals and 1-3 affected members were recruited for the study. Analyzing variants using exomeSuite along with stringent filtering criteria revealed 7 novel and 4 reported mutations. A novel homozygous missense mutation was found in *ARL6* in a Middle Eastern family. A novel homozygous frame shift mutation causing premature truncation of the *CERKL* protein was found in a pedigree of Indian origin. A previously reported homozygous mutation was detected in the *ABCA4* gene in a Pakistani pedigree. Recessive compound heterozygous causative variants were observed in the remaining four Caucasian pedigrees. One of these had a novel damaging missense and a novel frame shift variant in the *RPE65* gene. In a second Caucasian pedigree, two heterozygous nonsense changes (novel and reported) in *PDE6B* were detected. In two additional pedigrees causative changes were detected in the *USH2A* and *ABCA4* genes. A novel nonsense and a novel damaging missense variants were detected in the *USH2A*. A missense mutation and a second intronic variant reported to be associated with RD were detected in the *ABCA4* gene in a Caucasian pedigree. All these variants segregated with disease and the novel variants were not detected in ethnicity-matched controls. Conclusions: Exome capture and sequencing revealed seven novel and four known variants as the underlying cause of retinal degeneration in seven unrelated pedigrees.

2994T

Trio-based exome sequencing approach to identify candidate genes for phenotypic variability of Incontinentia pigmenti. *F. Fusco[1], M.I. Conte[1], M. Paciolla[1], M.B. Lioi[2], M.G. Miano[1], M.V. Ursini[1].* 1) Institute of Genetics and Biophysics 'Adriano Buzzati-Traverso', IGB-CNR, Naples, Italy; 2) University of Basilicata, Potenza 85100, Italy.

Incontinentia pigmenti (IP, MIM308300, 1/10.000) is an X-linked dominant neuroectodermal disorder. The skin defects are always present, while the extracutaneous manifestation (ocular, dental, hair, nail and central nervous system-CNS- defects) may occur at variable frequency. In 30% of IP patients CNS anomalies (seizures, encephalopathy, encephalomyelitis, ischemic stroke) are reported. IP patients carry a mutation in the IKBKG/NEMO gene (IKBKG/Nuclear Factor kappaB, Essential MOdulator) that encodes for NEMO/IKKgamma regulatory protein of the IKK complex, required for the activation of the canonical NF-kappaB pathway. We collected a large cohort of IP patients with an high variability of clinical phenotype. A variable CNS defect was observed, even in IP patients with an identical NEMO mutation. The NEMOdel4_10 deletion, identified in 72% of IP patients, can be associated with a wide range of CNS symptoms also in the same IP family. The skewed X-chromosome inactivation could only partially explain this variability thus modifier loci may contribute to the severity of IP phenotype. Here we present a trios-based exome sequencing approach to identify candidate genes for phenotypic variability of IP. We selected three severe IP cases carrying the NEMOdel4_10 deletion, one case with extreme intra-familial phenotypic variability: in the mother only skin defects were present, in the children also a severe mental retardation with neuromuscular defects were reported; and two sporadic cases. We performed whole exome-sequencing of trios samples. From sequencing of an exome-enriched library a list of single nucleotide/indels variants was produced. By applying combined filtering method we select SS/I variants and NS variants predicted to be damaging by polyphen software. We sorted all variants by two model of inheritance: de novo dominant, and recessive (homozygous, compound heterozygous). We will present a list of candidate genes and their associated pathways prioritizing variants fitting to recessive model of inheritance.The most representative category of alterated genes is the methabolic pathway in which we identified two genes acting as modifiers in known diseases. Trio-based exome sequencing has proven to be a powerful approach for identifying candidate genes that may contribute to the phenotypic variability of IP. The identification of modifier genes of IP will be useful to anticipate the outcome of the IP disease in order to apply an early targeted therapy.

2995S

NEMO Deficiency: Mutation in 5' leader sequence of *IKBKG* causes adult onset mycobacterial skin disease. *A.P. Hsu, G. Uzel, S.M. Holland.* Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD.

Uniallelic hypomorphic mutations of X-linked *IKBKG*, encoding the nuclear factor-κB essential modulator (NEMO), lead to ecdodermal dysplasia with immunodeficiency (MIM 300248). Onset of infections in affected males is typically in infancy or childhood, with *Pneumocystis jirovecii*, pyogenic bacteria and mycobacteria. Severe systemic infections cause high mortality during the first decades of life. We identified 3 patients who presented as adults with disseminated *Mycobacterium avium* complex (MAC) and granulomatous skin disease, but limited prior infectious histories. Patient A.I.1 had *Haemophilus influenzae* pericarditis at age 5 then diffuse cutaneous MAC at age 34. Full-length cDNA analysis of the major isoform showed a 110 base pair (bp) deletion in the 5' leader sequence due to a G>C transversion at the last base of the first exon, c.1-16G>C, weakening the consensus splice site and leading to utilization of a cryptic site 110 bases 5' to the end of the exon. Patient B.II.1 had the same change (c.1-16G>C) in *IKBKG* with similar consequence. He had *Mycobacterium marinum* at age 6 and then disseminated MAC at age 27. His maternal cousin, B.II.3 had recurrent respiratory infections in childhood and then developed disseminated MAC at 31 years. Intracellular staining of patient lymphocytes showed <50% of normal NEMO protein. Patients A.I.1 and B.II.1 were lymphopenic (total lymphocyte counts <800), with low naïve CD4+ and CD8+ cells as well as few CD20+/CD27+ memory B cells. Stimulation of PBMCs from patient A.I.1 showed decreased production of TNFα in response to LPS. Bioinformatic analysis revealed the deletion is wholly contained within a conserved 153 bp upstream open reading frame (uORF). This deletion causes a frameshift of the uORF leading to 8 original and 5 missense codons in frame with the initial coding ATG. The uORF ATG has a stronger predicted Kozak sequence than the canonical ATG, possibly leading to a 13 residue N-terminal extension of the protein. Luciferase constructs are being used to determine whether reduced NEMO protein is due to a shortened 5' leader sequence, loss of the uORF, or the 13 additional residues. These late onset mycobacterial infections are caused by mutations in *IKBKG* that affect NEMO protein levels. The identification of similar mutations in 2 unrelated kindreds suggests that 5'UTR *IKBKG* mutations should be sought in late onset MAC infections.

2996M

Inherited UNC13D or PRF1 Mutations in patients with PTLD and severe HHV viremia after HSCT. *H. Liu[1], Y. Zhang[1], F. Wang[1], W. Teng[1], X. Chen[2], P. Zhu[2], C. Tong[1], T. Wu[1], Y. Zhao[1], J. Zhang[1], D. Lu[1].* 1) Medical Laboratory Division, Ludaopei Hematology & Oncology Center, Beijing, China; 2) Department of Hematology, Peking University First Hospital, China.

Background: Post-transplant lymphoproliferative disorder (PTLD) is a rare but life-threatening disease after hematopoietic stem cell transplantation (HSCT) and 90%are EBV positive. Severe human herpes virus (HHV) viremia is the more common form of HHV activation after HSCT. Familial hemophagocytic lymphohistocytosis (FHL) is a mainly autosomal recessive inherited disorder and always triggered by HHV (esp. EBV), the most common mutated genes are UNC13D and PRF1. **Methods and Cases:** To identify the association of genetic defects and HHV disease after HSCT, 32 acute myeloid or lymphoblastic leukemia patients underwent HSCT were enrolled, including 7 PTLD cases and 25 suffered from severe HHV viremia. Genomic DNA was obtained from peripheral blood of the patients before and after HSCT and their donors. Qualitative and quantitative analysis of HHV in peripheral blood was performed by PCR. Molecular genetic studies were performed by direct sequencing of all coding exons and flanking sequences of UNC13D and PRF1. **Results:** 6 paired cases carried UNC13D or PRF1 mutations. Including 4 cases with monoallelic mutations in the recipients' blood sample after HSCT and their donors, one with PRF1 p.P22RfsX29, one with UNC13D p.L202M and two with UNC13D c.2553+5C>G mutations. In the other 2 cases, mutations were inherited carried by the patients' own somatic cells, one with UNC13D p.T1045M/c.2553+5C>G biallelic mutations, and the other with UNC13D c.2553+5C>G mutation.**Conclusion:** UNC13D and PRF1 mutations might be the inherited predisposing factors of severe HHV viremia or PTLD after HSCT.

2997T

Exome sequencing of a family with Wiskot-Aldrich syndrome reveals a mutation in the WIPF1 gene. *A. Hawwari[1], M. Dasouki[1], L. Al Baik[1], S. Al-Hiss[1], O. Khier[1], H. Al-Mousa[1, 2, 3], R. Arnaout[2].* 1) Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 2) Department of Pediatrics, King Faisal Specialist Hospital & Research Center, Riyadh, Saudi Arabia; 3) Al-Faisal University, Riyadh, Saudi Arabia.

This Family we are reporting in this abstract is the first family with WIP deficiency in Saudi Arabia and it would be the second WIP deficiency case reported in the literature. This family originates from the south province of Saudi Arabia, highly consanguineous, with more than fourteen members reported to have clinical presentation suggestive of WIP deficiency. Most of whom had fatal courses. We had the opportunity to evaluate five of these patients from this family. The first case was a boy which presented more than 10 years ago with low platelets, recurrent infection and eczema. He was initially diagnosed as Wiskot-Aldrich case, except that his platelets were of normal size (normal MPV). The Patient underwent stem cell transplantation successfully. Two years later, a second degree female cousin of the first case presented to our hospital with low platelets also normal MPV, and fungal brain infection. Being a female, we started questioning the diagnosis of WAS and WIP deficiency, theoretically, was a logical alternative diagnosis. However, there were no reported human cases of WIP deficiency then. Over the following years, 3 more members of the family all females presented to us soon after birth all with low normal sized platelets, and severe CMV infection in the first months of life. All underwent stem cell transplantations, two survived transplant and the third, who is a sister of our second case, died post transplantation from CMV infection and post transplant complications. We decided to perform Whole Exome Sequencing of two sisters along with their parents which revealed a nonsense mutation (Q237X) in the WIPF1 gene. Full molecular evaluation of many of members of this family confirmed our long suspected diagnosis of WIP deficiency.

2998S

Erythroid Krüppel-like factor mutations are relatively more common in a thalassemia endemic region and ameliorate the clinical and hematological severity of β-thalassemia. *X. Xu[1], D. Liu[1], X. Zhang[2], L. Yu[1], R. Cai[3], X. Ma[1], C. Zheng[4], Y. Zhou[5], Q. Liu[6], X. Wei[1], L. Lin[1], T. Yan[3], J. Huang[1], N. Mohandas[7], X. An[7].* 1) Southern Medical University, Guangzhou, Guangdong, China; 2) Department of Hematology, 303rd Hospital of the People's Liberation Army, Nanning, China; 3) Department of Birth Health and Heredity, Liuzhou Women and Children Care Hospital, Liuzhou, China; 4) Prenatal Diagnostic Center, Guangxi Zhuang Autonomous Region Women and Children Care Hospital, Nanning, China; 5) Department of Birth Health and Heredity, Zhuhai Women and Children Care Hospital, Zhuhai, China; 6) Key Lab for Experimental Teratology of the Ministry of Education and Department of Medical Genetics, School of Medicine, Shandong University, Jinan, China; 7) Red Cell Physiology Laboratory, New York Blood Center, New York City, NY, United States.

The erythroid transcriptional factor 1 (KLF1) has recently emerged as one of the key regulators of the γ- to β-globin gene switching. Mutations in human KLF1 have recently been reported to be responsible for increased fetal hemoglobin (HbF) and hemoglobin A2 (HbA_2). As increased HbF and HbA_2 levels are important features of β-thalassemia, we examined whether there is any relationship between KLF1 mutation and β-thalassemia in China. For this, we first studied the incidence of *KLF1* mutations in two Chinese populations: 3839 individuals from a thalassemia endemic region in south China and 1190 individuals from a non-thalassemia endemic region in north China. Interestingly we found that the prevalence of *KLF1* mutations is significantly higher in thalassemia endemic region than that in non-thalassemia endemic region (1.25% versus 0.08%). Furthermore, a total of 64 mutant alleles were documented in the present study, of which 41 are *KLF1* heterozygotes alone, 11 are *KLF1* mutations co-inherited with β-thalassemia heterozygotes, and 12 are *KLF1* mutations co-inherited with β-thalassemia homozygotes or compound heterozygotes. We identified seven functional variants including four previously reported (p.Gly176AlafsX179, p.Ala298-Pro, p.Thr334Arg and c.913+1G>A) and three novel ones (p.His299Asp, p.Cys341Tyr and p.Glu5Lys) in southern China. The two most common mutations, p.Gly176AlafsX179 and p.His299Asp, accounted for 90.6% of the total. This study also provides an overview of the important hematological features of individuals with *KLF1* heterozygous mutations. In addition, we found it surprising that two KLF1 zinc-finger mutations were selectively represented in 12 β-thalassemia intermedia patients, who were identified from 922 β-thalassemia patients with known β-globin genotypes and various modifier genes. Both multivariate and univariate analysis showed that the modifier of *KLF1* mutations had an effect on amelioration of the severity in β-thalassemia. Our findings suggest that *KLF1* mutations occur selectively in the presence of β-thalassemia to increase the production of HbF which in turn ameliorates the clinical severity of β-thalassemia. The knowledge gained from this study should help in clinical accurate diagnoses and genetic counseling, as well as enable designing of appropriate and personalized transfusion program for thalassemia patients with *KLF1* mutations.

2999M

Homozygous loss of *DIAPH1* causes a rare, complex syndrome with epilepsy, blindness, immune deficiency and lymphoma. *M. Kaustio[1], R. Hinttala[2,3,4], H. Almusa[1], M. Renko[2,3], H. Tuominen[5], R. Herva[5], J. Uusimaa[2,3,4], J. Saarela[1].* 1) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 2) Institute of Clinical Medicine, Department of Pediatrics, University of Oulu, Oulu, Finland; 3) Department of Pediatrics, Oulu University Hospital, Oulu, Finland; 4) Medical Research Center, University of Oulu and Oulu University Hospital, Oulu, Finland; 5) Department of Pathology, Oulu University Hospital and University of Oulu, Oulu, Finland.

Using whole-exome sequencing (WES), we investigated a Finnish family with two affected male siblings, who presented with a novel, complex syndrome including an early-onset intractable epilepsy, blindness, severe mental retardation, deficiency in cellular immunity and B-cell lymphoma. Muscle biopsy showed atrophy, mild increase of fat, a few SDH-COX negative muscle fibres, and an increased number of mitochondria suggesting a mitochondrial disease. Brain MRI and spectroscopy revealed focal occipital leukoencephalopathy. The disease course was progressive leading to death in both siblings at the age of 5 and 18 years. The neuropathologic examination showed severe gliosis, atrophy and vascular abnormality affecting the primary visual cortex. Through WES we identified a rare, homozygous splice donor variant (NM_005219: c.684+1G>A) in the gene *DIAPH1*, which encodes for the mammalian diaphanous-related formin mDia1. Capillary sequencing of reverse transcribed patient RNA with amplicons spanning the mutated site showed that the mutation results in the use of several different cryptic splice sites, all of which ultimately result in the introduction of a premature stop codon in the Drf/GBD domain and likely cause lack of any functional protein. The absence of a protein product in patient cells or tissues was confirmed by western blotting and immunostaining for mDia1. mDia1 acts downstream of Rho GTPases promoting actin polymerization and microtubule stabilization, and it is known to play a role in various processes such as cell migration and tumour metastasis, neuronal development, mitochondrial trafficking and the function of several types of immune cells. Hence, the several aspects of mDia1 function seem to explain most if not all of the features of the phenotype observed in the affected individuals. The *DIAPH1* splice donor variant is not present in databases such as the 1000 Genomes, the Exome Variant Server or dbSNP, but instead was found in the Sequencing Project Suomi (SISu) database with exome sequencing data from about 3300 Finnish individuals with a frequency of 0.075%. This led us to hypothesize that this variant may be enriched in the Finnish population and hence, we will further assess the prevalence and distribution of this putative novel disease variant in Finland by using genotyping data from ~11400 Finns. Furthermore, functional studies addressing the effect of this mutation on mitochondrial function are also ongoing.

3000T

Copy number variations in a cohort of Brazilian sickle cell anemia patients with and without cerebrovascular accident. *P.RS. Cruz[1], G. Ananina[1], F. Menaa[1], A.S. Araujo[2], G.P. Gil[1], W.M. Avelar[3], F. Cendes[3], F.F. Costa[4], M.B. Melo[1].* 1) Genetics and Molecular Biology, University of Campinas, Campinas, São Paulo, Brazil; 2) Hematology and Hemotherapy Center of Pernambuco/HEMOPE, Recife, Pernambuco, Brazil; 3) Neuroimaging Laboratory, Department of Neurology, University of Campinas, Campinas, São Paulo, Brazil; 4) Hematology and Hemotherapy Center/HEMOCENTRO, University of Campinas, Campinas, São Paulo, Brazil.

Although sickle cell anemia (SCA) results from homozygosity for a single mutation at position 6 of gamma-hemoglobin locus, this disease presents high heterogeneity in phenotype, so that different patients may have significantly different clinical outcomes. Virtually all organs may be affected by vascular occlusion, with emphasis on the Central Nervous System, where are observed transient ischemic attacks, stroke and cerebral hemorrhage, which affect approximately 25% of patients with SCA. Early identification of patients with sickle cell anemia, susceptible to stroke (CVA) could reduce the risk, possibly preventing the recurrence of infracts and potentially reducing their incidence. Thus, we propose to investigate the presence of copy number variation in alleles or "Copy Number Variation" (CNV), using high density microarray (Genome-Wide Human SNP Array 6.0) in order to identify genomic regions potentially involved in the increased risk of stroke in sickle cell patients. Among individuals evaluated by microarray so far, 32 are categorized as non-affected by stroke, while 39 have reported at least one episode in their medical history. An exploratory analysis of distribution of copy number events was performed to compare all autosomal CN events larger than 1 kb between groups. Distribution of CN events per individual in the affected and non-affect cohorts, respectively, was: max. number - 499/511; min. number - 62/51; median - 159/147 for losses; while for gains we observed max. number - 1437/ 1166; min. number - 54/56; median - 160/213. Further, by means of permutation analysis using PLINK software, we found a CNV on chromosome 22 to differ between cases and controls (0.15% vs. 0 for deletions, and 38% vs. 10% for duplications, respectively; p-value=$9e^{-9}$).

3001S

HLA confer the risk of familial Mediterranean fever in Japanese population. *M. Yasunami[1], H. Nakamura[1], K. Agematsu[2], M. Yazaki[3], K. Migita[4].* 1) Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan; 2) Department of Infection and Host Defense, Shinshu University Graduate School of Medicine, Matsumoto, Japan; 3) Department of Medicine, Shinshu University School of Medicine, Matsumoto, Japan; 4) Clinical Research Center, National Hospital Organization Nagasaki Medical Center, Omura, Japan.

BACKGROUND Familial Mediterranean fever (FMF) has been considered to be an autosomal recessive trait which is characterized by self-limiting recurrent fever and serositis (OMIM #249100) and classified into a category of autoinflammatory disease. *MEFV*, identified as the responsible gene for FMF, encodes cytosolic protein pyrin (also known as marenostrin) which regulates the activity of NLRP3 inflammasome. Mutations in MEFV gene have been registered to “Infevers” database (http://fmf.igh.cnrs.fr/ISSAID/infevers/); 296 *MEFV* sequence variants have been reported so far (as of June 4th, 2014) including hot spots for pathogenic amino acid substitutions in the C-terminus region of the protein. Recently, dominant form (OMIM #134610) of the disease was reported (Koné-Paut I, *et al.*, *Rheumatology* 2009), and overlapping and continuum to other autoinflammatory diseases were proposed (Touitou I, *et al.*, *J Med Genet* 2013). We performed a nationwide surveillance of FMF in Japan and found that only M694I is commonly identified among hot spot mutations in the exon 10, homozygotes of which comprise as many as 10 % of Japanese FMF (Migita K, *et al.*, *Medicine* 2012). It is possible that the weak genotype-phenotype correlation observed in Japanese FMF is under the strong influence of modifier gene(s). *HLA* is one of the candidates for such modifiers because it is involved in many immune-related and inflammatory diseases. METHODS Genotypes of *HLA-B* and *-DRB1* loci were determined for 267 probands of FMF pedigree, who satisfied modified Tel-Hashomer criteria, by WAKFlow HLA-B and HLA-DR typing reagent kits (Wakunaga Pharmaceutical, Japan) after obtaining written informed consent. Frequencies of allele carriers in the patients were compared with those of the controls to obtain odds ratio for the evaluation of risk/protective effects of *HLA* alleles. RESULTS Carriers of *B*39:01* were increased (10.5% vs. 3.3%, OR=3.39, $p=7.1\times10^{-4} < 0.05/13=3.8\times10^{-3}$) in the patients, whereas those of *B*52:01* were decreased (15.7% vs. 25.4%, OR=0.55, $p=4.7\times10^{-3}$, not significant after Bonferroni’s correction). The decrease of *B*52:01* appeared secondary to the decrease of *DRB1*15:02* (12.8% vs. 24.1%, OR=0.46, $p=6.5\times10^{-4}$, $< 0.05/14=3.6\times10^{-3}$), which is in tight linkage with *B*52:01* in Japanese population. CONCLUSION The *HLA class I*-associated genetic risk and the *HLA class II*-associated genetic resistance were identified for Japanese FMF.

3002M

A pathogenic haplotypes of the g6pd gene correlating with enzyme activity. *D. Nantakomol[1], M. Chaowanathikhom[1], P. Nuchnoi[2].* 1) Department of Clinical Microscopy, Faculty of Allied Health Sciences, Chulalongkorn University, Bangkok, Thailand; 2) Department of Clinical Microscopy, Faculty of Medical Technology, Mahidol University.

The single nucleotide polymorphisms (SNP) in glucose-6-phosphate dehydrogenase (G6PD) gene caused enzyme deficiency leading to acute hemolytic anemia and neonatal jaundice. Over 400 SNPs in G6PD have been deposited in G6PD database. Most of the G6PD SNP showing clinical significance caused amino acid substitution. However, the association of G6PD SNP and enzyme activity could not be well established. Molecular characterization of 107 Thai adults revealed 11 reported mutations and 2 novel mutations (IVS-12: G+50A and 3’UTR: A*99G) of G6PD gene were detected. Mostly four SNPs consisted of IVS-11 (T93C) 50.5%, 3’UTR (A*357G) 50.5%, silent mutation-2 (C1311T) 49.5%, following by Viangchan variant (G871A) 21.5%. Haplotype analysis using Haploview software revealed the haplotype block of G6PD gene that consisted of four haplotypes (G1: GCTA, G2: GTCG, G3: ATCG and G4: GCCG). Only Viangchan variant (G871A) and haplotype G3 (ATCG) showed significance by association test and was found in G6PD deficient group more than normal group. Moreover, it showed high odds ratio (OR 7.90, 95% CI: 3.15-19.81) and affected G6PD mRNA secondary structure changing and microRNA binding that were predicted by CLC Main Workbench 6.9 and RegRNA 2.0 program, respectively. Altogether, this is the first report of haplotype association with reduction of G6PD enzyme expression. Thus, not only a single variation but also haplotype of G6PD gene should considered for the severity of G6PD deficiency.

3003T

Defective Dimerization of STAT3 causes Autosomal Dominant Hyper-IgE Syndrome. *M. Dasouki[1], S. Keles[2], T. Chatila[2].* 1) Neurology, Univ Kansas Med Ctr, Kansas City, KS; 2) Pediatric Immunlogy, Harvard University School of Medicine.

Background: Autosomal dominant and autosomal recessive hyper-IgE syndromes are primary immune deficiency syndromes caused by heterozygous mutations in STAT3 (Signal Transducer and Activator of Transcription 3) and homozygous mutations in DOCK8 & PGM3 respectively. STAT3 is involved in IL6 and related cytokines induced signal transduction by JAK kinases and represses macroautophagy. STAT3 forms a homodimer or a heterodimer with a related family member (at least STAT1) and mediates cellular responses to interleukins, KITLG/SCF and other growth factors. It also binds to (IL-6)-responsive elements identified in the promoters of various acute-phase protein genes. Upon tyrosine phosphorylation and dimerization, and in response to signaling by activated FGFR(1-4) it is translocated into the nucleus. Most previously reported STAT3 mutations involve its SH2 domain. Aim: To characterize the causal mutation and its functional effects in a 40 year old Dutch man with AD-HIES (severe chronic eczema, onychomycosis, asthma and bilateral cataract, lymphopenia, hypereosinophilia and hyper IgE). Methods: STAT3 mutation analysis by DNA sequencing. Activation of STAT3 by IL-6 and IL-21 and STAT1 by IFN-α was assessed by intracellular staining with anti-phospho (p)STAT3 and -pSTAT1 antibodies. TH17 and TH1 cell differentiation was assessed by measuring the production of IL-17 and IFN-γ, respectively. Results: Patient had normal B&T cell flow cytometry profiles and normal TREC count. A novel heterozygous STAT3 mutation (c.491G>A; p.R84Q) which is predicted to impair its dimerization was identified. Lymphocytes functional studies showed reduced phosphorylation of STAT3 in response to cytokine stimulation, whereas pSTAT1 activation was unaffected. Also, impaired TH17 responses and (early steps) in TH17 differentiation was found. Conclusion: This novel amino terminal domain STAT3 mutation causes defective dimerization and impaired TH17 responses indicating a novel mechanism for AD-HIES.

3004S

WES detects disease causing SNVs and CNVs in Primary immunodeficiencies. *H.S. Sorte[2], A. Stray-Pedersen[1,2,3], T. Gambin[1], P.S. Samarakoon[2], M.D. Vigeland[2], O.K. Rodningen[2], I.C. Hanson[4], L.R. Forbes[3,4], J.S. Orange[3,4], R. Lyle[2], J.R. Lupski[1,5,6], Baylor-Hopkins Center for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; 3) Center for Human Immunobiology, Texas Children's Hospital, Houston, TX; 4) Department of Pediatrics, Section of Immunology, Allergy, and Rheumatology, Baylor College of Medicine and Texas Children's Hospital, Houston, TX; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Human Genome Sequencing Center of Baylor College of Medicine, Houston, TX.

Primary immunodeficiencies (PIDDs) constitute a heterogeneous group of genetic diseases affecting the immune system. Depending on the genetic etiology, symptoms range from mild to severe and life threatening. Knowledge of the molecular genetic cause and disease mechanism is important and can direct targeted and curative therapy. However, subtype classification is difficult as patients often have overlapping phenotypes. In addition, more than 250 PIDD genes have been reported, and few are offered for diagnostic genetic testing. We examined the utility of whole-exome sequencing (WES) to detect single nucleotide variants (SNVs) and copy number variations (CNVs) in the diagnosis of PIDDs. As of June 2014, 275 patients with extensive immunological and genetic testing from 241 families have been recruited from Texas Children Hospital (Houston, USA) and Oslo University Hospital (Norway). Strategies for genetic analysis were tailored based on clinical data, immunophenotyping and family history, in 25 families more than one person WES tested, but for most families only the proband was subjected to WES. Initially, WES data were systematically screened for variants in reported and potential PIDD genes. In addition, a computational CNV prediction pipeline was applied to enable identification of potential disease-causing CNVs from the WES data. Analysis of the first 126 families identified PIDD relevant variants in 60 percent of the cases; half of these attaining a definitive molecular PIDD diagnosis. The other half had previously reported PIDD-causing variants, but with an unexpected or extended clinical phenotype, or heterozygote, potential deleterious variants in recessively inherited PIDD genes. In two families the patients had co-existing disease causing and modifying variants in two genes (*ZAP70/RNF168* and *SH2D1A/FANCB*, respectively). In one family two different PIDDs occurred (RAG1 SCID and DOCK8 related Hyper IgE syndrome). Other interesting findings include PIDD-causing CNVs in 7 families, somatic revertant mosaicism (*IL7R, FANCA, IL2RG*), de novo AK2 lymphocyte mosaicism, and 13 novel disease genes (5 definitive *NUDCD3, PGM3, COPA, MCM10, SIGIRR*, and 8 potential).

3005M

***PGM3* Mutations Cause a Congenital Disorder of Glycosylation with Severe Immunodeficiency and Skeletal Dysplasia.** *A. Stray-Pedersen[1,2,3,4], H.S. Sorte[4], P.H. Bache[5,6,7], L. Mørkrid[6,7], N.Y. Chokshi[3,8], H.C. Erichsen[9], T. Gambin[1], K.B.P Elgstøen[5,6], M. Bjørås[5,7], M.W. Wlodarski[10], M. Kruger[10], S.N. Jhangiani[1, 11], D.M. Muzny[1,11], A. Patel[12], K.M. Raymond[13], C.A. Martinez[14], S.M. Abraham[15], P. Hall[16], L.R. Forbes[2,3], E. Merckoll[17], J. Westvik[17], G. Nishimura[18], C.F. Rustad[4], O.K. Rødningen[4], E.A. Boerwinkle[1,11,19], R.A. Gibbs[1,11,], E. Lausch[10], I.C. Hanson[3], J.S. Orange[2,3], J.R. Lupski[1,8,11,12], Baylor-Johns Hopkins Center for Mendelian Genomics.* 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Center for Human Immunobiology, Texas Children's Hospital, Houston, TX 77030, USA; 3) Section of Immunology, Allergy, and Rheumatology, Department of Pediatrics, Baylor College of Medicine and Texas Children's Hospital, Houston, TX 77030, USA; 4) Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway; 5) 5Department of Microbiology, Oslo University Hospital, 0424 Oslo, Norway; 6) Department of Medical Biochemistry, Oslo University Hospital, 0424 Oslo, Norway; 7) Institute of Clinical Medicine, University of Oslo, 0318 Oslo, Norway; 8) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 9) Department of Pediatrics, Oslo University Hospital, 0424 Oslo, Norway; 10) Department of Pediatrics, Freiburg University Hospital, 79106 Freiburg, Germany; 11) Human Genome Sequencing Center, Baylor College ofMedicine, Houston, TX 77030, USA; 12) Medical Genetics Laboratories, Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 13) Department of Laboratory Medicine and Pathology, Mayo College of Medicine, Rochester,MN 55905, USA; 14) Center for Cell and Gene Therapy and Texas Children's Cancer and Hematology Centers, Baylor College of Medicine and Texas Children's Hospital, Houston, TX 77030, USA; 15) Pediatric Hematology Oncology, University of New Mexico, Albuquerque, NM 87106, USA; 16) Emory Genetics Laboratory, Department of Human Genetics, Emory University, Decatur, GA 30033, USA;; 17) Department of Radiology, Oslo University Hospital, 0424 Oslo, Norway;; 18) Department of Pediatric Imaging, Tokyo Metropolitan Children's Medical Center, 2-8-29 Musashidai, Fuchu, Tokyo 183-8561, Japan;; 19) Human Genetics Center, University of Texas Health Science Center, Houston, TX 77030, USA.

Human phosphoglucomutase 3 (PGM3) catalyzes the conversion of N-acetyl-glucosamine (GlcNAc)-6-phosphate into GlcNAc-1-phosphate during the synthesis of uridine diphosphate (UDP)-GlcNAc, a sugar nucleotide critical to multiple glycosylation pathways. We identified three unrelated children with recurrent infections, congenital leukopenia including neutropenia, B and T cell lymphopenia, and progression to bone marrow failure. Whole-exome sequencing demonstrated deleterious mutations in *PGM3* in all three subjects, delineating their disease to be due to an unsuspected congenital disorder of glycosylation (CDG). Functional studies of the disease-associated PGM3 variants in E. coli cells demonstrated reduced PGM3 activity for all mutants tested. Two of the three children had skeletal anomalies resembling Desbuquois dysplasia: short stature, brachydactyly, dysmorphic facial features, and intellectual disability. However, these additional features were absent in the third child, showing the clinical variability of the disease. Two children received hematopoietic stem cell transplantation of cord blood and bone marrow from matched related donors; both had successful engraftment and correction of neutropenia and lymphopenia. We define PGM3-CDG as a treatable immunodeficiency, document the power of whole-exome sequencing in gene discoveries for rare disorders, and illustrate the utility of genomic analyses in studying combined and variable phenotypes.

3006T

IPEX and IPEX-like syndromes: *FOXP3* and *FOXP3*-pathway related genes. *M. Vignoli[1,2], S. Ciullini Mannurita[1,2], G. Colarusso[3], F. Barzaghi[4], R. Bacchetta[4], M. Cecconi[5], A. Tommasini[6], A. Gennery[7], S.M. Holland[8], A.J. Cant[7], E. Gambineri[1,2].* 1) Dept of NEUROFARBA, section of Child's Health, University of Florence; 2) "Anna Meyer" Children's Hospital, Florence, Italy; 3) Santo Stefano Hospital, Paediatric Unit, Prato, Italy; 4) San Raffaele Telethon Institute for Gene Therapy (HSR-TIGET), Milan, Italy; 5) Human Genetic Laboratories, Ospedale Galliera, Genoa, Italy; 6) IRCCS Burlo Garofalo, Paediatric Immunology Laboratory, Trieste, Italy; 7) Department of Paediatric Immunology, Newcastle upon Tyne Hospitals, Newcastle upon Tyne, UK; 8) Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda.

IPEX (MIM #304790) is characterized by severe early-onset enteropathy, endocrinopathy, dermatitis and other autoimmune phenomena. It is due to *FOXP3* (Xp11.23) gene mutations. Interestingly an increasing number of patients show a phenotype consistent with IPEX but do not harbour *FOXP3* mutations. *IL2Rα* (10p15.1) and *STAT5b* (17q11.2) are two genes involved in *FOXP3* pathway whose mutations have been associated with IPEX-like penotype, nevertheless many patients remain without a molecular diagnosis. Our cohort of patients is composed of 76 cases with a phenotype compatible with IPEX syndrome, that were referred to us during the past 10 years for molecular analysis. We decided to perform a retrospective study evaluating clinical and laboratory findings to better delineate the clinical spectrum of multiple autoimmune diseases and identify different subgroups of patients to be addressed to specific molecular investigations. We analyzed the clinical history, T regulatory cells (Tregs) and FOXP3 expression by flow cytometry and the gene sequencing results of *FOXP3*, *IL2Rα*, *STAT5b*, *STAT1*, *IL10*, *IL10Ra*, *IL10Rb*. In 16 cases we identified *FOXP3* mutations confirming IPEX syndrome diagnosis. The 60 IPEX-like patients were divided in three distinct groups. The first group includes 31 cases with a phenotype very similar to IPEX, they showed enteropathy with onset within six months of life, growth retardation, eczema and at least one autoimmune phenomena. The second group comprises 22 cases with a later onset of the clinical symptoms (mainly diarrhoea and at least two autoimmune signs). The third group includes 7 cases with severe early-onset diarrhea together with perianal abscesses or enteric fistulae, these cases are more similirar to Inflammatory Bowel Disease manifestation. Autoimmune enteropathy is the key clinical sign for IPEX-like patients of all groups, while in subjects without enteropathy, endocrinopathy and cytopenias are the main clinical symptoms. High IgE levels is often reported. In the first group the molecular analysis resulted in 3 patients harboring *IL2Rα* mutations leading to the absence of CD25 expression, a patient with a *STAT5b* variation and one *STAT1* deficiency case, confirming the association between mutations in *FOXP3*-related genes and clinical features very similar to IPEX. Therefore the clusterization in distinct subgroups can help to address the molecular analysis of different genes and to select patients for whole exome sequencing analysis.

3007S

Aicardi-Goutières syndrome is caused by IFIH1 mutations. *H. Oda[1,2], K. Nakagawa[1], J. Abe[1,3], T. Awaya[1], M. Funabiki[4], A. Hijikata[5], R. Nishikomori[1], M. Funatsuka[6], Y. Oshima[7], Y. Sugawara[8], T. Yasumi[1], H. Kato[4,9], T. Shirai[5], O. Ohara[2,10], T. Fujita[4], T. Heike[1].* 1) Department of Pediatrics, Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Integrative genomics, RIKEN-Institution for Integrative Medical Sciences, Yokohama, Japan; 3) Pediatric Department, Kitano Hospital, Osaka, Japan; 4) Laboratory of Molecular Genetics, Institute for Virus Research, Kyoto University, Kyoto, Japan; 5) Department of Biosciences, Nagahama Institute of Bio-Science and Technology, Nagahama, Japan; 6) Pediatric Department, Tokyo Women's Medical University, Tokyo, Japan; 7) Pediatric Department, Faculty of Medical Sciences, University of Fukui, Fukui, Japan; 8) Pediatrics and Developmental Biology Department, Tokyo Medical and Dental University, Tokyo, Japan; 9) PRESTO, JST, Kawaguchi, Japan; 10) Human Genome Research Department, Kazusa DNA Research Institute, Kisarazu, Japan.

Aicardi-Goutières syndrome (AGS) is a rare, genetically determined early-onset progressive encephalopathy. Individuals affected with AGS typically exhibit severe neurological symptoms associated with basal ganglia calcification and type I interferon elevation in the cerebrospinal fluid. To date, mutations in six genes (TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, and ADAR1) have been identified as etiologic for AGS. However, our previous Japanese nationwide AGS survey (Abe, 2013) as well as other reports showed a group of AGS individuals without mutations in the six genes above, suggesting the existence of additional AGS associated genes. We performed trio-based whole exome sequencing in three AGS individuals without molecular diagnosis. After the removal of common polymorphisms, de novo missense mutations in IFIH1 were identified in all three. IFIH1 encodes MDA5, one of the cytosolic pattern recognition receptors that recognize double-stranded RNA (dsRNA). MDA5 activation by its oligomerization on dsRNA results in the enhancement of type I interferon transcription, leading to the activation of a large number of interferon-stimulated genes (ISGs). Indeed, transfection of IFIH1 plasmids containing the mutations in the AGS individuals into Huh7 cells showed significantly stronger promoter activities of IFNB1 as well as downstream IFIT1 transcription, suggesting that these IFIH1 mutations induce activation of type I interferon pathway and are responsible for the phenotypes of AGS. Furthermore, we retrovirally reconstituted Ifih1-null mouse embryonic fibroblasts (MEFs) with the IFIH1 mutants, and subsequently infected these MEFs with encephalomyocarditis virus (EMCV), an MDA5-specific ligand. Interestingly, these MDA5 mutants lacked ligand-specific responsiveness, which was similar to the lack of ligand responsiveness of p.Gly821Ser mouse MDA5 mutant, previously reported in the SLE mouse model (Funabiki, 2014). Our results, currently under revision, showed that AGS is caused by the mutations in IFIH1, which has also been recently reported by Rice et al (Rice, 2014). Further analysis remains to elucidate the precise mechanism how these IFIH1 mutations cause the type I interferon overproduction.

3008M

The comprehensive genetic analysis of congenital anomalies of kidney and urinary tract (CAKUT) in Japan. *N. Morisada[1], M. Taniguchi-Ikeda[1], K. Nozu[1], A. Shono[1], K. Kamei[2], S. Ito[2], R. Tanaka[3], K. Iijima[1].* 1) Pediatrics, Kobe University Graduate School of Medicine, Kobe, Hyogo, Japan; 2) Nephrology and Rheumatology, National Center for Child Health and Development, Tokyo, Japan; 3) Nephrology, Hyogo Prefectural Children's Hospital, Kobe, Japan.

Purpose: Congenital anomalies of kidney and urinary tract (CAKUT) are the most common cause of pediatric end-stage renal disease over the world. The detailed mechanisms of development of CAKUT were still unclear; however genetic mutations in various genes may play pivotal roles. We have conducted the comprehensive genetic analyses of CAKUT in Japan. Methods: This study was approved by the Institutional Review Board of Kobe University Graduate School of Medicine. Genomic DNA samples of patients were provided from peripheral blood mononuclear cells. We performed direct sequencing, multiplex ligation-dependent probe amplification (MLPA) analysis, array comparative genome hybridization (aCGH) analysis and/or next-generation sequencing (NGS). Results: We examined 206 patients from 187 families. Ninety-one patients out of 82 families were non-syndromic CAKUT (NSC) and 115 patients out of 105 families were syndromic CAKUT (SC). We identified the responsible genes in 62 patients from 44 families. The detection rates in NSC and SC families were 14.6% and 30.5%, respectively. The most common responsible gene found in this cohort was *PAX2* (11 families). Other responsible genes were *EYA1* (9), *HNF1B* (7), *UMOD* (1), *OFD1* (1), *SALL1* (1) and *CHD7* (1). MLPA analysis enabled us to identify *TSC2-PKD1* contiguous gene deletions in a patient in our cohort. aCGH is useful to reveal the responsible chromosomal lesion in CAKUT with intellectual disability (ID), and we have identified copy number variations in 9 patients with CAKUT and ID such as 22q11.2 microdeletion, 16q microdeletion and 1q21.1 microdeletion. We identified several responsible genes in the patients by NGS. Conclusions: This is the first nation-wide study for the genetic approach of CAKUT in Japan.

3009T

Deciphering Molecular Etiology of the Mayer-Rokitansky-Küster-Hauser (MRKH) Syndrome. *A.B. Ekici[1], C. Büttner[1], J. Pschirer[1], P. Strissel[2], P. Oppelt[2], S. Renner[2], M. Beckmann[2], R. Strick[2].* 1) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany; 2) Department of Gynecology and Obstetrics, Laboratory for Molecular Medicine, University-Clinic Erlangen, Germany.

The Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome describes women with congenital aplasia of the uterus, cervix and the upper two-thirds of the vagina due to absent paramesonephric (Müllerian) ducts. Women with MRKH have a normal female karyotype (46, XX), normal female external development, internal normally regressed mesonephric (Wolffian) ducts and have correctly timed pubarche and thelarche. The incidence of the MRKH syndrome is one in 4,500 female newborns. MRKH patients represent the second most frequent cause of primary amenorrhea after Turner syndrome and commonly have associated malformations, like skeletal and renal abnormalities. The MRKH syndrome occurs isolated as a genital malformation (type 1), but also with associated malformations (especially malformations of the kidneys and urinary tract (type 2) and MURCS association (Müllerian renal and cervicothoracic somite abnormalities; partial or complete absence of the spinal column, ribs, or arms; asymmetric or improperly developed ribs or arms). To date the genetic or molecular etiology of the MRKH syndrome is completely unknown. We performed copy number variation analysis with high density SNP microarrays, whole exome sequencing and sequenced the enriched methylome from two pairs of discordant monozygotic twins and their family members. Since the results of this integrative analysis of these three high-throughput analysis data sets did not resulted in clear causative variants even with a combined genetic-epigenetic disease model for the data analysis, we now made the next step in the extent of data depth. We now used two in depth analysis methods: sequencing whole genomes and bisulfite converted whole genomes in parallel. The first method enables in addition to comprehensive SNV and small InDel analysis also a thoroughly analysis for structural variants like large InDels, inversions, translocations etc. The latter method is the gold standard for DNA methylation analysis and provides single nucleotide resolution map of 5-mC of the genome. So far the first tier of data analysis is very promising and a complete comparative analysis is ongoing. We believe that the methods in the present application will help to unravel the genetic or molecular causes of MRKH and be of clinical relevance in order to help clinicians for an accurate molecular diagnosis and council patients and their families in order to understand the syndrome.

3010S

Utility of whole exome sequencing in diagnosis of individuals with congenital anomalies of kidney and urinary tract. *M. Bekheirnia, N. Bekheirnia, M. Bainbridge, N. Janzen, S. Jhangiani, D. Muzny, M. Michael, E. Brewer, E. Elenberg, A. Kale, A. Riley, S. Swartz, P. Srivaths, S. Wenderfer, R. Alan Lewis, E. Gonzales, R. Gibbs, J. Lupski, J. Belmont, D. Roth, M. Braun, D. Lamb.* Baylor College of Medicine, Houston, TX.

Introduction: Congenital Anomalies of Kidney and Urinary Tract (CAKUT) are the most common cause of pediatric end stage renal disease (ESRD). The discovery of underlying genetic etiologies would not only improve diagnostic precision for non-syndromic forms, optimize management of affected patients, prevent disease recurrence by PGD (preimplantation genetic diagnosis), but also aid prenatal diagnosis in forms that result in early-onset ESRD. Here, we show the utility of whole exome sequencing (WES) in a diagnostic setting and in the genomic research of CAKUT. Methods: We completed research WES in 65 families with CAKUT. Approximately 25% of the cases were familial and 10% were syndromic. We used the ACMG (American College of Medical Genetics and Genomics) standards to classify genetic variants. We focused on analysis of the cases with mutations in 20 genes that are known to cause CAKUT and are central to development of the kidney and urinary tract. Results: Mutational analysis was performed and the following deleterious or possibly deleterious mutations were identified: *PAX2* (G24fs, S171P), *HNF1B* (Q378fs), *RET* (S401C), *EYA1* (c.867+5G>A), *SIX2* (P241L), and *DSTYK* (R592Q, D120N). No mutation was identified in *GDNF, SIX1, SOX17*, or *GATA3*. Benign variants or variants of unknown significance were identified in *BMP7, CDC5L, CHD1L, SALL1, SIX5, ROBO2, UPK3A, BMP4, KAL1*, and *TNXB*. The G24fs mutation occurs in a family with multiplex renal dysplasia and membranous nephropathy that has not been reported previously with *PAX2* mutations. Q378fs is a novel mutation in *HNF1B*. Phenotypes of some of these families expand our current knowledge about the corresponding genes. Conclusion: This is the first report utilizing WES for diagnosis in patients with CAKUT. At least 12% of individuals with CAKUT have deleterious mutations in known genes that can be identified by WES. Therefore, WES should be considered as a diagnostic tool in such patients, especially in the families with complex medical histories. Identification of novel mutations and expansion of the phenotypes are additional important benefits of WES as a diagnostic tool. Acknowledgement: Supported by: a-K12 DK0083014, the Multidisciplinary K12 Urologic Research (KURe) Career Development Program to DJL (MRB is a KURe Scholar). b- NHGRI Mendelian grant to JRL c- RAL is a Senior Scientific Investigator of Research to Prevent Blindness, New York, whose unrestricted funds supported part of these investigations.

3011M

Focal segmental glomerulosclerosis exomes reveal candidate variants highly enriched in cell movement and cell adhesion related genes. *J. Suh[1], G. Genovese[1,2], V. Charoonratana[1], A. Knob[1], M. Pollak[1,2].* 1) Beth Israel Deaconess Medical Center, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA.

The findings of focal segmental glomerulosclerosis (FSGS)-causing genes established that genetic components contribute to the development of FSGS during recent two decades. However, in majority of families even with the evidence of strong inheritance, underlying genetic cause is unknown. Here, we analyzed exome of families and sporadics which are unexplained by known FSGS genes in an extended effort to find the gene responsible for the development of FSGS. We performed exome sequencing of 403 probands from 311 familial and sporadic cases of FSGS. We also sequenced 96 unaffected family members. After variants calling, variants are filtered against common variants from the 1,000 Genomes Project, dbSNP, and Exome Sequencing Project except known FSGS-causing variants, and then novel non-synonymous or splice variants are kept. Exome analysis revealed that 30.3 percent of all cases contain known FSGS-causing variants with 47.9 percent of familial cases and 16.0 percent of sporadic cases, respectively. After excluding explained cases, we filtered genes showing more than 4 novel non-synonymous and splice variants, and grouped genes according to 1) their cellular component, molecular function, and biological process using gene ontology service, AmiGO2, and 2) their protein-protein interaction using DAPPLE. GO terms of genes from unexplained cases were significant enriched in cellular component movement and cell adhesion both with p-value less than 0.0001. DAPPLE results consistently showed two big protein-protein interactions each involving cell movement and cell adhesion, respectively. Here we show sporadic FSGS cases are less explained by known FSGS-causing genes compared to familial cases. These data show that novel non-synonymous or splice variants are highly enriched in cell movement- and cell adhesion-involved genes in the exome of FSGS patient samples, suggesting the importance of podocyte movement and interaction with neighboring cells and extracellular matrix.

3012T

The role of *MAZ* in the regulation of genitourinary development via modulation of WNT signaling. *M. Haller, D.J. Lamb.* Baylor College of Medicine, Department of Molecular and Cellular Biology, Center for Reproductive Medicine, Houston, TX.

Genitourinary (GU) birth defects comprise some of the most common yet least studied congenital malformations and range in severity from conditions such as undescended testes (cryptorchidism) and ventrally misplaced urethral meatus (hypospadias) to highly complex malformations such as bladder exstrophy epispadias complex (BEEC) and ambiguous genitalia. Congenital anomalies of the kidney & urinary tract are also common and include phenotypes such as duplicated tract components, kidney agenesis, congenital hydronephrosis, horseshoe kidney, and cystic kidneys. Genomic aberrations such as copy number gains and losses can result in congenital malformations of the GU tract, among other organ systems. Genome wide array comparative genomic hybridization (aCGH) together with an extensive literature review allowed identification of genomic hotspots of GU development and delineated the smallest CNV regions of maximum overlap. CNV mapping revealed over 30 patients with GU defects harboring duplications or deletions in the syndromic genomic region, 16p11.2 - the most common known pathogenic gene dosage region in humans. The only gene covered collectively by all the mapped CNVs in patients with GU defects was MYC-associated zinc finger (*MAZ*). *MAZ* encodes a transcription factor with a similar consensus sequence to that of *WT1*, and is implicated in WNT signaling. *In situ* hybridization on mouse embryos and isolated mouse GU tracts defined the expression profile of *Maz* during development. *In situ* experiments showed robust staining of the GU tract including the kidneys, ureters, testes, bladder and genital tubercle. Immunohistochemistry confirmed this expression pattern at the protein level. Analysis of the lab's cohort of genomic DNA from GU-abnormal patients and GU-normal fertile controls via CNV qPCR determined the frequency of *MAZ* CNVs in these groups. Validated by two independent probe sets, 6% of our GU abnormal cohort displayed CNVs in *MAZ* (n=258) compared to 0% in controls (n=57). As predicted based on its consensus sequence, knockdown or overexpression of *MAZ* in human embryonic kidney (HEK293) cells results in differential expression of several WNT pathway genes including *DKK1/2/4, LRP5/6, FZD1/9/10, DACT1, SFRP4/5, WNT4, WNT11* and *GSK3A*. *MAZ*, once thought to be simply a housekeeping gene, encodes a dosage sensitive transcription factor that may play a key role in urogenital development and contribute to the congenital malformations of the 16p11.2 phenotype.

3013S

Functional Characterization of Renin Variants Identified in African Americans in Exome Sequencing Project. *S. Kmoch[1], M. Zivna[1], K. Hodanova[1], P. Vyletal[1], V. Stranecky[1], B.I. Freedman[2], A.J. Bleyer[2].* 1) First Medical Faculty, Charles University of Prague, Prague, Czech Republic; 2) Section of Nephrology, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.

Renin is a key component of blood pressure regulation. It is encoded by the *REN* gene and produced in juxtaglomerular cells in kidney. Its primary structure consists of a 23 amino acid N-terminal signal sequence, 43 amino acid "pro" domain, and the mature renin comprising 340 amino acids. The *REN* mutations lead to two different human diseases. Dominant mutations located within the signal sequence of *REN* cause familial juvenile hyperuricemic nephropathy type 2 (HNFJ2; OMIM 613092), characterized by early onset hyperuricemia, anemia and chronic kidney disease. Recessive mutations, mostly nonsense, affecting synthesis or the activity of renin cause renal tubular dysgenesis (RTD; OMIM 267430). To further study *REN* variation, we interrogated the Exome Variant Server (http://evs.gs.washington.edu/EVS/) and identified three nonsynonymous variants P8A, R33W and G217R. These variants are present exclusively in African-Americans with 0.005; 0.001 and 0.062 allele frequencies, respectively. Localization of these variants in the context of *REN* mutations associated with HNFJ2 and RTD suggested, that if causal, the P8A variant may lead to symptoms of HNFJ2 in 1.17 % and R33W and G217R variants to RTD in 0.35 % of African Americans. To asses functional impact of these variants, we cloned individual cDNAs and transiently expressed wild-type and corresponding mutant proteins in HEK293 cells. We studied proteolytic processing and glycosylation patterns of expressed proteins, measured renin activities in cell lysates and cultured media and investigated cellular localization of expressed proteins. Our analysis revealed, that the P8A variant does not affect biosynthesis and activity of prorenin and renin but reduces their secretion rate by ~ 50 % compare to wild-type protein. The R33W variant reduces biosynthesis of prorenin and renin to ~ 25 % of the wild-type protein and also affects formation of characteristic intracellular renin secretory granules. The G217R variant had no effect on prorenin and renin properties. The results of the analysis suggest that P8A and R33W variants significantly affect renin biosynthesis and secretion and may thus contribute to low plasma renin activity and a higher incidence of kidney diseases in populations of African origin. Exact population frequencies and clinical effects of these two variants should be therefore further studied using association studies in appropriate population cohorts and by detailed phenotyping of individual carriers.

3014M

Detection of genes causing polycystic kidney in Saudi Arabian Fetuses and Neonates. *M.H. Al-Hamed[1], W. Kurdi[2], Z. Abdullah[3], M. Tulbah[2], M. Al-Nemer[2], R. Khan[2], N. Al-Sahan[2], M. Albaqumi[1,4].* 1) Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia; 2) Department of Obstetrics and Gynecology, KFSHRC, Riyadh, Saudi Arabia; 3) Obstetrics and Gynaecology Dept, PSMMC, Riyadh, Saudi Arabia; 4) Department of Medicine, KFSHRC, Riyadh, Saudi Arabia.

Polycystic kidney disease (PKD) is common in the Saudi population and is associated with significant morbidity and mortality. Recent reports estimated the prevalence of PKD in Saudi population to be 0.57 per 1000 births. The majority of polycystic disease in fetuses arises from recessively inherited mutations of genes expressed in the primary cilium of renal epithelial cells. To investigate the genetic basis of polycystic disease in Saudi families, we recruited forty fetuses presenting with PKD diagnosed by ultrasound. Molecular karyotype to exclude aneuploidy was negative for all cases. Sequencing of the following genes (PKHD1, NPHP1, INVS, NPHP3, NPHP4, IQCB1, CEP290, GLIS2, RPGRIP1L, AHI1, CC2D2A, MKS1, TCTN2, TMEM67, and TMEM216) resulted in identification of molecular causes of PKD in 16 cases (40%). Mutations of the CC2D2A gene are most common and account for 44% of mutations detected. Mutations in the CC2D2A gene are always associated with cystic kidneys and Encephaloceles indicating Meckel-Gruber syndrome diagnosis. Novel mutations in INVS and TCTN2 genes were detected. Oligogenic inheritance has been observed in two families. A third family had homozygous mutation in TMEM67 gene and another homozygous mutation in the RPGRIP1L gene. These types of inheritance (Oligogenic and homozygosity in two Cilia genes) are likely aggravating the phenotype.

3015T

Loss of *Zeb2* Causes Glomerulocystic Kidney Disease in Mice. *H. Milo Rasouly, S. Chan, A. Pisarek-Horowitz, W. Lu.* Renal, Boston University, Boston, MA.

BACKGROUND: ZEB2 is a zinc finger E-box-binding homeobox transcription factor. Mutations in *ZEB2* cause the Mowat-Wilson syndrome (MIM #235730), an autosomal dominant disorder characterized by multiple congenital anomalies including kidney anomalies. However, the role of ZEB2 during kidney development is unknown. **METHODS**: ZEB2 expression was analyzed by immunostaining of the embryonic kidneys during mouse development. Kidney specific *Zeb2* conditional knockout mice were generated by crossing the *Zeb2* floxed allele ($Zeb2^{flox/flox}$) with the *Pax2-cre* and *Six2-cre* alleles. The conditional knockout mice were analyzed from embryonic stage E14.5 to postnatal 8 weeks old. The mouse kidneys were analyzed by H&E staining and immunostaining for apoptosis. Gene expression analysis was performed by TaqMan assays. Kidney function was assessed by measuring urine protein levels, serum creatinine and blood urea nitrogen. **RESULTS**: ZEB2 is highly expressed during mouse kidney development. Both $Zeb2^{flox/flox}$*;Pax2-cre* and $Zeb2^{flox/flox}$*;Six2-cre* conditional knockout mice developed kidney glomerular cysts starting at E16.5 days. $Zeb2^{flox/flox}$*;Pax2-cre* mice died at birth and $Zeb2^{flox/flox}$*;Six2-cre* mice developed abnormal kidney function at 5 weeks old. The cysts originate from glomeruli with dilated Bowman's capsules and collapsed glomerular tufts. Reduced apoptosis was detected in the S-shaped and C-shaped bodies of *Zeb2* knockout mice compared to the wild type controls. Gene expression analysis revealed increased levels of *Pkd1* mRNA in the developing kidney of both *Zeb2* conditional knockout mice compared to the wild-type controls. Interestingly, overexpression of *Pkd1* has been reported to cause glomerular cysts in mice. **CONCLUSIONS**: Loss of *Zeb2* causes glomerulocystic kidney disease in mice with increased expression of the *Pkd1* transcript.

3016S

Whole exome sequencing identifies a homozygous mutation in *SOHLH1* in two sisters with non-syndromic hypergonadotropic hypogonadism. *Y. Bayram[1], T. Guran[2], G. Yesil[3], Z. Atay[2], E. Karaca[1], T. Gambin[1], S. Turan[2], D. Pehlivan[1], S.N. Jhangiani[4], D. Muzny[4], B. Halioglu[2], A. Bereket[2], R.A. Gibbs[4], J.R. Lupski[1,5,6].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 2) Department of Pediatric Endocrinology and Diabetes, Marmara University Hospital, Istanbul, Turkey; 3) Department of Medical Genetics, Bezmialem University, Istanbul, Turkey; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 6) Texas Children's Hospital, Houston, TX, USA.

Patients with hypergonadotropic hypogonadism (HH) present with primary or secondary amenorrhea or delayed puberty secondary to primary gonadal dysfunction and characterized by elevated gonadotropins and low sex steroids. HH is always a pathological state; therefore, assessment should include history of surgery, irradiation and chemotherapy, examination for features of the Turner syndrome and consideration of a karyotype. Several genetic defects have been shown to cause HH including genes having a role in gonadal development or maintenance and defects in sex steroid synthesis, or end organ resistance to gonadotropins. However, elucidation of novel genetic defects causing HH will provide major insights into the biology of this condition - especially knowledge of the regulation of human reproductive function. Here we report two sisters who were born to a first degree cousin marriage presenting with HH. Both of them were referred to a pediatric endocrinology clinic because of primary amenorrhea and lack of secondary sex characteristics at 16 and 14 years of age, respectively. Further hormonal investigations revealed elevated FSH and LH and decreased estradiol concentrations. On the pelvic ultrasound imaging both patients had hypoplastic ovaries and prepubertal-sized uteri. Karyotype analyses were normal and no other significant clinical or dysmorphic features were identified; the patients were diagnosed with non-syndromic HH. We performed whole exome sequencing (WES) for both affected patients which identified a homozygous frame-shift deletion (c.705delT; p.Lys236Argfs*3) in the *SOHLH1* gene. Segregation analysis within the family revealed that both parents were heterozygous carriers while the unaffected sister was wild type for the same variant consistent with Mendelian expectations. *SOHLH1* encodes a basic helix-loop-helix transcription factor with homologues in humans and other placental mammals. Previous mouse studies revealed that *Sohlh1* is preferentially expressed in mouse oocytes during early folliculogenesis and required for oogenesis. In a *Sohlh1* knock-out mouse model, *Sohlh1* -/- female mice have been observed to be infertile with atrophic ovaries that lacked oocytes while the *Sohlh1* +/- females were normal. In conclusion; we describe the first homozygous *SOHLH1* mutation in humans and clinical features in our patients together with findings observed in a mouse model strongly suggest that homozygous mutations in *SOHLH1* cause ovarian dysfunction.

3017M

IMAGe syndrome mutations in the PCNA-binding site of *CDKN1C* cause in increased protein stability and inhibition of the cell cycle. *V.A. Arboleda[1,2], K.S. Borges[3], E. Vilain[1,4,5].* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 2) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA; 3) Department of Genetics, Ribeirão Preto Medical School, University of São, Ribeirão Preto, Brazil; 4) Department of Pediatrics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 5) Department of Urology, David Geffen School of Medicine at UCLA, Los Angeles, CA.

Mutations in CDKN1C (a.k.a. P57KIP2) can cause opposing genetic syndromes: the undergrowth syndromes IMAGe Syndrome (Intrauterine growth retardation, Metaphyseal dysplasia, Adrenal hypoplasia and Genital anomalies) and Russell Silver Syndrome (RSS) and the overgrowth syndrome Beckwith-Wiedemann syndrome (BWS). Gain-of-function mutations located in the PCNA-binding domain of CDKN1C result in a spectrum of undergrowth disorders ranging from IMAGe Syndrome and RSS, while loss-of-function mutations in CDKN1C cause BWS. CDKN1C is a cyclin-dependent kinase inhibitor that functions as a negative regulator of cell proliferation. Here, we investigate the effects of gain-of-function CDKN1C mutations on protein stability during cell cycle progression and on cell proliferation to understand the molecular mechanisms underlying IMAGe syndrome. 293T cells transfected with wild type and IMAGe-mutant (p.K278E or p.F276V) and treated with de novo protein synthesis inhibitor, cyclohexamide (CHX), were subject to immunoblotting and demonstrated an increased stability of IMAGe-mutant CDKN1C compared to wild-type CDKN1C. To demonstrate that IMAGe-mutant CDKN1C resulted in impaired progression through the cell-cycle, we transfected them with wild-type, IMAGe-mutant (p.K278E), or BWS-mutant (p.L42P), synchronized the cells, and then assessed the levels of CDKN1C protein over a 24-hour period. Both the wild-type and BWS-mutant cells showed peak levels of CDKN1C at 3- and 15- hours post thymidine release while the IMAGe-mutant showed stable CDKN1C protein expression over 24-hours, indicating a lack of progression through the cell cycle. Cell-cycle phase distribution at the time T0 showed that in IMAGe-mutant transfected cells, a larger proportion of cells were in the G1 phase (36.94%) compared to wild-type CDKN1C, empty plasmid, and BWS-mutant (32.31%, 14.88% and 19.85 respectively). Finally, clonogenic assays demonstrated that cells transfected with IMAGe-mutant CDKN1C had smaller and fewer colonies than cells transfected with wildtype CDKN1C plasmids. In conclusion, gain-of-function mutations in the PCNA-binding site of CDKN1C increase protein stability by inhibiting degradation of CDKN1C during cell-cycle progression. Therefore, cells are stalled in G1-phase and have decreased clonogenic potential. Our findings shed light on the molecular mechanism behind IMAGe mutations on cell cycle, proliferation, and undergrowth disorders in humans.

3018T

Hypopituitarism caused by dysregulation of pituitary progenitors and epithelial to mesenchymal transition. *M.Ines. Perez Millan, A.H. Mortensen, M.L. Brinkmeier, S.A. Camper.* Human Genetics, University of Michigan, Ann Arbor, MI.

Mutations in PROP1 and several other transcription factors, including POU1F1 (PIT1), HESX1, and OTX2, cause Multiple Pituitary Hormone Deficiency (MPHD) in humans and mice. PROP1 is co-expressed with stem cell markers including the growth factor receptor, GFRA2 and transcription factors SOX2, SOX9, and OCT4 (1, 2). In mouse fetal development *Prop1* expression coincides spatially and temporally with the delamination, migration, and differentiation of progenitor cells in Rathke's Pouch (RP) into hormone producing cells of the adenohypophysis. *Prop1* mutant progenitors fail to undergo these processes, leading to organ dysmorphology, poor vascularization, and postnatal pituitary hypoplasia (3). *Prop1* is necessary for regulating expression of two other MPHD genes; repressing *Hesx1* and activating *Pou1f1*. Mutations in *PROP1* are the most common known cause of MPHD, but most of the patients have no molecular diagnosis. The aim of this work is to identify additional downstream target genes of *Prop1* to reveal new candidate genes for the cases of hypopituitarism that remain unexplained and to better understand the pathophysiology of MPHD. We chose an in vitro approach of identifying DNA sequences bound by PROP1 in an immortalized pituitary cell line by ChIP-Seq technology. Because the efficiency and specificity of ChIP depends on the quality of the antibody, we developed a biotin-tagged PROP1 system that permits precipitation of PROP1 with an avidin based detection assay. We identified enrichment of PROP1 binding at the *Pou1f1* gene and in genes that encode components of the pathways of cell junction signaling and regulation of epithelial to mesenchymal transition (EMT). We examined expression of several candidate genes in *Prop1* mutant mice and confirmed that their expression was altered. These data suggest that *Prop1* promotes the transition of progenitors to differentiation by suppressing expression of progenitor markers like SOX genes and repressing claudin gene expression, in order to release tight junctions and permit EMT. In addition, *Prop1* activates expression of cyclin E, a marker of the transition state, and *Pou1f1*, a marker of the differentiation state. This study establishes new mechanisms underlying PROP1's role in pituitary progenitor cell regulation and offers new candidate genes for MPHD. References 1. Fauquier et al. Proc Natl Acad Sci 105:2907-12, 2008. 2. Garcia-Lavandeira et al. PLoS One 4:e4815, 2009. 3. Ward et al. Mol Endo 20:1378-1390, 2006.

3019S

Cell Specific Biochemical Changes in Snyder Robinson Syndrome. *J. Albert[1,2], L. Wolfe[1,2], W. Bone[1], T. Markello[1,2], D. Adams[1,2], C. Schwartz[3], R. Gafni[4], M. Collins[4], L. Tosi[5,6], N. Bhattacharyya[4], W. Gahl[1,2], C. Boerkoel[1].* 1) Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, MD, USA; 2) Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD, USA; 3) J.C. Self Research Institute, Greenwood Genetics Center, Greenwood, SC 29646, USA; 4) Skeletal Clinical Studies Unit, Craniofacial and Skeletal Disease Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD 20892, USA; 5) George Washington University School of Medicine, Washington, DC, USA; 6) Children's National Medical Center, Washington, DC, USA.

Spermidine and spermine are simple positively charged ubiquitous molecules synthesized by spermidine synthase and spermine synthase (SMS), respectively. Spermine and spermidine homeostasis is crucial for ion channel regulation, transcription and translation. Deficiency of the SMS enzyme disturbs this homeostasis in Snyder-Robinson syndrome (SRS), a form of X-linked intellectual disability. Individuals with SRS have developmental delay, dysmorphic facies, osteoporosis, seizures, and hypotonia. A major cause of morbidity in these individuals is bone disease. To elucidate the mechanism of osteoporosis, we obtained osteoblasts from an individual with SRS and our results demonstrate that there is an osteoblast specific effect of deficient SMS enzyme activity and/or aberrant polyamine homeostasis. SMS fibroblasts and osteoblasts have similar decreases in SMS protein compared with normal controls. However, normal control osteoblasts have less SMS protein compared with normal control fibroblasts suggesting that osteoblasts may be more sensitive to more subtle changes in SMS protein. In both cell lines, we observed a slight decrease in both cytoplasmic and nuclear SMS protein as detected by cellular fractionation and western blot. Furthermore, the differences in spermine and spermidine levels were significantly lower and higher, respectively in the osteoblasts. These osteoblast specific changes in polyamine metabolism may exemplify the mechanism underlying the cellular specificity of disease in SRS.

3020M

Genetic and enzymatic characterization of the anti-oxidant enzyme GPx1 in sickle cell disease patients. *M. Beaudoin[1], G. Galarneau[1], A. Binding[2], M. Reid[3], I. Hambleton[4], C. McKenzie[3], J.N. Hirscchorn[5,6], M. Warner[2], G. Lettre[1].* 1) Montreal Heart Institute and Université de Montréal, Montreal, Quebec, Canada; 2) McGill University, Montreal, Quebec, Canada; 3) Tropical Medicine Research Institute, The University of the West Indies, Mona, Jamaica; 4) The University of the West Indies, Cave Hill, Barbados; 5) Children's Hospital Boston, Boston, MA, USA; 6) Broad Institute, Cambridge, MA, USA.

Introduction: Sickle cell disease (SCD) is a genetic blood disorder characterized by anemia, pain crises, strokes and early death. Affected individuals inherit two copies of a mutation in the gene encoding β-globin, one of the two subunits of hemoglobin. Despite of this simple mode of inheritance, SCD is characterized by patient-to-patient clinical heterogeneity. Alpha-thalassemia and fetal hemoglobin levels are two known modifiers of SCD severity. The goal of our study was to identify new genetic modifiers of SCD using whole-exome sequencing followed by functional characterization. Methods: We sequenced the whole-exome of 19 patients selected from the Jamaican SCD birth cohort. Over 18 years of follow-up observations, these patients presented with few severe SCD-related complications when compared to the rest of the cohort (0.07 vs. 0.57 severe events per year, $p<0.001$). Analysis of the whole-exome sequence data identified a common missense variant in GPX1 (Pro200Leu, rs1050450) that is enriched in Jamaican patients with mild SCD. We obtained DNA and red blood cell (RBC) membranes from 40 SCD patients recruited at the Royal Victoria Hospital in Montreal, Canada. We genotyped GPX1 Pro200Leu using a high resolution melting protocol. We measured GPx1 protein level and enzymatic activity with ELISA and absorbance assays, respectively. Results: The GPX1 Leu200 allele is enriched in 19 “mild” Jamaican SCD patients when compared to populations of African ancestry from the 1000 Genomes Project (47% vs. 25-34%). In 95 patients from the Jamaican SCD birth cohort, GPX1 Pro200Leu is associated with a composite score that summarizes severe SCD-related events collected from birth until adulthood ($p=0.02$). In RBC membranes from SCD patients, GPX1 Pro200Leu is not associated with GPx1 protein amount ($p=0.95$), but is correlated with GPx1 enzymatic activity ($p=0.049$). Conclusion: GPX1 encodes glutathione peroxidase, an enzyme that protects RBC from free radical damages. Previous studies have implicated free radical production in SCD severity, and in particular in the etiology of vaso-occlusive crises. Our results indicate that a common missense variant in SCD patients regulates the activity of an important enzyme implicated in the RBC defense against oxidation. Larger studies are required to determine how this enzymatic variation may impact clinical heterogeneity in SCD.

3021T

Structural, molecular and cellular impact of the Ogden syndrome Ser3-7Pro mutant N-terminal acetyltransferase Naa10. *G.J. Lyon[1], L. Myklebust[2], P. Van Damme[3], S. Stove[2], N. Reuter[2], M. Doerfel[1], Y. Wu[1], G. Liszczak[4], A. Abboud[2], T. Kalvik[2], C. Grauffel[2], V. Jonckheere[3], H. Kaasa[2], R. Marmorstein[4], K. Gevaert[3], T. Arnesen[2].* 1) Cold Spring Harbor Laboratory, New York City, NY; 2) Department of Molecular Biology, University of Bergen, N-5020 Bergen, Norway; 3) Department of Medical Protein Research, VIB, B-9000 Ghent, Belgium; 4) Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA.

The infantile X-linked lethal Ogden syndrome was the first reported human genetic disorder associated with a mutation in an N-terminal acetyltransferase (NAT) gene. The eight affected and now deceased males in two unrelated families had a Ser37Pro mutation in the gene encoding Naa10, the catalytic subunit of NatA, the major human NAT. NATs transfer an acetyl moiety from acetyl coenzyme A (Ac-CoA) to the primary α-amino group of nascent polypeptides as they emerge from ribosomes, and proteomic analyses have revealed that 50-90% of yeast, plant, fruit fly and human proteins are Nt-acetylated. In order to understand the detrimental impact of Naa10 Ser37Pro in humans, we performed structural, molecular and cellular investigations. Structural modeling and molecular dynamics simulations of the human NatA and its Ser37Pro mutant suggest differences in regions involved in catalysis and at the interface between Naa10 and the auxiliary subunit Naa15. In agreement, biochemical data demonstrate a reduced catalytic capacity and a reduced complex formation of Naa10 Ser37Pro with Naa15 and Naa50, indicating a disruption of the NatA complex. Patient derived Naa10 mutant cells show reduced Nt-acetylation for a subset of NatA-type substrates compared to wild type Naa10 cells. More specifically, N-terminal acetylome analyses of B-cells and fibroblasts reveal an in vivo perturbation of Naa10 (NatA) mediated Nt-acetylation in Ogden syndrome males as some NatA-type substrates are less Nt-acetylated, with some of these substrates previously found to be affected in their Nt-acetylation status by siNatA-mediated knockdown. Ogden syndrome fibroblasts displayed reduced cell proliferation and migration capacity as well as abnormal growth patterns. Introduction of human NatA into yeast that lack endogenous NatA could rescue a specific heat stress phenotype, whereas NatA Ser37P failed to. Additionally, proteomic analyses of these strains revealed a reduced expression of ribosomal proteins in the Ogden strains. The impaired NatA S37P complex formation and catalytic capacity, together with reduced in vivo Nt-acetylation and abnormal cellular phenotypes provide new insights into the underlying molecular and cellular mechanisms of Ogden syndrome. More broadly, these studies broaden our understanding of the role of N-terminal acetylation of proteins, which is a vastly understudied but very prevalent co- and post-translational modification.

3022S

Analyzing KMT2D/MLL2 missense mutations as causative in Kabuki syndrome and testing U1snRNAs-based approaches to revert KMT2D/MLL2 splicing defects. *L. Micale[1], B. Mandriani[1], P. De Nittis[1], B. Augello[1], C. Fusco[1], A. Romano[2], B. Piccini[2], M.T. Pellico[1], C. Rinaldi[1], A. Di Lauro[1], T. Verri[3], G. Merla[1].* 1) Medical Genetics Unit, IRCCS Casa Sollievo Della Sofferenza Hospital, San Giovanni Rotondo, Italy; 2) Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy; 3) Department of Biological and Environmental Sciences and Technologies, University of Salento, Lecce, Italy.

Histone methylation is an epigenetic mechanism by which spatial and temporal expression of distinct genes and pathways are regulated at precise developmental stages. The discovery of histone methyltransferase KMT2D and demethylase KDM6A genetic alterations in Kabuki syndrome patients expanded and highlighted the role of histone modifiers in causing congenital anomalies and intellectual disability syndromes. Kabuki syndrome is a multiple congenital malformation syndrome characterized by facial features, skeletal anomalies, dermatoglyphic abnormalities, mental retardation, and postnatal growth deficiency. KMT2D and KDM6A, play important roles in the epigenetic control of active chromatin states modulating the expression of genes essential for embryogenesis and development. We performed a mutational screening on about 400 Kabuki patients by sequencing, MLPA, and qPCR. Among the KMT2D mutations we identified 49 missense variants across the entire length of the KMT2D gene and 9 splicing site variants that result in a frameshift with the generation of premature stop codons. In this study we ascertain the pathogenicity of KMT2D missense mutations through an integrative analysis of bioinformatics tools and biochemical and cellular assays. We used an innovative in silico approach that combines comparative analysis and motif/domain search tools with threading/homology modeling protocols to predict functional/structural effect of KMT2D missense variants. Due to the huge size of KMT2D gene, we devised a strategy where minigenes carrying KMT2D missense variants were generated. Their biological effects were evaluated by measuring KMT2D methyltransferase activity and KMT2D-mediated target genes expression during neuronal differentiation in NT2/D1 cells silenced for KMT2D. In this study we also used U1 small RiboNuclearProteins (U1snRNAs)-based approaches to restore normal splicing process of KMT2D 5' splice site. This strategy, exploited using minigenes and adapted U1snRNAs complementar to the mutated splice site, allowed us to revert the splicing defects for all the analysed mutations. The experimental approaches proposed offer a valuable support to estimate the real deleterious effect of KMT2D missense variants, a main issue in diagnostic counseling, and evaluate the effectiveness of modified U1snRNAs in reverting KMT2D splicing defects as therapeutic tool.

3023M

Infantile myofibromatosis: Gene discovery leads to novel treatment paradigm based on *PDGFRB* gain-of-function mutations. *K. Oishi[1,2], B.R. Evans[1], T. Silvers[1], J.A. Martignetti[1].* 1) Dept Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Dept Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY.

Background: Infantile myofibromatosis (IMF; MIM 228550) is an autosomal dominant disorder characterized by solitary or multicentric fibrous tumors of skin, muscles, viscera and bones. Tumor growth may begin during fetal development and can be present throughout adult life. Occasionally, these tumors may spontaneously regress without treatment. Recently, our group and another identified germline missense mutations in the platelet-derived growth factor receptor beta gene (*PDGFRB*) as causing the majority of IMF cases. The effect of IMF mutations on PDGFRB signaling and the molecular pathway(s) which are dysregulated and result in IMF tumors are unknown but will need to be understood if a targeted therapy is to be devised. **Objective**: To characterize the molecular functions of the IMF-causative mutant PDGFRB. **Design/methods**: Wild-type (WT) human *PDGFRB* cDNA was cloned into pcDNA3.1 expression vector. Two IMF-causing mutations, R561C and P660T, were introduced using site-directed mutagenesis. Mutant proteins were transiently expressed in COS-7 cells, which lack PDGFRB, and the resultant cell lines were stimulated with PDGF-BB for 30 min after serum starvation. Phospho-PDGFRB and phospho-AKT antibodies were used for western blot. **Results**: Cells expressing mutant PDGFRB proteins had an increase in basal phosphorylation of PDGFRB at Y751 and Y857, compared to cells expressing WT protein. Following PDGF-BB stimulation, all cells demonstrated an increased phosphorylation of PDGFRB, but the IMF mutant cells still maintained markedly higher phosphorylation than controls. While basal AKT phosphorylation was minimal in all cell lines, IMF mutant PDGFRB-expressing cells had increased activation of AKT following PDGF-BB stimulation. Levels of phospho-PDGFRB and phospho-AKT were both consistently higher in P660T than R561C. Imatinib, a tyrosine-kinase inhibitor, suppressed the PDGFRB activation. **Conclusions**: IMF-causing *PDGFRB* mutations result in constitutive gain-of-function (GOF). It was suggested that the mutations induce excessive kinase activity and binding capacities with other signaling molecules through increased phosphorylation of Y857 and Y751, respectively. Of note, there was a ligand-dependent increased AKT activation in the cells expressing mutant PDGFRB. There was GOF strength difference between the two mutations, P660T and R561C. Suppression of PDGFRB activation by imatinib suggests a new therapeutic approach for IMF.

3024T

Biochemistry of UBA1 Mutations that Cause Infantile X-Linked Spinal Muscular Atrophy. *C.D. Balak[1], J.M. Hunter[1], G. D'Urso[2], D. Wiley[2], L. Baumbach-Reardon[1].* 1) Translational Genomics Research Institute, Phoenix, AZ; 2) University of Miami, Miller School of Medicine, Miami, Florida.

Spinal muscular atrophy (SMA) is the leading genetic cause of infantile death. Mutations in the SMN1 gene account for a large percentage of SMA. However, in 2008, Dr. Lisa Baumbach-Reardon's group discovered that a rare and lethal form of X-linked SMA can be caused by mutations in the UBA1 gene. UBA1 is an essential non-redundant gene found highly conserved in all organisms from yeast to man; coding for the Ubiquitin Activating Enzyme-1 (Uba1). Uba1 is the pinnacle enzyme in the Ubiquitin Proteasome System (UPS) that is responsible for the turnover and degradation of most proteins in every living cell. The role of Uba1 in the UPS is to activate and transfer a Ubiquitin (Ub) molecule to an E2 Ubiquitin-conjugating enzyme. These Ubiquitin-E2 complexes, in conjunction with E3 ubiquitin ligases, transfer the activated Ub to targeted proteins "tagging" them for degradation by the proteasome. This process of tagging and degrading proteins is essential for the survival of every cell. Our goal is to understand how mutations in Uba1 alter its biochemical function. This is fundamental in understanding the disease mechanism in XL-SMA and developing treatments for XL-SMA. Uba1 activates Ub in a 2 step process. First, Uba1 adenylates Ub in a reaction that transfers the energy of ATP to Ub. The activated Ub is then covalently bound to a cysteine residue of Uba1 to form a thiolester bond in the second step. The Ub is then transferred to a cysteine residue of an E2 enzyme. To date, all of the mutations that have been found in UBA1 cluster in one domain of Uba1 thought to be important for ATP and Mg2+ binding and adenylation, or for protein-protein interactions. Therefore we have developed a continuous kinetic assay to measure the adenylation activity of wild type (WT) and mutant (MT) forms of Uba1. Our initial experiments demonstrate Uba1 mutants retain adenylation activity, thus suggesting alterations in other catalytic activities or protein-protein interactions. We are investigating the rate of thiolester bond formation between Ub and Uba1, and transthiolation of activated Ub to E2 enzymes. Finally we are measuring interactions between mutant forms of Uba1 and key SMA proteins, Smn and gigaxonin. The results of our studies will identify the causal biochemical defect in XL-SMA and pave the way forward for identifying pharmaceutical targets and treatments for XL-SMA.

3025S

Comparative proteomic analysis of different fragile X syndrome cell lines. *S. Lanni[1], F. Palumbo[1], M. Goracci[1], G. Mancano[1], A. Vitali[2], V. Marzano[3], F. Iavarone[3], F. Vincenzoni[3], M. Castagnola[3], P. Chiurazzi[1], E. Tabolacci[1], G. Neri[1].* 1) Institute of Medical Genetics, Catholic University, Rome, Italy; 2) Institute of Chemistry of Molecular Recognition-UOS Rome, CNR, Rome, Italy; 3) Institute of Biochemistry and Clinical Biochemistry, Catholic University, Rome, Italy.

Fragile X syndrome (FXS), the leading cause of inherited intellectual disability, is caused by absence of the FMRP protein due to expansion over 200 repeats of the CGG tract at the 5' UTR of the FMR1 gene and subsequent DNA methylation. We have already characterized rare individuals of normal intelligence with CGG expansion over 200 repeats without DNA methylation (unmethylated full mutation, UFM), which have relatively normal transcription and translation, and represent the status of FXS cell lines prior to gene silencing. We have compared the three types of cell lines (normal control WT, FXS and UFM fibroblasts) with a proteomic approach in order to demonstrate possible differences that might clarify the mechanisms through which the rare UFM cells remain unmethylated and transcriptionally active. Protein extracts were compared by LC-ESI LTQ Orbitrap MS/MS analysis after mono-dimensional SDS-PAGE and trypsin digestion. Interrogation of the dataset for differential protein expression shows that some metabolic pathways are deregulated in UFM cells when compared to FXS cells. Among these pathways, mitochondrial metabolism (oxidative stress) is of particular interest considering its role in neurodegenerative diseases (like FXTAS) and particularly in epigenetic regulation. The deregulated proteins, specifically mitochondrial SOD (SOD2), were validated by Western blot. The interaction of target mRNAs with FMRP was assessed by RNA immunoprecipitation. Preliminary data suggest that mitochondrial metabolism is likely to have a role in DNA hypomethylation of UFM cell lines. Supported by Telethon Onlus, FRAXA Foundation, Italian Association for fragile X syndrome and Novartis Institute for Biomedical Research.

3026M

Utilization of Cas9/CRISPR to understand the genetic disease mechanism: dissecting the functions of NIPBL in the pathogenesis of Cornelia de Lange syndrome. *K. Izumi[1], M. Bando[1], I. Krantz[2], K. Shirahige[1].* 1) Research Center for Epigenetic Disease, The University of Tokyo, Tokyo, Japan; 2) Division of Human Genetics, The Children's Hospital of Philadelphia, PA, USA.

Cornelia de Lange syndrome (CdLS) is a multisystem developmental disorder characterized by craniofacial dysmorphisms, intellectual disabilities, growth retardation, limb anomalies and several other systemic abnormalities. Heterozygous germline mutations in cohesin structural and regulatory components cause CdLS, and mutations in NIPBL, a cohesin regulatory component, have been identified in nearly 60% of CdLS probands. The cohesin complex possesses multiple roles such as maintaining sister chromatid cohesion during cell division and transcriptional control. Given the absence of premature sister chromatid separation in CdLS patient cell lines, the pathogenesis of CdLS is considered to be associated with abnormal gene expression. To reveal the underlying transcriptome abnormality of CdLS, we have previously performed transcriptome analysis using patient-derived cell lines, however, the influence of genetic background difference cannot be eliminated. To overcome this problem, we created 293FT cell line, a common cell line originating from human embryonic kidney, with NIPBL frameshift mutations using Cas9/CRSPR system. Among the clones screened, we discovered a 293FT clone with bi-allelic NIPBL frameshift mutation in exon 3 (NIPBL null clone). Western blot with N-terminus NIPBL antibody demonstrated the complete absence of full length NIPBL. Interestingly, sister chromatid separation defects were not obvious in this clone, suggesting that for the prevention of premature sister chromatid separation, full length NIPBL is not required. RNA-seq using this clone discovered 3732 genes (out of 35636) whose expression was significantly altered from wild type 293FT line. The expression of many HOX genes was altered in the NIPBL null clone. Since abnormal expression of HOX genes were often seen in CdLS patient samples, this NIPBL null clone recapitulates the some features of transcriptome alteration seen in CdLS. Surprisingly, RNA-seq demonstrated the normal expression level of NIPBL transcript, suggesting the presence of hitherto un-described shorter NIPBL isoform(s). These data suggests that full length NIPBL primarily possesses gene expression regulatory role, however, the full length NIPBL appears not to be required for the maintenance of sister chromatids cohesion during cell division. Our results suggest the differential functional roles of NIPBL isoforms, and full-length NIPBL plays a cardinal role in transcriptional regulation, whose disruption causes CdLS.

3027T

Incontinetia pigmenti: identification of IKBKG/NEMO mutation revealing a novel mechanism of cell death acting as major triggering of the disease. *A. Pescatore[1], E. Esposito[1], G. Calculli[1], G. Courtois[2], M.V. Ursini[1].* 1) Institute of Genetics and Biophysics "A. Buzzati-Traverso" CNR, Via P. Castellino, 111 80125 Naples, Italy; 2) Institut de Recherches en Technologies et Sciences pour le Vivant (iRTSV) CEA Grenoble 17 avenue des Martyrs 38054 Grenoble, France.

Incontinentia pigmenti (IP, MIM#308300) is an inherited neuroectodermal disorder lethal in male and caused in heterozygous female by mutation in the X-linked NEMO/IKBKG, which encodes for a key regulatory protein of the NF-kappaB pathway. Such pathway is involved in many fundamental biological functions regulating the expression of essential proteins for innate and adaptive immunity, inflammation and cell survival. In agreement with the crucial function of NEMO/IKBKG, the embrionic lethality in hemizigous male is considered to be due to massive liver apoptosis. Instead, IP females survive thanks to the skewed X-inactivation that produced counter selection of NEMO mutated cells. Currently it is thought that the sensitivity to apoptosis observed in NEMO-lacking cells in response to TNF is mainly a consequence of impaired transcription of NF-kappaB-dependent anti-apoptotic genes. We will present data on a novel NEMO/IKBKG mutation found in a female affected by a classical although severe form of IP associated to microcephaly, brain atrophy and mental retardation. In vitro analysis of Nemo (-) complemented with the NEMO/IKBKG mutation revealed that these cells prevented NF-kappaB activation upon TNF treatment. Unexpectedly, NEMO-mutant cells were only poorly sensitive to TNF-induced cell death when compared to Nemo (-) cells, while they were very sensitive to TNF-induced necroptosis. We will illustrate how the presence of mutated-NEMO could affect the molecular mechanisms governing the balance between apoptosis and necroptosis, following TNF-receptor stimulation, paying a special attention to the role of NEMO to this regard. These findings open new avenues for future treatment of IP severe form.

3028S

The *cblX* gene, *HCFC1*, regulates craniofacial development by modulating *MMACHC* expression. *T.H. Shaikh, E.A Geiger, B. Appel, A.M. Quintana.* Dept. of Pediatrics, University of Colorado School of Medicine, Aurora, CO.

Mutations in *HCFC1* (MIM300019), which encodes a transcriptional co-regulator, have been recently associated with methylmalonic acidemia and homocysteinemia *cblX* type (MIM309541), an X-linked, recessive disorder characterized by defects in cobalamin metabolism. The metabolic imbalance results from the aberrant transcriptional regulation of *MMACHC* (MIM609831) expression, a downstream target of HCFC1 known to be involved in cobalamin metabolism. Mutations in *MMACHC* cause a highly related disorder, commonly referred to as *cblC*. Both *cblX* and *cblC* have overlapping clinical manifestations including developmental delay, microcephaly, and failure to thrive, but individuals with *cblX* are additionally characterized by severe neurological defects that primarily include intractable epilepsy. According to global ChIP-seq analysis there are over 5000 genes, including *MMACHC*, whose expression is potentially regulated by HCFC1. Thus, it is unclear if all of the clinical manifestations observed in *cblX* are mediated by aberrant expression of *MMACHC* or if some of them result from the dysregulation of other target genes. To begin to elucidate the mechanism by which HCFC1 modulates disease phenotypes, we have carried out loss of function analyses in the developing zebrafish. Of the two *HCFC1* orthologs in zebrafish, *hcfc1a* and *hcfc1b*, the loss of *hcfc1b* specifically led to defects in craniofacial development. Subsequent analysis revealed that *hcfc1b* is essential for the expression of *col2a1* and *sox9a*, two genes that are essential for neural crest cell differentiation into chondrocyte progenitors. Further, the *hcfc1b*-mediated craniofacial abnormalities were rescued by expression of exogenous human *MMACHC*. These data raise the possibility that HCFC1 regulates craniofacial development by modulating expression of *MMACHC*. Furthermore, distinct *HCFC1* mutations demonstrated variable effects on *MMACHC* expression in humans and craniofacial development in zebrafish. Notably, several individuals with mutations in either *HCFC1* or *MMACHC* have been reported to have mild to moderate facial dysmorphia. Thus, our data demonstrate that *HCFC1* plays a role in craniofacial development, which is in part mediated through the regulation of *MMACHC* expression. Further, these observations highlight facial dysmorphism, a previously under-reported phenotype, as a potential outcome of mutations in both *HCFC1* and its downstream target, *MMACHC*.

3029M

Disruption of *FMR1*, *RAI1*, *MBD5*, and *TCF4* results in abnormal expression of circadian rhythm genes and sleep disturbances in fragile X, Smith-Magenis, 2q23.1 deletion, and Pitt-Hopkins syndromes. *S.V. Mullegama[1], L. Pugliesi[2,3], B. Burns[2], R. Tahir[2], Z. Shah[2], Y. Gu[2], D.L Nelson[1], W-H. Tan[4], S.H. Elsea[1,2].* 1) Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Human and Molecular Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA; 3) St. Joseph's Regional Medical Center, Paterson, NJ; 4) Division of Genetics, Boston Children's Hospital; Harvard Medical School, Boston, MA USA.

Sleep disturbance is prevalent in neurodevelopmental disorders such as fragile X syndrome (FXS, *FMR1*), Smith-Magenis syndrome (SMS, *RAI1*), 2q23.1 deletion syndrome (del 2q23.1, *MBD5*), and Pitt-Hopkins syndrome (PTHS, *TCF4*). The frequent comorbidity of sleep disturbance with neurodevelopmental disorders implies overlapping pathways exist that lead to shared sleep phenotypes. We hypothesize that sleep phenotypes in FXS, SMS, del 2q23.1, and PTHS have common etiologies and that FMRP, RAI1, MBD5, and TCF4 function in overlapping circadian pathways that lead to these sleep disturbances. While sleep disturbance is well described in SMS and FXS, we sought to delineate the sleep anomalies in del 2q23.1 and PTHS. Through caregiver surveys and review of the literature, we found that difficulty maintaining sleep was a feature of both disorders. Furthermore, del 2q23.1 patients have difficulty falling asleep and experience night or early awakenings, while PTHS patients have frequent night awakenings. Thus, common sleep issues in these 4 disorders include reduced total sleep time and frequent nighttime and early morning awakenings. We then evaluated expression of the causative genes for these 4 neurodevelopmental disorders in patient lymphoblastoid cell lines (LCLs), which revealed altered levels of *MBD5*, *TCF4*, *RAI1*, and *FMR1* across each of the 4 syndrome cell lines, suggesting a possible common etiology in all four syndromes. To support involvement of these genes in circadian gene expression, we identified putative E-boxes in *MBD5*, *RAI1*, and *FMR1*, while TCF4 regulates gene expression at E-box containing promoters. Expression levels of circadian genes (*NR1D2* and *CRY2*) were reduced in patient LCLs, suggesting key roles for MBD5, TCF4, RAI1, and FMRP in the regulation of circadian gene expression. *MBD5*, *RAI1*, and *TCF4* siRNA knockdown in SH-SY5Y cell lines showed that levels of *CLOCK*, *PER1*, *CRY2*, *ATF2*, and *ATF4* were altered, supporting a role for *MBD5*, *RAI1*, and *TCF4* in circadian gene expression. Pathway analysis of *MBD5*, *RAI1*, and *TCF4* knockdown microarray data show dysregulation of pathways associated with sleep disturbance, including circadian rhythm signaling and mTOR signaling pathways, corroborating published *FMR1* studies, while *mTOR* expression was altered in LCLs from all 4 disorders. These data demonstrate that MBD5, TCF4, RAI1, and FMRP function in overlapping circadian rhythm pathways and point to a common etiology for sleep disturbance in these disorders.

3030T

ITM2B implicated in familial dementias and retinal dystrophy associates with ciliary-centrosomal protein TOPORS. *B. Czub, A.Z. Shah, P. Kruczek, G. Alfano, C. Chakarova, S.S. Bhattacharya.* Institute of Ophthalmology, University College London, London, United Kingdom.

Purpose: *TOPORS* (MIM 609507) is a ubiquitously expressed gene implicated in autosomal dominant retinitis pigmentosa (RP [MIM 268000]); mutations are known to cause only RP with no systemic symptoms. We performed yeast two-hybrid (Y2H) screens of human retinal cDNA libraries to identify proteins interacting with TOPORS, which could explain the restricted phenotype. This led to the isolation of a soluble fragment of integral membrane protein 2B (*ITM2B* [MIM 603904]), mutations in which are associated with inherited forms of dementia and, most recently, a dominant form of retinal dystrophy. **Methods**: Matchmaker™ Gold Y2H System (Clontech) was used for library screening and testing direct protein-protein interactions (PPIs). Results were validated in HeLa and hTERT-RPE1 (RPE1) cell extracts by co-immunoprecipitation. Immunofluorescence methods were used for co-localisation studies in RPE1 cells and mouse retina cryo-sections. ProteoExtract® Subcellular Proteome Extraction Kit (Merck) was used to prepare cellular fractions (HeLa). **Results**: The TOPORS-ITM2B PPI, identified by screening, was selected for further study due to the role of *ITM2B* in retinopathy. It was detected in complexes with TOPORS precipitated from HeLa and RPE1 extracts, and both proteins co-localised at the centrosome in RPE1 cells. In mouse retina sections the strongest ITM2B signal was observed in rod photoreceptor inner segment, with a weaker signal in the ganglion cell layer. Direct PPI experiments in yeast with artificial TOPORS fragments (corresponding to different protein domains) implied that its mutation hotspot region may be required for association with ITM2B. Interaction with the TOPORS' RING finger region was also consistently observed, suggesting ITM2B could be modified by ubiquitination. Preliminary Western blot results suggest differential expression of proteolytically cleaved ITM2B peptides in cellular fractions. **Conclusions**: ITM2B is a novel interacting partner of TOPORS; defining the cellular processes in which they both associate could explain their roles and the phenotypes resulting from their mutations. Furthermore, evaluation of their relationship could lead to a better understanding of neurodegeneration mechanisms overall. Subsequent work will aim to delineate the roles of the various ITM2B peptides, generated by physiological proteolytic cleavage of the membrane-bound precursor protein, which could clarify the centrosomal co-localisation with TOPORS.

3031S

Disruption of the basal body protein POC1B results in autosomal recessive cone-rod dystrophy. *S. Roosing[1,2,3,7], I.J.C. Lamers[2,3,7], E. de Vrieze[2,4,7], L.I. van den Born[5,7], S. Lambertus[6], H.H. Arts[2,3], T.A. Peters[4], C.B. Hoyng[6], H. Kremer[2,3,4], L. Hetterschijt[2], S.J.F Letteboer[2,3], E. van Wijk[4,8], R. Roepman[2,3,8], A.I. den Hollander[2,3,6,8], F.P.M. Cremers[2,3,8], Study group.* 1) Department of Neurosciences, University of California San Diego, La Jolla, San Diego, CA; 2) Department of Human Genetics, Radboud university medical center, Nijmegen, The Netherlands; 3) Radboud Institute for Molecular Life sciences, Radboud University Nijmegen, Nijmegen, The Netherlands; 4) Department of Otorhinolaryngology, Radboud university medical center, Nijmegen, The Netherlands; 5) The Rotterdam Eye Hospital, Rotterdam, The Netherlands;; 6) Department of Ophthalmology, Radboud university medical center, Nijmegen, The Netherlands; 7) Shared first author; 8) Shared last author.

Exome sequencing revealed a homozygous missense mutation (c.317C>G, p.Arg106Pro) in *POC1B*, encoding POC1 centriolar protein B, in three siblings with autosomal recessive cone dystrophy or cone-rod dystrophy, and compound heterozygous *POC1B* mutations (c.199_201del [p.Gln67del] and c.810+1G>T) in an isolated case with cone-rod dystrophy. Upon overexpression of POC1B in human TERT-immortalized retinal pigment epithelium 1 cells, the encoded wild-type protein localized to the basal body of the primary cilium, while this localization was lost for p.Arg106Pro and p.Gln67del variant POC1B proteins. Morpholino oligonucleotide induced knock-down of poc1b translation in zebrafish resulted in a dose-dependent small eye phenotype, impaired optokinetic responses and decreased photoreceptor outer segment length. These ocular phenotypes could partially be rescued by wild-type human *POC1B* mRNA, but not by c.199_201del and c.317C>G mutant human *POC1B* mRNAs. Yeast two-hybrid screening of a human retinal cDNA library revealed FAM161A as a binary interaction partner of POC1B. This was confirmed in coimmunoprecipitation and co-localization assays, which both showed loss of FAM161A interaction of p.Arg106Pro and p.Gln67del variant POC1B. FAM161A was previously implicated in autosomal recessive retinitis pigmentosa and shown to be located at the base of the photoreceptor connecting cilium where it interacts with several other ciliopathy-associated proteins. Taken together, this study demonstrates that *POC1B* mutations result in a defect of the photoreceptor sensory cilium affecting the cone and rod photoreceptors.

3032M

TOPORS, a centrosomal and ciliary protein implicated in Retinitis Pigmentosa, associates with the actin cytoskeleton. *A.Z Shah, B. Czub, P.M. Kruczek, G. Alfano, C.F. Chakarova, S.S. Bhattacharya.* Department of Genetics, Institute Ophthalmology, UCL, London, United Kingdom.

Purpose: *TOPORS* (MIM 609507) is a ubiquitously expressed gene, encoding a protein localising to the nucleus and centrosome of dividing cells, and to the basal body of ciliated cells (including connecting cilium of photoreceptor cells in the retina). Interestingly, mutations in *TOPORS* are only implicated in dominant retinitis pigmentosa (MIM 268000). This study was undertaken to better understand the role of TOPORS in the cell. **Methods:** Subcellular localisation of TOPORS was examined in hTERT-RPE1 (RPE1) cell line by confocal immunofluorescence in cultures of increasing confluence. Stress experiments were performed to test observations. Co-immunoprecipitation was performed on RPE1 total cell extract; cellular fractions were prepared using ProteoExtract® Subcellular Proteome Extraction Kit (Merck); both analysed by Western blotting. RT-PCR on cDNA from human cell lines and retina was performed to identify alternatively spliced variants. TOPORS antibody was validated using RNAi knockdown. All experiments were performed on endogenous proteins. **Results:** Confluence studies showed TOPORS localises primarily to the nucleus but is found in the cytoplasm, centrosome and close to the cell membrane with increasing cell density. Furthermore it localises along actin filaments. Stress conditions (such as changes in osmolarity) do not appear to effect TOPORS localisation. Results suggest TOPORS could interact with actin and other cytoskeletal components. There is one alternatively spliced isoform reported for TOPORS. The commercially available TOPORS antibody is able to detect both long and short isoforms, as confirmed by RNAi knockdown experiments. The short isoform for TOPORS was detected in all cell lines tested and in human lymphocytes, and could account for the cytoplasmic localisation observed. However, to date it has not been detected in human retina. Preliminary evidence suggests differential expression of both TOPORS isoforms in different cellular compartments. **Conclusions:** In addition to being a nuclear and centrosomal protein, TOPORS is a novel component of the actin cytoskeleton. This could be due to the short TOPORS isoform. Our findings suggest a potential role of TOPORS in mediating actin-related cellular changes through its E3 ubiquitin- and/or SUMO-1 ligase activities and could indicate a role in cell signalling. This work highlights the increasing complexity of protein function and the importance of understanding their role(s) in health and disease.

3033T

Loss of the ribosome-associated factor Sbds in murine models of Shwachman-Diamond syndrome leads to aberrant polysome profiles. *H. Liu[1, 2], M.E. Tourlakis[1, 2], R. Gandhi[1], P. Hu[3], J.M. Rommens[1, 2].* 1) Program in Genetics & Genome Biology, SickKids Research Institute, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Department of Biochemistry & Medical Genetics, University of Manitoba, Winnipeg, Manitoba, Canada.

Shwachman-Diamond syndrome (SDS) is an autosomal recessive disease characterized by growth retardation, exocrine pancreatic dysfunction, skeletal dysplasia, cognitive impairment and hematological abnormalities with increased susceptibility for leukemia. SDS is caused by mutations in *SBDS*. A current model proposes that SBDS/Sbds functions together with EFTUD1/Eftud1 and is involved in the release of EIF6/Eif6 from the pre-60S complex, enabling translation initiation. Mouse embryos deficient for Sbds ($Sbds^{R126T/R126T}$) show notable growth retardation and cell culture models reveal reduced protein synthesis. To assess translation deficiency, polysome profiles of fetal organs from this SDS murine model (E18.5) were examined and found to show anomalies compared to matched controls (*Sbds*+/+ and $Sbds^{+/R126T}$), with reduction in 80S peaks as well as preservation of polysomes. The low 80S peak suggested impairment of translation initiation; however the prominent polysomes were unexpected and appeared inconsistent with deficiency in protein synthesis. To investigate how the translatome is affected in SDS, total and polysomal mRNAs of mutant and control fetal liver samples were studied using cDNA microarray analyses. By comparing individual polysomal transcript to respective total transcript level, we could establish which genes were being actively translated and/or retained by the polysomes with loss of Sbds function. With normalization, 173 genes showed reduced association with polysomes, while 664 showed increased association. Typically, long transcripts with lower GC content were favored for polysome association. However, the changes of polysome association did not correspond to respective steady state protein levels, as shown by western immunoblotting and label-free mass spectrometry. Further, the proteome revealed that relative levels of ribosomal proteins and translation factors are unaltered in SDS fetal livers. The observations of prominent polysomes and the large number of transcripts with high polysome association in absence of corresponding increase in the protein levels suggest that loss of Sbds function leads to the formation of ribosomes which may be deficient in elongation or ribosome recycling, possibly due to impaired removal of Eif6 prior to subunit joining.

3034S

Examining the Molecular Mechanisms Underlying SRCAP Mutations in Floating-Harbor Syndrome. *R.L. Hood[1,2], K.M Boycott[2,3,4], W.L. Stanford[1,5,6], D.E Bulman[2,4].* 1) Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Ontario, Canada; 2) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa; 3) Department of Genetics, Children's Hospital of Eastern Ontario; 4) Department of Pediatrics, University of Ottawa; 5) The Sprott Centre for Stem Cell Research, Regenerative Medicine Program, Ottawa Hospital Research Institute; 6) Department of Cellular and Molecular Medicine, University of Ottawa.

Floating-Harbor Syndrome (FHS; OMIM# 136140) is a rare genetic disorder which is characterized by short stature, delayed osseous maturation, language deficits, and unique dysmorphic facial features. FHS typically occurs sporadically; however a few instances of autosomal dominant transmission have been reported. Since our first report of mutations in SRCAP as the cause of FHS, we have identified a total of 50 mutation positive patients. The mutations identified were shown to be de novo in all cases where parental DNA was available. SRCAP encodes a SNF2-related chromatin-remodeling factor which is known to be a coactivator of CREB-binding protein (CREBBP or CBP) and CBP-mediated transcription. Additionally SRCAP is part of a large complex which catalyzes ATP-dependent substitution of the variant histone H2A.Z into nucleosomes. Based on the existence of patients with deletions encompassing all of SRCAP who do not have FHS, we postulate that the mechanism of disease is not due to haploinsufficiency. Furthermore we expect that FHS mutations result in widespread gene dysregulation; however the mechanisms and biological pathways underlying the FHS phenotype are currently unknown. We are directing our efforts at determining the mechanism by which truncating mutations of SRCAP cause disease. Sequencing of FHS patient cDNA indicated the presence of SRCAP mutant transcript and the relative abundance of this transcript was found at relatively comparable levels to the wild type. RNAseq analyses were used to compare expression differences in FHS patients compared to non-patient controls. Initial findings support the hypothesis that FHS is not caused by a haploinsufficient mutation mechanism. To more specifically examine the growth deficiency phenotype of FHS we reprogrammed FHS patient and gender matched control fibroblast samples into induced pluripotent stem cells (iPSCs) using an episomal vector approach. iPSCs were subsequently characterized by pluripotency cell surface marker expression, in vitro embryoid body differentiation, and in vivo teratoma formation assays. We plan on differentiating these clones into chondroprogenitors so we can begin elucidating the molecular mechanisms underlying the bone pathogenesis of FHS. Taken together these studies will increase our understanding of how truncating mutations of SRCAP alter its function and gain insight into the pathways which contribute to the development of Floating-Harbor syndrome.

3035M

Functional characterization and pharmacological correction of eleven novel mutations identified in Indian CF population. *R. Prasad[1], H. Sharma[1,2], M. Jollivet Souchet[2], F. Becq[2].* 1) Department of Biochemistry, PGIMER, Chandigarh, Chandigarh, India; 2) Laboratoire Signalisation et Transports Ioniques Membranaires, Université de Poitiers, CNRS, Poitiers, France.

Cystic Fibrosis an autosomal recessive disorder is usually considered as rare disease for Indian population hence much less is known about this disease in context with Indian sub continent. Recently we have established a spectrum of mutations in CFTR gene from classical Cystic Fibrosis as well as from infertile male patients with CBAVD/CUAVD in Indian populations. Among them S549N, L69H in classical CF and G126S, F87I, S118P, H139Q, F157C, F494L, E543A, Y852F, D1270E in CBAVD males were among the rare missense mutations (Sharma et al ., 2009). In this study we have attempted to conduct in vitro gene expression analysis to well establish genotype and phenotype correlation and to characterize these four rare missense mutations according to the mechanism that disrupt CFTR protein function. All four mutations from Indian population were characterized by expressing pEGFP-CFTR constructs in BHK-21 cells via 3 step technique viz; CFTR cellular localization was determined by confocal microscopy, where as Western blot analysis and automated iodide efflux assays was used to determine CFTR maturation processes and its channel activity respectively. In Western blot analysis only b-band is obtained for L69H substitution as for F508del whereas in the case of other mutants both b and c bands were found, indicating L69H mutation impair CFTR maturation process, the finding was again confirmed by confocal imaging. Iodide Efflux assay revealed significant decrease in channel activity for L69H and S549N mutants CFTR expressing cells in comparison to WT, although this decrease in channel activity was rescued when cells were incubated at 27 C. When the effect of CFTR correctors was checked on different mutants, it was found that VX809 significantly ameliorate the defect caused by L69H mutation. Mutations G126S, F87I, S118P, H139Q, F157C, F494L, E543A, Y852F, D1270E have no impact on CFTR maturation and function. Thus we can conclude that L69H mutation is a class II CF mutation causing impaired maturation leading to protein degradation and Cl^-ions impermeability as observed in F508del mutation. This defect is rescued by the corrector VX 809. Whereas S549N mutation can be categorized into a class II/III mutation causing impaired maturation and reduced channel activity.

3036T

Understanding the role of EYS by identification and characterisation of its retinal interacting partners. *P.M. Kruczek[1], G. Alfano[1], B. Czub[1], A.Z. Shah[1], A.C. Zelhof[2], S.S. Bhattacharya[1].* 1) Department of Genetics, UCL Institute of Ophthalmology, London, United Kingdom; 2) Department of Biology, Indiana University, 1001 East Third St, Bloomington, IN 47405, USA.

Purpose: The eyes shut homolog (Drosophila) gene (*EYS* [MIM 612424]) is one of the largest genes known to be expressed in the human retina. Mutations in *EYS* are the most common cause of autosomal recessive retinitis pigmentosa (arRP) (RP25 [MIM 602772]; RP [MIM 268000]). *EYS* encodes a protein named EYS whose biological role in humans is presently unclear, however, a *Drosophila* ortholog of EYS has been shown to be essential for the organization of *Drosophila* photoreceptors. In addition, *Drosophila* EYS interacts with Prominin, a transmembrane protein associated with microvilli of the *Drosophila* photoreceptor rhabdomere. Prominin is highly conserved throughout evolution and interestingly, mutations in human Prominin-1 (*PROM1* [MIM 604365]) cause retinopathies. This study was undertaken to investigate the relationship of EYS and Prominin-1 in humans, and to identify novel interacting partners for EYS in the retina. **Methods:** Putative interaction of EYS and Prominin-1 was examined by immunochemistry in cultured Y79 and ARPE19 cell lines. Novel potential binding partners of EYS were identified using Yeast 2-hybrid technology (Y2H) - baits, consisting of full-length EYS and three fragments covering N-terminal, middle and C-terminal parts of the protein, were screened against human retinal cDNA libraries. **Results:** Preliminary data suggest that EYS and Prominin-1 may co-localise in Y79 cell line. Furthermore, exogenous expression of EYS and Prominin-1 in ARPE19 cell line suggests that Prominin-1 might be required for proper localisation of EYS. Y2H screens revealed six proteins that potentially interact with EYS: AIPL1, SERF2, FBN1, PPT2, UBA52 and COX7C. **Conclusions:** Investigation of interaction between EYS and Prominin-1 suggests that these proteins may possess similar functions to their *Drosophila* orthologs and could play a role in maintaining the integrity of human photoreceptors. Y2H screens have identified six potential interactors of EYS which may help us elucidate its role in the retina. Additional investigation is required to validate and further analyse the aforementioned results. This will include verification using direct interaction in the Y2H system as well as experiments such as co-immunoprecipitation and immunocytochemistry. Not only is this crucial to the understanding of the role of EYS in health and disease, but it will also shed light on physiology and maintenance of human photoreceptors.

3037S

A pathologic genomic rearrangement in Incontinentia Pigmenti locus reveals a transcriptional regulation of NEMO gene by p63 family proteins. *M.I. Conte[1], M. Paciolla[1], E. Candi[2], A. Lena[2], G. Melino[2], M.V. Ursini[1], F. Fusco[1].* 1) Institute of Genetics and Biophysics, Naples, Italy; 2) Universita' di Tor Vergata, Rome, Italy.

Incontinentia pigmenti (IP, MIM308300, 1/10.000) is an X-linked dominant neuroectodermal disorder, lethal in males and caused in female by a mutation in the IKBKG/NEMO gene that encodes for NEMO/IKKgamma regulatory protein required for the activation of the NF-kappaB pathway. The typical IP skin defects, always present in IP patients, are due to the mosaic expression of NEMO/IKBKG mutation related to X-chromosome inactivation in heterozygous females. NEMO/IKBKG is transcribed by the bidirectional promoterB, shared by the two overlapping IKBKG/NEMO and G6PD genes and promoterA, unidirectional for NEMO expression. We recently described an IP de novo case due to a genomic deletion, named IKBKGdel, that eliminates the NEMO principal promoterB and preserves its expression by the secondary promoterA. The patient with IKBKGdel, showed only skin abnormalities suggesting that an early X-inactivation of mutated allele occurred in the other extracutaneous tissues, or the promoterA activity of NEMO prevented the IP alterations. We analysed promoterA transcription in vivo on RNA from different layers of the skin from healthy donors and in vitro on RNA from differentiated and proliferating HaCat keratinocyte to obtain a better insight into its role in the skin. Bioinformatic sequence comparison using the multiple alignment software m-Vista of the 5' of the NEMO promoter A in different species predicted several p53-like binding sites in the highly conserved regions. Here, we present our findings by in vivo and in vitro assays on the different role of p53-related proteins, p63 and p73, on promoterA-driven NEMO expression. Our data suggest that members of p63 family, DNp63alpha and TAp63alpha, which are key transcription factors in epithelial development and differentiation, are able to regulate the expression of promoter A of NEMO during the skin differentiation. Moreover, we found that the isoform TAp73alpha while regulating G6PD expression has not effects on NEMO transcription. The differential effects of p53-related proteins, p63 and p73 on NEMO and G6PD will be illustrated.

3038M

Chinese family segregating isolated diffuse oesophageal leiomyomatosis: a new COL4A5/COL4A6 deletion and a case of gonosomal mosaicism. *K.L. Wong[2], W. Liu[1], H.E. Qiuming[1], H.M. Wong[2], S.T. Tang[2,3], R. Zhang[3], M.T. So[3], K.Y. Wong[3], J. Nicholls[4], S. Cherny[2,5], P.C. Sham[2,5], P.K. Tam[3,5], M.M. Carcia-Barcelo[3,5], H. Xia[1].* 1) Guangzhou Women and Children's Medical Center, Guangzhou, China; 2) Department of Psychiatry, The University of Hong Kong, Hong Kong SAR, China; 3) Department of Surgery,The University of Hong Kong, Hong Kong SAR, China; 4) Department of Pathology,The University of Hong Kong, Hong Kong SAR, China; 5) Center for Genomic Sciences,The University of Hong Kong, Hong Kong SAR, China.

Diffuse oesophageal leiomyomatosis (DOL) is a rare disorder characterized by tumorous overgrowth of the muscular wall of the oesophagus. DOL is present in 5% of Alport syndrome (AS) patients. AS is a rare hereditary disease that involves varying degrees of hearing impairment, ocular changes and progressive glomerulonephritis leading to renal failure. In DOL-AS patients, the genetic defect consists of a deletion involving the COL4A5 and COL4A6 genes on the X chromosome. We present a two-generation family with two members (mother and son) affected with DOL and none of the features of AS. Exome sequencing and Copy Number Analyses revealed a new 40kb deletion encompassing from intron 2 of COL4A5 to intron 1 of COL4A6 at Xq22.3. The breakpoints were also identified. Possible confounding pathogenic exonic variants in genes known to be involved in other extracellular matrices disorders were also shared by the two affected individuals. Meticulous analysis of the maternal DNA revealed a case of gonosomal mosaicism. This is the first case where a "classic" DOL-AS COL4A6 deletion is described in patients with no AS and the first report of gonadosomal mosaicism associated to this condition. COL4A5/COL4A6 genes should also be investigated in cases of isolated DOL.

3039T

Atypical Microvillous Inclusion Disease (MVID) in a newborn with intractable diarrhea: clinical, pathological and molecular characterization. *A. Iglesias[1], D. Ball[1], S. Wontakal[2], V. Aggarwal[2].* 1) Division of Medical Genetics, Department of Pediatrics, Columbia University Medical Center, New York, NY; 2) Department of Pathology & Cell Biology, College of Physicians & Surgeons of Columbia University, New York, NY.

Microvillus inclusion disease (MVID) is characterized by intractable life-threatening watery diarrhea. Treatment includes total parental nutrition (TPN) or small bowel transplantation, both associated with high morbidity and mortality. Electron microscopy (EM) of the small bowel typically displays characteristic loss of microvilli as well as pathognomonic microvillus inclusion bodies in intestinal epithelial cells. Mutations in the *MYO5B* gene, which encodes a type Vb myosin motor protein, are associated with MVID. We report an 8 month-old male that presented at birth with intractable watery diarrhea. He was born prematurely at 35 weeks to a 30 year old mother by NSVD with APGARS of 6/8. Birth weight was 2850 grams. Diarrhea started on day 1 of life. Despite being properly fed since birth his weight on day 7 was 1900 grams. He was found to be severely dehydrated and in renal failure with a hyperchloremic metabolic acidosis with a normal anion gap. Metabolic evaluation and newborn screening were normal. An initial endoscopic intestinal biopsy revealed non-diagnostic alterations, however, based on clinical presentation, a primary intestinal disorder was strongly suspected. An immunological evaluation with normal results was done. To determine a possible molecular etiology for the intractable congenital diarrhea, whole exome sequencing (WES) was performed. WES identified a paternally inherited variant (c.4168C>T;p.Q1390X) and a maternally inherited alteration (c.4852+3A>T) in the *MYO5B* gene. These variants have not been previously reported in patients with MVID; however, given this intriguing finding, EM biopsy samples were re-examined. Re-evaluation of the EM images identified rare atypical findings in the small intestine, but unexpectedly, atypical findings in the colon were also noted. These findings were interpreted as consistent with the clinical presentation and therefore likely represent an atypical case of MVID. Currently we are conducting RNA studies to confirm the pathogenicity of these changes. The patient is currently on TPN and trophic enteral feedings. He is thriving and doing well, however, long-term prognosis is uncertain and the possibility of small bowel transplant has been already discussed with the family. In summary, we are presenting clinical, pathological and molecular characterization of an atypical case of MVID. This case expands the phenotype of this rare disorder and highlights the utility of WES in clinical practice.

3040S

Mutations in ACTG2 are associated with sporadic congenital chronic intestinal pseudo-obstruction and megacystis-microcolon-intestinal hypoperistalsis syndrome. *M. Yourshaw, A. Sicolo, R. Venick, L. Reyen, D. Farmer, J. Vargas, S. Nelson, M. Martín.* Pediatrics, UCLA Geffen School of Medicine, Los Angeles, CA.

Chronic intestinal pseudo-obstruction (CIPO) is a serious motility dysfunction syndrome characterized by symptoms of intestinal obstruction in the absence of any mechanical blockage. It is a major cause of intestinal failure. With whole-exome sequencing in a cohort of 20 patients with congenital CIPO or MMIH, we identified a subset of 10 cases with potentially damaging de-novo mutations at highly conserved loci in the ACTG2 gene, which encodes actin, gamma-enteric smooth muscle precursor. We conclude that ACTG2 governs a significant proportion of cases of sporadic congenital CIPO and MMIH.

3041M

Functional analysis of genes carrying de novo mutations in 24 sporadic Hirschsprung cases revealed 7 unexpected genes relevant to ENS development. *R. Hofstra[1,2], D. Schriemer[3], H. Gui[4], W. Cheng[4], P. Griseri[5], A. Pelet[6], M. Ruiz-Ferrer[7], C. Berrios[8], W. van Ijcken[9], M. van den Hout[9], Y. Sribudiani[1], R. Chaudan[1], P. Tam[4], C. Tang[4], I. Matera[5], I. Antinolo[7], A. Chakravarti[8], S. Borrego[7], S. Lyonnet[6], I. Ceccherini[5], B. Eggen[3], J. Amiel[6], M. Garcia-Barcelo[4].* 1) Clinical Genetics, ErasmusMC, Rotterdam, Netherlands; 2) Institute of Child Health, UCL LOndon UK; 3) University Medical Centre Groningen, Netherlands; 4) University of Hong Kong, China; 5) Instituto G. Gaslini, Genoa, Italy; 6) Universite Paris5-Descartes, Paris France; 7) Hospital Universptario, Virgen del Rocio, Sevilla, Spain; 8) Johns Hopkins University, Baltimore MD USA; 9) Biomics, EramusMC, Rotterdam, Netherlands.

Hirschsprung disease (HSCR) is a complex genetic disorder, characterized by the absence of the enteric neurvous system (ENS) of the distal part of the gastrointestinal-tract. HSCR most commonly presents sporadically, however it can be familial (~20% of the patients). The sporadic form of the disorder is believed to be genetically complex. To assess the role of de novo variations in sporadic HSCR, we performed exome sequencing on DNA of 24 sporadic HSCR patients and their unaffected parents. In total we confirmed 26 de novo variations (23 SNV, 3 Indels) in 19 genes. Eight variants were present in RET, the major HSCR gene. The remaining 18 genes all carried one de novo variant and none of the genes could directly be linked to any of the known HSCR associated gene network. Knockdown experiments in zebrafish, by both splice- and translation-blocking morpholino's, resulted in a HSCR like phenotype for 7 of these 18 genes (CKAP2L, DENND3, NCLN, NUP98, TBATA, MED26, MAP40). The zebrafish work is complimented by proliferation, migration and differentiation assays using stable, siRNA expressing neural crest derived cell lines. Our data show that de novo, as well as inherited mutations, contribute to the development of the ENS and thereby to HSCR. Moreover, as the identified 7 novel HSCR genes are all unrelated to the previous gene networks, our study suggests that functional analysis of all genes carrying de novo mutations, is warranted to genetically delineate the genetic architecture of a genetically complex diseases.

3042T

Identification of a novel gene causing a recognizable and distinct autosomal recessive cerebellar ataxia and intellectual disability syndrome, associated with early onset cerebellar atrophy and relative macrocephaly. *P. Stanier[1], A.C. Thomas[1], H. Williams[1], N. Setó-Salvia[1], C. Bacchelli[1], M. O'Sullivan[1], L. Ocaka[1], K. Mengrelis[1], M. Ishida[1], G. Anderson[2], D. Morough[2], M. Rytan[3], J.M. Saraiva[4], F. Ramos[4], B. Farren[2], D. Saunders[2], P. Gissen[1], A. Straatmaam-Iwanowska[1], F. Baas[5], N. Wood[3], R. Robinson[2], J. Hersheson[3], H. Houlden[3], R. Hennekam[5], J. Hurst[2], R. Scott[2], M. Bitner-Glindzicz[1,2], G. Moore[1], S.B. Sousa[1,4].* 1) University College London, Institute of Child Health, London, United Kingdom; 2) Great Ormond Street Hospital, London, United Kingdom; 3) University College London, Institute of Neurology, London, United Kingdom; 4) Hospital Pediátrico de Coimbra, Portugal; 5) University of Amsterdam, Holland.

A large number of conditions have been described where severe intellectual disability and ataxia are found in patients with cerebellar hypotrophy. Individually, most types are rare and without known molecular pathology. Here we describe three similar but unrelated consanguineous families with presumed autosomal recessive inheritance from Turkey and Portugal. The seven affected individuals share overlapping phenotypes, including severe intellectual disability, ataxia, early onset cerebellar atrophy, relative macrocephaly, coarse face, hearing loss, absent speech and 5th finger clinobrachydactyly and camptodactyly. To facilitate gene identification, two families with multiple affected individuals were selected for homozygosity mapping followed by exome sequencing. In family 1, a unique homozygous nonsense variant was identified in a novel gene contained within a large region of shared homozygosity. The variant was not subject to nonsense-mediated decay but produced reduced transcript levels and a truncated protein. The same gene was also found to contain a homozygous, multiexon deletion within the largest region of homozygosity in family 2. Compared to controls, this mutation resulted in normal transcript levels in patient fibroblasts, although a correspondingly truncated protein was predicted. The third case was selected from a large cohort of patients purely based on the phenotypic characteristics of the two index families. Sanger sequencing in this patient for the novel gene revealed a homozygous splice site mutation, which was demonstrated to result in an in-frame skipping of an exon that was present in all isoforms. Transcript levels were similar to those found in controls and a truncated protein was predicted to be present. This third family supports the proposal that the phenotype is sufficiently distinct to identify this condition as a novel and recognizable syndrome. Patient fibroblasts from each family were found to contain unusual vesicular inclusions supporting a likely role for the mutant protein in endosomal trafficking. Further characterization of the cellular function of the novel protein is ongoing.

3043S

The role of SRSF10 in *SMN1/2* splicing. *S. Brøner[1], Y. Hua[2], T.K Doktor[1], G.H. Bruun[1], M.R. Larsen[1], A.R Krainer[2], B.S. Andresen[1].* 1) Department of Biochemistry and Molecular Biology, University of Southern Denmark, 5230 Odense M, Denmark; 2) Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA.

SR-proteins are important regulators of pre-mRNA splicing through sequence specific binding to motifs in splicing silencers and enhancers, but their *in vivo* binding motifs are still insufficiently characterized. The intronic splicing silencer, ISS-N1, located in intron 7 of the *SMN* genes is the target of splicing correction by Antisense Oligonucleotide blocking, which is currently in clinical trials for treatment of the neurodegenerative disease Spinal Muscular Atrophy (SMA). Until now, only members of the heterogeneous nuclear ribonucleoprotein family, like hnRNPA1, are known to bind to the ISS-N1 silencer and inhibit splicing. We have performed RNA affinity binding assays combined with iTRAQ labelling and LC-MS/MS analyses in order to elucidate the binding of different splicing factors to this ISS-N1 motif. One new candidate is the splicing factor SRSF10, which belongs to the well characterized SR-protein family. SRSF10 was found to bind significantly stronger to the wild type ISS-N1 motif compared to a mutant motif in which two A>C substitutions destroys the two hnRNP A1 motifs. Our preliminary data using over expression of SRSF10 and SMN minigenes have shown that the long isoform of SRSF10 binds to the ISS-N1 motif and can affect the splicing of exon 7 in the *SMN2* gene similar to hnRNP A1. Interestingly, the presumed SRSF10 binding site overlaps the hnRNP A1 binding sites located in the ISS-N1 silencer. Whereas the hnRNP A1 is more constitutively expressed in all tissues SRSF10 shows a more neuronal tissue specific expression pattern. Thus, SRSF10 could be an important factor affecting the splicing of the *SMN* genes in a neuronal context. We are currently trying to elucidate the role of SRSF10 in determining SMN splicing efficiency and we employ a more genome-wide approach in order to characterize the role and the binding motifs of SRSF10 in other disease genes.

3044M

A functional role for BDNF in Familial Dysautonomia. *M. Nilbratt[1], M. Salani[1], E. Morini[1], F. Urbina[1], G. Lee[2], S. Haggarty[1], S. Slaugenhaupt[1].* 1) Neurology, Massachusetts General Hospital, Boston, MA; 2) Institute for Cell Engineering, John Hopkins University, Baltimore, MD.

Familial Dysautonomia (FD), or Riley-Day syndrome, is an autosomal recessive disorder with extensive sensory and autonomic nervous system involvement present at birth. FD is characterized by recurrent episodes of hypertension with tachycardia, impaired thermoregulation, and deceased sensitivity to pain and temperature. Neuropathological studies show a marked reduction of neurons in the sympathetic and sensory ganglia. All FD patients have an intronic splice site mutation in the *IKBKAP* gene, the scaffolding member of the Elongator protein complex involved in transcriptional elongation. This mutation results in tissue-specific skipping of exon 20 in the mRNA with aberrant splicing most pronounced in neuronal tissues. The alternative splicing defect leads to reduced production of IKAP protein. Complete embryonic loss of *Ikbkap* in mice causes early embryonic lethality, proving that this gene is required for effective transcription of genes involved in early neural development. The molecular mechanisms leading to FD are poorly understood. Neurotrophic factors are implicated in the survival and differentiation of several neuronal populations. Although there is evidence for a role of neurotrophins in various human neuropathies, their relevance to the disease process is not fully understood. Previous *in vitro* studies have indicated reduced neurotrophic activity in human fibroblasts from FD patients and defective expression of genes encoding neurotrophic factors in mice leads to similar neuronal phenotypes to those observed in FD patients. Here we show that reduced IKAP expression impacts brain-derived neurotrophic factor (BDNF) transcription in FD patient fibroblast cells, and we demonstrate a functional consequence of BDNF misregulation using our induced pluripotent stem (iPS) cell model system of human neural development. We found insufficient neurotrophic support from FD cells on the development of human iPS cell-derived neurons. Compromised development of human neurons was also observed by inhibition of neurotrophic activity from normal fibroblasts by neurotrophin antagonists. Interestingly, the reduced biological activity from FD patient cells was rescued pharmacologically using the plant cytokinin kinetin, which corrected aberrant mRNA splicing of *IKBKAP* and restored BDNF protein expression. Our data suggest that aberrant regulation of BDNF due to IKAP reduction may contribute to the developmental defects in neural development observed in FD patients.

3045T

Mutations in LAMA1 cause cerebellar dysplasia and cysts with and without retinal dystrophy in Poretti-Boltshauser syndrome. *K.A. Aldinger[1,2], S.J. Mosca[3], M. Tétreault[4], J.C. Dempsey[2], G.E. Ishak[9], T. Hartley[5], I. Phelps[2], R.E. Lamont[3], D. Basel[6], K.W. Gripp[7], J.S. Hogue[8], M.J. Stephan[8], F.P. Bernier[3], K.M. Boycott[5], J. Majewski[4], J.S. Parboosingh[4], A.M. Innes[3], D. Doherty[1,2].* 1) Seattle Children's Research Institute, Seattle, WA; 2) Department of Pediatrics, University of Washington, Seattle, WA; 3) Department of Medical Genetics, University of Calgary, Alberta; 4) Department of Human Genetics, McGill Univeristy, Montréal, Quebec.Montréal, Quebec; 5) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ontario. Care4Rare, Canada; 6) Division of Genetics, Medical College of Wisconsin, Milwaukee, WI; 7) Division of Medical Genetics, A.I. duPont Hospital for Children, Wilmington, DE; 8) Department of Pediatrics, Madigan Army Medical Center, Tacoma, WA; 9) Department of Radiology, Seattle Children's Hospital, University of Washington, Seattle, WA.

Cerebellar dysplasia with cysts (CDC) is an imaging finding typically seen in combination with cobblestone cortex and congenital muscular dystrophy in patients with dystroglycanopathies. More recently, CDC was reported in a series of seven children without neuromuscular involvement (Poretti-Boltshauser syndrome). Using a combination of homozygosity mapping and whole exome sequencing, we identified biallelic LAMA1 mutations as the cause of CDC in seven additional affected individuals from five families. Most of these patients also have high myopia, and some have retinal dystrophy and patchy increased T2/FLAIR signal in the cortical white matter. Two siblings with truncating LAMA1 mutations have retinal dystrophy and mild cerebellar dysplasia without cysts, further expanding the phenotype associated with LAMA1 mutations. This work adds another phenotype to the spectrum of laminopathy disorders and highlights the tissue-specific roles played by different laminin genes.

3046S

Whole genome sequencing of a balanced translocation reveals new gene candidates for epilepsy, learning difficulties and risk of acute myeloid leukaemia. *S-K. Chung[1,2], S. Ali[1], W.O. Pickrell[1,2], J.G. Mullins[1], A. Fry[3], D.T. Pilz[2,3], M. Kerr[2,4], M.I. Rees[1,2].* 1) Neurology and Molecular Neuroscience, College of Medicine, Swansea University, Swansea, UK; 2) Wales Epilepsy Research Network, College of Medicine, Swansea University, UK; 3) Institute of Medical Genetics, University Hospital Wales, Cardiff, UK; 4) Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, UK.

Whole-genome sequencing (WGS) was commissioned in a patient with generalised epilepsy, learning difficulties (LD), a familial history of treatment-resistant acute myeloid leukaemia (AML) and a constitutional balanced translocation t(1;14) (q32;q31). WGS revealed the exact translocation coordinates and confirmed breakpoints on 1q32 and 14q31 that were validated by Sanger sequencing. On chromosome 1 the breakpoint interrupts a gene of undetermined function (TAT protein family) which is poorly-documented in the literature. Pathway analysis indicates a 'DNase domain-containing protein' function with important endonuclease role in chromosome segregation and eye development in zebrafish. On chromosome 14, a more familiar gene was interrupted (*FOX* gene family) whose function includes neurodevelopmental regulation of the cell cycle and protein synthesis. Other FOX family genes have been linked to seizure disorders, cognition and haematological malignancies. Further mutation screening of these genes in unrelated epilepsy and LD cohorts (n=60) has revealed 4 non-synonymous variants that are not in controls or control dataset websites. Structure / function assays are underway to confirm pathogenic status of these variants as well as gene-fusion studies in the original index-case and screening these genes in cytogenetic-negative AML. These findings have predictive value for the at-risk members of the family and have further neuro-genetic impact in LD and genetic generalised epilepsy.

3047M

The complexity of *KDM5C* transcription: an XLID gene under the control of disease-related transcription factors. *A. Padula[1], L. Poeta[1], C. Shoubridge[2], F. Fusco[1], A. Ranieri[1], K. Helin[3], J. Gecz[2], C.E. Schwartz[4], M.V. Ursini[1], M.G. Miano[1].* 1) Institute of Genetics and Biophysics "Adriano Buzzati Traverso" CNR, Naples, Italy; 2) Department of Paediatrics, University of Adelaide, South Australia 5006, Australia; 3) Centre for Epigenetics, University of Copenhagen, Copenhagen DK-2200, Denmark; 4) J.C. Self Research Institute, Greenwood Genetic Center, Greenwood, SC, USA.

Lysine-specific demethylase 5C (*KDM5C*) is a chromatin remodelling regulator with histone demethylase activity for di- and trimethylated histone 3 lysine 4 (H3K4me2 and H3K4me3), acting as transcriptional repressor during brain development and neuronal maturation. Mutations in *KDM5C* are emerging as frequent causes of a spectrum of X-linked intellectual disorders (XLID), with cognitive handicap alone or in association with Epilepsy (XLID/Epilepsy). With Regulatory Element-1-Silencing Transcription factor (REST), a critical regulator of the spatio-temporal transition of neural progenitors to neurons, KDM5C co-occupies the promoters of a subset of REST target genes. We recently identified an epileptogenesis path, linking functionally *KDM5C* to another XLID/Epilepsy gene, encoding the homeotic transcription factor ARX, whose mutations severely impair *KDM5C* transcript regulation. Expanding our study, we analysed two additional XLID proteins that also bind the *KDM5C* promoter. They are PHD Finger Protein 8 (PHF8), a H3K9 demethylase belonging to JmjC protein family, and Zinc Finger Protein 711 (ZNF711), a transcriptional factor, whose role is unclear. We observed that PHF8 and ZNF711, which co-occupy the target promoter, cooperatively induce *KDM5C* stimulation. We therefore tested the impact of *ZNF711* mutants on *KDM5C* trans-activation in order to establish their functional effects. This activity seems to be ARX-independent and we propose that the transcriptional induction by ARX does not synergize with the action of the PHF8/ZNF711 complex and consequently they represent distinct levels in *KDM5C* regulation. Starting from these data, we are now able to break up a unique epigenetic road involved in ID and epilepsy. However, the framework of this complex regulation may become more clear once we will decode the role of other key elements, such as ncRNAs, chromatin modifiers and cofactors. Ongoing efforts to define this transcriptional path may help to identify useful tools towards research and drug discovery for XLID/Epilepsy phenotypes and many other companion disorders with malignant seizure.

3048T

Genetics of Joubert syndrome in the French Canadian population. *M. Srour[1], F.F. Hamdan[1], J. Schwartzentruber[2], L. Patry[1], C. Nassif[1], L.H. Ospina[3], E. Lemyre[1], C. Massicotte[1], D. Amrom[4], E. Andermann[4,5], R. Laframboise[6], B. Maranda[7], D. Labuda[1], J-C. Decarie[8], F. Rypens[8], C. Fallet-Bianco[8], J-F. Soucy[8], K. Boycott[9], B. Brais[5], R-M. Boucher[6], G.A. Rouleau[5], J. Majewski[2], J.L. Michaud[1] FORGE Canada Consortium.* 1) CHU Sainte-Justine Research Center, Montreal, PQ, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montréal, PQ, Canada; 3) Department of Ophthalmology, Sainte-Justine Hospital, Montréal; 4) Montreal Neurological Hospital, McGill University, Montreal, PQ, Canada; 5) Department of Neurology and Neurosurgery, McGill University, Montreal, PQ, Canada; 6) Centre Hospitalier Universitaire de Québec, Québec, PQ, Canada; 7) Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, PQ, Canada; 8) CHU Sainte-Justine, Montréal, PQ, Canada; 9) Children's Hospital of Eastern Ontario, Ottawa, ON, Canada.

Joubert syndrome (JBTS) is a primarily autosomal recessive disorder characterized by a distinctive mid-hindbrain/cerebellum malformation, oculomotor apraxia, irregular breathing, developmental delay, hypotonia and ataxia. JBTS is considered a ciliopathy; most of the known 21 causal genes are implicated in ciliary structure or function. We sought to characterize the genetic landscape associated with JBTS in the French Canadian (FC) population. We performed mutation analysis in 40 FC JBTS patients from 32 families using a stepwise approach, combining screening of recurrent FC mutations by Sanger sequencing and exome sequencing. We identified causal mutations in 28 families. Twelve families had mutations in *C5ORF42*, 9 in *CC2D2A*, 2 in *TMEM231*, 2 in *NPHP1*, one in *TCTN1*, one in *TMEM67*. Interestingly, we documented a complex founder effect with multiple recurrent mutations in 3 genes (*C5ORF42*, *CC2D2A*, *TMEM231*). In the remaining families, we identified potentially pathogenic mutations in known JBTS genes (*CEP290*, *OFD1*) and in cilia genes, which are therefore strong candidates for JBTS. We conclude that JBTS has substantial locus and allelic heterogeneity in FC populations.

3049S

Hereditary diffuse leukoencephalopathy with spheroids (HDLS): Novel CSF1R mutations and locus heterogeneity. *C. Toro[1], H. Cheung[1], M. Tzeng[1], D. Landis[1], A. Vanderver[3], R. Godfrey[1], M. Nehrebecky[1], C. Boerkoel[1], W. Gahl[1, 2].* 1) NIH Undiagnosed Diseases Program, National Institutes of Health, Bethesda, MD; 2) Office of the Clinical Director, NHGRI, National Institutes of Health, Bethesda, MD; 3) Children's National Medical Center, Washington, DC.

Hereditary diffuse leukoencephalopathy with spheroids (HDLS [MIM 164770]) is an uncommon, dominantly inherited, adult-onset neurological disorder with characteristic clinical and MRI features. Most often, patients present in their third to fifth decade of life with progressive personality changes, dementia, spasticity, gait difficulties and depression. For decades, the diagnosis was only possible by brain biopsy or autopsy. Recently, however, mutations in the protein tyrosine kinase (PTK) domain (exons 13 thru 19) of the colony stimulating factor receptor 1 (*CSF1R*) gene have been associated with this disorder. As part of the NIH Undiagnosed Diseases Program, we evaluated 8 patients in 5 families with typical clinical and imaging features of HDLS. All patients underwent clinical phenotyping, brain and spinal cord MRI and CSF examination. One patient had undergone a brain biopsy and subsequent autopsy brain examination, which confirmed the clinical diagnosis of HDLS. All patients underwent whole exome sequencing or direct Sanger sequencing of all *CSF1R* coding exons. Previously unreported, predicted damaging mutations in exons 13 and 19 of *CSF1R* were identified in 2 families. One family with documented vertical transmission and autopsy-proven HDLS did not harbor mutations in *CSF1R*, indicating locus heterogeneity for HDLS. Another patient harboring a novel predicted damaging mutation in exon 19 *(NM_005211.3:c.2512G>T:p.V838F),* with disease onset at age 35 years, shared the variant with his father, who at age 75 was reported to be clinically unaffected, indicating incomplete penetrance. Our study enhances understanding of the genetic basis of this devastating disorder and highlights the need to continue searching for additional causative genes; this could be aided by and improved insight into the cell biological mechanisms responsible for the emergence of signs and symptoms.

3050M

A variant in *DCTN2* causes intermediate Charcot-Marie-Tooth disease. A new Charcot-Marie-Tooth disease gene? *G.J. Braathen[1,2,3], H. Høyer[1,2,3], O.L. Busk[1], C.F. Skjelbred[1], M.B. Russell[2,3].* 1) Section of Medical Genetics, Department of Laboratory Medicine, Telemark Hospital, Skien, Norway; 2) Head and Neck Research Group, Research Centre, Akershus University Hospital, Lørenskog, Norway; 3) Institute of Clinical Medicine, Campus Akershus University Hospital, University of Oslo, Nordbyhagen, Oslo, Norway.

Introduction: Charcot-Marie-Tooth disease (CMT) is a heterogeneous inherited neuropathy. The number of known CMT genes is rapidly increasing mainly due to next generation sequencing technology, at present more than 60 CMT-associated genes are known. We investigated whether variants in the *DCTN2* could cause CMT. **Material and methods:** Fifty-nine Norwegian CMT families from the general population with unknown genotype were tested by targeted next-generation sequencing (NGS) for variants in *DCTN2* along with 33 CMT genes and 18 other genes causing other inherited neuropathies or neuronopathies, due to phenotypic overlap. **Results:** Targeted NGS identified in one family a variant of *DCTN2*, c.337C>T, segregating with the phenotype in five affected while it was not present in the three unaffected members. The *DCTN2* variant c.337C>T; *p.(His113Tyr)* was neither found in in-house controls nor in SNP databases. The result is further strengthened by the fact that 33 known CMT genes and 18 other peripheral neuropathy genes did not harbor pathogenic mutations. **Conclusions:** This is the first time a *DCTN2* variant has been linked to an inherited neuropathy in man.

3051T

A missense mutation in the PISA domain of HsSAS-6 causes autosomal recessive primary microcephaly in a large consanguineous Pakistani family. *V.M. Rupp[1], M.A. Khan[2], M. Orpinell[3], M.S. Hussain[4,5], J. Altmüller[4,6], M.O. Steinmetz[7], C. Enzinger[8], H. Thile[4], W. Höhne[4], G. Nürnberg[4], S.M. Baig[9], M. Ansar[10], P. Nürnberg[4,11,12], J.B. Vincent[13], M.R. Speicher[1], P. Gönczy[3], C. Windpassinger[1].* 1) Institute of Human Genetics, Medical University of Graz, Graz, Styria, Austria; 2) Gomal Centre of Biochemistry and Biotechnology, Gomal University D.I.Khan, Khyber- Pakhtoonkhuwa, Pakistan; 3) Swiss Institute for Experimental Cancer Research (ISREC), Swiss Federal Institute of Technology (EPFL), School of Life Sciences, Lausanne, Switzerland; 4) Cologne Center for Genomics (CCG), University of Cologne, Cologne, Germany; 5) Institute of Biochemistry I, Medical Faculty, University of Cologne, Cologne, Germany; 6) Institute of Human Genetics, University of Cologne, Cologne, Germany; 7) Laboratory of Biomolecular Research, Department of Biology and Chemistry, Paul Scherrer Institut, Villigen PSI, Switzerland; 8) Department of Neurology, Medical University of Graz, Graz, Austria; 9) Human Molecular Genetics Laboratory, Health Biotechnology Division, National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan; 10) Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan; 11) Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany; 12) Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany; 13) Molecular Neuropsychiatry and Development (MiND) Lab, The Campbell Family Brain Research Institute, The Centre for Addiction & Mental Health (CAMH), Toronto, Ontario, Canada.

Asymmetric cell division is essential for normal human brain development. Mutations in several genes encoding centrosomal proteins that participate in accurate cell division have been reported to cause autosomal recessive primary microcephaly (MCPH). By homozygosity mapping including three affected individuals from a consanguineous MCPH family from Pakistan, we delineated a critical region of 18.53Mb on chromosome 1p21.3-1p13.1. This region contains the gene encoding HsSAS-6, a centrosomal protein primordial for seeding the formation of new centrioles during the cell cycle. Both next-generation and Sanger sequencing revealed a homozygous c.185T>C missense mutation in the *HsSAS-6* gene, resulting in a p.Ile62Thr substitution within a highly conserved region of the PISA domain of HsSAS-6. This variant is neither present in any single nucleotide polymorphism or exome sequencing databases nor in a Pakistani control cohort. Experiments in tissue culture cells revealed that the p.Ile62Thr mutant of HsSAS-6 is substantially less efficient than the wild-type protein in sustaining centriole formation. Together, our findings demonstrate a dramatic impact of the mutation p.Ile62Thr on HsSAS-6 function and add this component to the list of genes mutated in primary microcephaly.

3052S

Consulsive seizures and SUDEP in a mouse model of SCN8A epileptic encephalopathy. *J.L. Wagnon[1], M.J. Korn[2], R. Parent[3], J.M. Jones[1], M.F. Hammer[4], G.G. Murphy[3,5], J.M. Parent[2,6], M.H. Meisler[1,2].* 1) Human Genetics, University of Michigan, Ann Arbor, MI; 2) Neurology, University of Michigan, Ann Arbor, MI; 3) Molecular and Integrative Physiology, University of Michigan, Ann Arbor, MI; 4) Arizona Research Laboratories, Division of Biotechnology, University of Arizona, Tuscon AZ; 5) Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor MI; 6) VA Ann Arbor Healthcare System, Ann Arbor, MI.

Mutations of the voltage-gated sodium channel gene SCN8A have recently been recognized as a cause of epileptic encephalopathy, which is characterized by refractory seizures with developmental delay and cognitive disability. We previously described the heterozygous SCN8A missense mutation p.Asn1768Asp in a child with epileptic encephalopathy that included seizures, ataxia, autism, and sudden unexpected death in epilepsy (SUDEP). This mutation results in increased persistent sodium current and hyperactivity of transfected neurons. We introduced the gain-of-function mutation into the mouse genome to investigate the pathology of the altered channel in vivo using TALEN technology with a targeting construct. Heterozygous Scn8aN1768D/+ mice exhibit seizures and SUDEP, confirming the causality of the de novo mutation in the proband. Using video/EEG analysis, we detected ictal discharges that coincide with tonic seizures, generalized tonic-clonic seizures, and myoclonic jerks. Prior to seizure onset, heterozygous mutants were not defective in motor learning or fear conditioning, but did exhibit mild impairment of motor coordination and social discrimination. Homozygous mutant mice exhibit earlier seizure onset than heterozygotes and more rapid progression to SUDEP. Analysis of the intermediate phenotype of functionally hemizygous Scn8aN1768D/- mice indicates that severity is increased by a double dose of mutant protein and reduced by the presence of wildtype protein. Scn8aN1768D mutant mice provide a model of epileptic encephalopathy that will be valuable for studying the in vivo effects of hyperactive Nav1.6 and the response to therapeutic interventions. (Support from NIH R01 NS34509).

3053M

FAR1 loss of function impairs the reduction of fatty acids in individuals with intellectual disability. *R. Abou Jamra[1], R. Buchert[1], H. Tawamie[1], J. Parboosingh[2], S. Uebe[1], B. Hallak[3], I. Micheil[2], C. Smitz[2], A.B. Ekici[1], H. Sticht[4], B. Schwarze[5], F. Bernier[2], A. Reis[1].* 1) Institute of Human Genetics, FAU Erlangen-Nürnberg, Erlangen, Bavaria, Germany; 2) Department of Medical Genetics, University of Calgary, Calgary, Canada; 3) Practice for Pediatrics, Kefrenbel, Syria; 4) Institute of Biochemistry, FAU Erlangen-Nürnberg, Erlangen, Germany; 5) Institute for Forensic Medicine, FAU Erlangen-Nürnberg, Erlangen, Germany.

Homozygosity mapping and exome sequencing in two Syrian siblings with severe intellectual disability, early onset epilepsy, congenital cataracts (in one patient), spasticity, microcephaly, and growth retardation revealed a homozygous mutation in the fatty acyl-CoA reductase 1 gene (FAR1: p.E165_P169delinsD). In addition, exome sequencing identified compound heterozygous mutations in FAR1 (p.Arg263X; p.Asp365Gly) in an unrelated 19 year old Canadian male with severe intellectual disability, congenital cataracts, microcephaly, spasticity and severe growth failure. In silico prediction programs and protein molecular modeling suggested a pathogenic effect for all identified variants. To further prove protein function impairment, we transfected HEK293 with either wild type or mutant FAR1 and performed gas-chromatography and mass-spectrometry on lipid extracts of cell cultures. All three variants abolished protein function since cells transfected with mutant FAR1 could not metabolize palmitic acid and stearic acid to hexadecanol and octadecanol, respectively. FAR1 is essential for the synthesis of plasmalogens, which are essential membrane components, especially in the nervous system, in the protection against reactive oxygen species. Mutations in other genes of the plasmalogen synthesis pathway in peroxisomes, PEX7, GNPAT and AGPS, cause rhizomelic chondrodysplasia punctata 1, 2 and 3, respectively. Our patients show overlapping features with rhizomelic chondrodysplasia punctata including severe intellectual disability, early onset epilepsy, cataracts, spasticity, microcephaly, and growth retardation. But they do not show chondrodysplasia. Altogether, we consider the identified mutations as causing for the phenotype in our patients and suggest FAR1 as a novel autosomal recessive peroxisomal disorder gene involved in plasmalogen synthesis.

3054T

Ten years apart: the second family with non-syndromic autosomal recessive intellectual disability due to a CRBN gene mutation. *B. Popp[1], F. Radwan[1], A. Muhammed[2], S. Uebe[1], A. Ekici[1], A. Reis[1], R. Abou Jamra[1].* 1) Institute of Human Genetics; FAU Erlangen-Nürnberg, Erlangen, Germany; 2) Practice for Pediatrics and Children with Special Needs, Lattakia, Syria.

In 2004 CRBN (cereblon) was only the second gene ever to be implicated in non-syndromic autosomal recessive intellectual disability (ARID). CRBN is part of the Cullin 4-Ring ubiquitin ligase complex CUL4 and functions as a substrate receptor which recruits proteins for ubiquitination. Extensive functional analyses of the nonsense mutation identified in single kindred with mild ARID showed altered CRBN degradation as the likely pathomechanism. Nevertheless, despite the broad application of whole-exome sequencing in ARID families no second mutation was reported to date, which would have confirmed its relevance beyond the initial family reported. Now, 10 years later we report this second CRBN mutation. We performed homozygosity mapping and exome sequencing in a family with two affected siblings of a consanguineous family with mild non-syndromic intellectual disability. Both affected individuals show a similar phenotype to the previously described cases with an estimated IQ of 50 to 70 and without dysmorphic or autistic features. We identified a homozygous splice-donor mutation in intron 8 of CRBN (c.835+1G>A) co-segregating with the phenotype in the family. In silico analysis using multiple splice site programs predicted disruption of the splice donor. Through cDNA analysis we could exclude the existence of full length CRBN transcript and characterize three aberrant transcripts in the affected individuals. One transcript leads to a frame-shift and a premature stop-codon before the last exon and thus is most likely subject to degradation through NMD. Two of these transcripts lead to an in-frame deletion and to an abruption of a C-terminal ubiquitination site. Thus this mutation likely interferes with the normal regulation of CRBN and leads to an altered formation or activity of the complex. The identification of this new mutation confirms the implication of the CRBN gene in autosomal recessive non-syndromic intellectual disability, possibly due to CRBN deficiency caused by altered autoubiquitination. The rareness of CRBN mutations underlines the heterogeneity in ARID and emphasizes that, although a second mutation is still needed to confirm pathogenicity, a gene implicated in ARID may not be discarded only because of pending replication studies. Pathway enrichment methods along with large scale collaborations may offer an opportunity to accelerate this confirmation process.

3055S

De novo mutations in _NALCN_ cause a new syndrome with congenital contractures, hypotonia, and early death. *K.M. Shively[1], J.X. Chong[1], M.J. McMillin[1], A.E. Beck[1,2], C.T. Marvin[1], K.J. Buckingham[1], J.R Armenteros[1], A. Monteil[3], J. Shendure[4], D.A. Nickerson[4], M.J. Bamshad[1,2,4], University of Washington Center for Mendelian Genomics.* 1) Department of Pediatrics, University of Washington, Seattle, WA, USA; 2) Seattle Children's Hospital, Seattle, WA, USA; 3) Department of Physiology, Institute for Functional Genomics, Montpellier, France; 4) Department of Genome Sciences, University of Washington, Seattle, WA, USA.

Mutations in multiple genes, including *TPM2*, *TNNI2*, *TNNT3*, and *MYH3*, cause distal arthrogryposis (DA) but ~60% of cases remain unexplained. Analysis of exome sequencing data from six kindreds with an atypical form of DA with neurological abnormalities including hypotonia and developmental delay identified six different missense mutations in the gene sodium leak channel, non-selective, or *NALCN*. All six mutations were confirmed by Sanger sequencing to be *de novo*. Recently, it has been reported that mutations in *NALCN* cause autosomal recessive infantile hypotonia with psychomotor retardation and characteristic facies (IHPRF) and infantile neuroaxonal dystrophy (INAD) with facial dysmorphism in consanguineous Turkish and Saudi Arabian families. Two of these mutations were predicted to result in a truncated protein (or no protein if subject to nonsense-mediated decay), while the third mutation was a missense variant located in the 3rd helix of domain IV. In contrast, all six of the mutations we discovered are in the helices of the pore-forming regions of *NALCN*. These findings demonstrate that mutations in *NALCN* cause both recessive and dominant conditions, perhaps depending on the domain(s) affected by the mutations. Moreover, while there is some phenotypic overlap between the individuals with INAD with facial dysmorphism or IHPRF and the individuals with atypical DA, there are unique clinical features that distinguish the individuals with *de novo* mutations including congenital contractures of the hands, mouth, and feet. Last, the widely varied severity of developmental delay and contractures in the six kindreds we studied suggests that *NALCN* mutations could account for individuals with other congenital contracture disorders and/or unexplained developmental delay.

3056M

Application of array painting and next generation mate-pair sequencing (MPS) for improved mapping of chromosomal breakpoints in a familial translocation segregating with a particular phenotype. *P.M. Kroisel[1], M. Auer[1], K.M. Roetzer[2], R. Birnbacher[3], K. Wagner[1], M.R. Speicher[1], M.M. Mehrjouy[4], M. Bak[4], J.B. Geigl[1], N. Tommerup[4].* 1) Institute of Human Genetics, Medical University of Graz, Graz, Austria; 2) Hanusch-Krankenhaus, Vienna, Austria; 3) Department of Pediatrics, Regional Hospital Villach, Villach, Austria; 4) Wilhelm Johannsen Centre for Functional Genome Research, Department of Cellular and Molecular Medicine, Faculty of Health Science, University of Copenhagen, Copenhagen, Denmark.

In a familial translocation t(1;6)(p36.3;p21.2) identified in a mother and one of her two sons as well as in the maternal grandmother we previously confirmed by array CGH that no genomic unbalance is present. The translocation cosegregates with distinct phenotypic features including a mild to moderate intellectual disability and microcephaly. No other genetic disorders have been reported in this family. Array painting performed by using laser micro dissected and amplified chromosomal fragments carrying the translocation breakpoints followed by hybridization to standard 60k CGH arrays allowed the characterization of both chromosomal breakpoints at an achieved resolution that made it possible to identify genes potentially affected by the translocation. A number of less than ten chromosomal fragments was sufficient to obtain an adequate signal to noise ratio for this analysis. The CAMTA1-gene (Calmodulin binding transcription activator 1) at 1p36.31 and the BTBD9 or KIAA1880 gene at 6p21.2 were found to be located very close to or at the chromosomal breakpoints. Next generation mate-pair sequencing (MPS) will further improve the resolution of the breakpoint mapping by pin-pointing the location of the breakpoints at base pair level, allowing a detailed characterization of the genotype-phenotype relationship and of the DNA repair mechanism involved in the rearrangement. This approach might also allow a more immediate and better interpretation of the molecular and functional effects of particular structural genomic aberrations.

3057T

Deletion of the 5' exons of the TCF4 gene in patients without classical Pitt-Hopkins syndrome. *S. YU[1,3], D. GAI[2], N. Jillian[1].* 1) GENETIC Medicine, SA Pathology, North Adelaide, South Australia, Australia; 2) Royal Melbourne Hospital, Parkville, Victoria, Australia; 3) School of Paeddiatrics and Reproductive Health, University of Adelaide, South Australia, Australai.

Pitt-Hopkins syndrome (PHS) is characterised by severe intellectual disability (ID) and typical facial features. Other features include episodic hyperventilation, seizures, eye abnormalities and constipation. The TCF4 gene was identified as disease-causing. Haplo-insufficiency of TCF4 has been reported to cause classical PHS. We describe two patients with deletions of the first two 5' exons of TCF4 without classical PHS. The first patient is a 16 year old boy with high functioning autism spectrum disorder and constipation. Array CGH identified a 128 kb deletion [chr18: 53,278,787-53,526,749 (hg19)]. He also has a 227kb duplication of 11p15.5 (chr15:1,448,960 -1,222,378) with unknown clinical significance. His mother does not carry the TCF4 deletion but the father is not available for testing. The second patient is a 14 year old girl who had mild ID with perceptual reasoning problems. She had a similar deletion (chr18: 51,429,785-51,677,747). Parents are not available for testing. No dysmorphism was reported in either patient. DECIPHER database recorded at least four cases with a deletion only involving the 5' exons of the TCF4 gene without classical PHS. Two cases (269234 and 269230) had mild intellectual disability. The third (250498) had ID, microcephaly, constipation, and seizures. The fourth (1024) had ID, microcephaly, delayed speech and other features. In addition, disruption of the TCF4 gene proximal to exon 4 by a translocation was reported in a girl with ID but without classic PHS (Kalscheuer et al). TCF4 has 20 exons with multiple isoforms using alternative 5' exons. Mutations and deletions causing classical PHS have been reported in most coding exons sparing the 5' exons (exons 1-3) (Whalen et al). TCF4 mRNA expression study showed that the isoforms containing the 5'exons were restricted to testis, prostate and placenta, and not expressed in brain, while other isoforms expressed in multiple brain regions (Sepp et al). It is possible that deletions of the 5' exons of the TCF4 gene only affect isoforms not expressed in brain and therefore lead to a less severe phenotype. This is the first report to suggest that deletion of the 5' exons of the TCF4 gene may not cause classical PHS. Tissue specific expression of particular isoforms may contribute to the milder phenotype. 1.Kalscheuer VM et al. (2008) Am J Med Genet Part A 146A:2053-2059. 2.Sepp M et al. (2011) PLoS ONE 6:e22138 3.Whalen S et al. (2011) Human Mutation 33:64-72.

3058S

A genetic dosage study of DYT1 Dystonia using an inducible knock-in ΔE-Tor1a mouse model. *C. Weisheit[1], W. Dauer[1,2,3].* 1) Graduate Program in Cellular and Molecular Biology, University of Michigan, Ann Arbor, MI; 2) Department of Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI; 3) Department of Neurology, University of Michigan Medical School, Ann Arbor, MI.

DYT1 dystonia is a common inherited form of primary dystonia caused by a dominant mutation in TOR1A that removes a single glutamic acid ("ΔE") residue from torsinA, the encoded protein. We generated a novel line of inducible knock-in (I-KI) ΔE-Tor1a knock-in mice to evaluate whether the ΔE mutation acts via a gain- or loss-of-function mechanism. The ΔE mutation is located in the last exon (exon 5) of Tor1a. I-KI mice contain a floxed exon 5, and an additional downstream exon 5 containing the ΔE mutation; this allele is converted from a wild type to DYT1 allele following Cre deletion, allowing spatial and temporal induction of the pathogenic protein from the endogenous locus. We intercrossed these mice with mice carrying a floxed Tor1a gene to generate a gene dosage series, hypothesizing that if the ΔE mutation causes a gain of toxic function, animals harboring two mutant alleles would fare worse than littermates with one mutant and one null allele. We evaluated Nestin Cre Tor1aI-KI/Flx and Tor1aI-KI/I-KI littermates for viability, motor phenotype, and neuropathology. We found that Nestin Cre Tor1aI-KI/I-KI mice do not experience a worsening phenotype compared to Nestin Cre Tor1aI-KI/Flx mice. Both Nestin Cre Tor1aI-KI/I-KI and Tor1aI-KI/Flx mice exhibit abnormal motor behavior and histological investigations show increased reactive gliosis, ubiquitin accumulation, and neurodegeneration in several motor brain regions recapitulating pathology observed in conditional knockout mouse models. To this end, we developed a primary cortical culture model that successfully recapitulates torsin LOF abnormalities observed in mice. With this tool, we are addressing the hypothesis that layer 5 projection neurons are distinctly susceptible to torsinA LOF. In addition, we aim to further characterize the implications of gene dosage by comparing cultures from Nestin Cre Tor1aI-KI/I-KI and Tor1aI-KI/Flx mice. This project will provide an essential cell culture model that recapitulates specific phenotypes observed in a disease-manifesting mouse. These are necessary tools currently lacking in the field that will undoubtedly lead to a better understanding of torsinA dysfunction and DYT1 pathogenesis.

3059M

Whole exome sequencing identified the first STRADA point mutation in a patient with polyhydramnios, megalencephaly, and symptomatic epilepsy syndrome (PMSE). *W. Bi[1], I. Glass[2], A. Sun[2], D.M. Muzny[3], A.L. Beaudet[1], R.A. Gibbs[1,3], C.M. Eng[1], Y. Yang[1].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics, University of Washington, Seattle, WA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

Polyhydramnios, megalencephaly and symptomatic epilepsy syndrome (PMSE) is an ultra rare neurodevelopmental disorder characterized by severe, infantile-onset intractable epilepsy, neurocognitive delay, macrocephaly and craniofacial dysmorphism. The molecular diagnoses of this condition had only been made in 16 Old Order Mennonites who carry a homozygous 7-kb founder deletion mutation affecting exons 9 to 13 of the STRADA gene, which encodes an upstream inhibitor of mammalian target of rapamycin complex 1 (mTORC1). We recently performed clinical whole exome sequencing (WES) on a 4-year-old South Asian male with global developmental delay, a history of failure to thrive, infantile spasms, seizures, repetitive behaviors, hypotonia, low muscle mass, marked joint laxity in large and small joints, and dysmorphic facial features including tall forehead, long face, arched eyebrows, small chin, and carp shaped mouth. Concurrent cSNP array and WES genotype data revealed absence of heterozygosity (AOH) in multiple chromosomal regions (total AOH ~186 Mb) consistent with the fact that parents are first cousins. A homozygous single nucleotide duplication, c.842dupA (p.D281fs), in exon 10 of the STRADA gene of this patient was identified by WES. The frameshift change is located within one of the AOH regions on 17q23.3. Sanger sequencing confirmed the mutation in the individual and identified both parents as carriers. In light of the molecular discoveries, the patient's clinical phenotype was considered to be a good fit for PMSE. In addition, his unique features expand the spectrum of clinical findings associated with STRADA mutations. Because of the non-specific nature of the PMSE phenotype and the fact that the molecular basis of the disorder had only been previously reported once in the Old Order Mennonites, PMSE was not considered as a possible diagnosis during clinical evaluations. The previously reported mutation in STRADA is limited to the 7-kb founder deletion mutation in the Old Order Mennonites. We identified for the first time a homozygous point mutation in the STRADA gene causing PMSE in a patient of South Asian origin. Additional bi-allelic mutations related to PMSE are not seen in other WGL data, supporting the rarity of this disorder. Our findings may also have treatment implications for the patient since previous studies have shown rapamycin as a potential treatment for the seizures and cognitive problems in PMSE patients (PMID 23616120).

3060T

The expanding role for chromatin remodeling in epilepsy: Gene discovery to pathogenic mechanisms. *G.L. Carvill[1], J.M McMahon[2], H. Wang[3], J. Stamatoyannopoulos[3], I.E. Scheffer[2], H.C. Mefford[1].* 1) Pediatrics, University of Washington, Seattle, WA; 2) Epilepsy Research Center and Department of Medicine, University of Melbourne, Austin Health, Australia; 3) Department of Genome Sciences, University of Washington, Seattle, WA.

Epilepsy is one of the most common neurological disorders, with a lifetime incidence of 3%. The epileptic encephalopathies (EEs) are the most severe of all the epilepsies. Patients typically present with refractory epilepsy with multiple seizure types, cognitive arrest or regression, and have a poor prognosis. *De novo* mutations have been increasingly recognized as causative for these disorders, and mutations in over 30 genes have been described. These gene discovery efforts have highlighted a role for genes involved in chromatin regulation in the pathogenesis of EE, including *CHD2, MBD5*, *MEF2C*, *FOXG1* and *ARX*. We aim to identify novel causes of epilepsy in genes that are chromatin regulators. We selected 26 known and candidate genes that act in this process and will perform targeted resequencing in a cohort of 868 patients to identify novel *de novo* mutations. To date we have completed mutation detection in 703 patients in 21 genes. Thus far we have identified 17 pathogenic mutations in seven genes with at least one novel gene for this disorder. Overall, mutations in chromatin regulator genes account for ~2% of cases in our cohort, and about 14% of all mutation positive patients. Our results illustrate the importance of this biological process in the pathogenesis of epilepsy and highlight the need for studies of the pathogenic mechanisms that underlie these conditions. To this end, we are using the chromatin remodeler, CHD2, as a model to develop novel epigenomic approaches to studying disease mechanisms. By performing ChIP-seq in four fetal tissues, we will identify those target genes that are brain-specific and integrating this data with chromatin state maps from fetal brain we will determine the likely function of CHD2. This experimental paradigm could be expanded to include other epilepsy genes. The role of chromatin remodeling is largely unexplored in epilepsy, but it is a dynamic process with many therapeutic targets, and presents a novel opportunity to develop new treatments for patients.

3061S

Processing of double-R-loops (RNA:DNA hybrids) in (CAG)•(CTG) and *C9orf72* (GGGGCC)•(GGCCCC) repeats causes instability. *M.H.M. Schmidt[1,2], K. Reddy[1,2], J. Geist[3], G.B. Panigrahi[1], T. Zu[4], L.W. Ranum[4], Y.H. Wang[5], C.E. Pearson[1,2].* 1) The Hospital for Sick Children, Canada; 2) University of Toronto; 3) Laurentian University; 4) University of Florida; 5) University of Virginia.

Expansion of gene-specific repeat sequences causes at least 40 neuromuscular and neurodegenerative diseases, 14 of which are associated with CAG/CTG repeats. Recently amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) was shown to be caused by an expanded hexanucleotide repeat. Continued somatic repeat expansions can contribute to disease onset and progression, thereby highlighting a need to understand the expansion process or processes that inhibit this. R-loops, transcriptionally-induced RNA:DNA hybrids, can occur at various trinucleotide repeat sequences including (CAG)n, (CTG)n, (CGG)n, (CCG)n, and (GAA)n, that are associated with diseases such as myotonic dystrophy, Huntington's disease, spinocerebellar ataxia, fragile X and Friedreich's ataxia. Many of these repeats are bidirectionally transcribed, allowing for single- and double-R-loop configurations, where either or both DNA strands of the repeat may be bound by RNA. R-loops can trigger repeat length instability at transcribed (CTG)/(CAG) repeat tracts, but the mechanism of how this occurs remains unclear. We demonstrate R-loop-mediated instability through direct processing of R-loops by human cell extracts (HeLa and the neuron-like SH-SY5Y). We find that double R-loops, derived from simultaneous convergent transcription across the repeats can trigger greater levels of repeat instability than single R-loops derived from transcription in either direction alone. Pretreatment with RNase H only partially suppresses R-loop-mediated instability, supporting a model in which R-loops may directly generate instability through their aberrant processing, or through formation of slipped-DNAs upon RNA removal from the R-loop and their subsequent aberrant processing. Removal of the RNA from R-loops can lead to slipped-DNA formation, as evidenced by electron microscopy. Since transcriptionally-induced R-loops can occur in the absence of DNA replication, R-loop processing, possibly involving a slipped-DNA intermediate, may be a source of repeat instability in non-proliferating tissues. Double-R-loop formation and processing to instability was extended to the expanded *C9orf72* (GGGGCC)•(GGCCCC) repeats, known to cause ALS and FTD, providing the first suggesting through which these repeats may become unstable. These findings extend our understanding of the mechanistic basis for R-loop-mediated instability at disease-associated repeat tracts.

3062M

Glutathione-S-transferase gene polymorphisms (GSTM1,GSTT1, GSTM3 &GSTP1) and and its correlation with GST enzyme activity in DM1. *A. Kumar, S. Agrawla, S. Pradhan, S. Phadke.* Sanjay Gandhi Post-Graduate Institute of Medical Sciences (SGPGIMS) Raebareli Road, Lucknow, Uttar Pradesh, India.

Background: Myotonic dystrophy type 1 (DM1) is due to trinucleotide sequence (CTG) in the 3' UTR region of DMPK gene located at 19q13.3 chromosome. Glutathione S-transferases (GSTs) activity involved in the pathogenesis of DM1 is because it contains polymorphic triplet repeat and GSTs belong to a super family of phase II detoxification enzymes and play an important role in protecting cells from damage. Aim: The intention of the present study is to assess the association of GST gene polymorphisms (GSTT1, GSTM1, GSTP1 and GSTM3) and its enzyme activity in DM1 affected Indian population Material and Methods: Clinically diagnosed 20 DM1 patients (16 men and 4 women; median age 32.8 years±9.3, range 17-52) and 75 age-sex matched controls (median age 31.0 years±8.6, range 16-54) were included in the study. DNA isolation (standard phenol chloroform method) and serum separation were performed. Serum GST level was assessed by using Glutathine-S-Transferase assay kit (Cat. No. 703302-96) of Cayman Chemical Company, USA. GSTM1 (null or present) & GSTT1 (null or present), GSTM3 (AA, AB and BB) & GSTP1 (Ile/Ile, Ile/Val and Val/Val -105) were analysed by PCR and PCR-RFLP method respectively. Results: Serum GST was reduced significantly in the patient group compared to the control group and were significantly correlated with diabetes only. Patients had significantly higher (except GSTM3A/B) GSTM1*0 (GSTM1 null genotype), GSTT1*0 and GSTP1 (ile/val) frequency than controls. The deletion frequencies (GSTM1 and GSTT1) and GSTP1 (ile/val) were not associated with higher risk while heterozygous frequency of GSTM3 (A/B) increased the risk ratio up to three fold. The GSTM1, GSTT1 and GSTM3 genotypes correlated with dyspepsia, age at presentation and duration of disease respectively. The group of combination genotype frequency had no impact on higher risk of disease. There was no correlation between GST activity and GST polymorphisms. Conclusion: Our data supports that a significant association occurs between DM1 and GSTT1 & GSTM1 polymorphism in Indian population. Further study of GST polymorphisms or other antioxidant enzymes are required to clarify the relationship between increased oxidative stress and DM1.

3063T

The whole genome sequences from a Rottweiler and Black Russian Terrier with overlapping neurological syndromes contain the same RAB3GAP1 frame shift mutation. *T. Mhlanga-Mutangadura[1], G.S. Johnson[1], G.C. Johnson[1], L. Hansen[1], G.V. Tamassia[1], J.F. Taylor[2], R.D. Schnabel[2], D.P. O'Brien[1].* 1) College of Veterinary Medicine, University of Missouri, Columbia, MO; 2) College of Agriculture, Food and Natural Resources, University of Missouri, Columbia, MO.

We generated whole-genome sequences (WGSs) with DNA from a 5-month-old male Rottweiler diagnosed with neuronal vacuolation and laryngeal paralysis and a 3-month-old female Black Russian Terrier diagnosed with juvenile laryngeal paralysis, polyneuropathy, microphthalmia, persistent pupillary membranes and cataracts. Initially we thought these two dogs had overlapping but distinct disease syndromes; however, both of their WGSs contained the same homozygous single base pair deletion: RAB3GAP1:c.743delC. The frame shift produced by this deletion predicts a premature stop codon and a truncated gene product RAB3GAP1:p.P248Lfs3* missing 730 C-terminal amino acids. In addition to the Rottweiler and Black Russian Terrier with the WGSs, the DNA samples from all 3 other affected Rottweilers and all 4 other Black Russian Terriers in our collection were homozygous for the c.743delC allele; whereas, 85 clinically normal Rottweilers and 41 clinically normal Black Russian Terriers tested either heterozygous (n = 18) or homozygous for the ancestral allele (n = 108). Truncating mutations in human RAB3GAP1 have been found in patients with Warburg micro syndrome. This severely debilitating developmental disorder is characterized by microcephaly, microphthalmia, microcornea, congenital cataracts, optic atrophy, cortical dysplasia, mental retardation, and hypogonadism. In contrast, Rab3gap1 knockout mice have exhibited only subtle abnormalities. The disease phenotype produced by the frame shift mutation in canine RAB3GAP1 appears to be less severe than that in the children with Warburg micro syndrome but much more severe than that of Rab3gap1 knockout mice. We are currently re-evaluating the disease phenotypes from affected dogs from the two breeds to see if they are, in fact, identical and to determine if they have other clinical signs in common with the Warburg micro syndrome patients.

3064S

The missing factors influencing spinal and bulbar muscular atrophy: evaluation of genetic polymorphisms. *C. Bertolin[1], G. Querin[1], M. Pennuto[2], E. Pegoraro[1], C. Gellera[3], D. Pareyson[3], C. Mariotti[3], G. Sorarù[1].* 1) Biology, University of Padova, Padova, Padova, Italy; 2) 2Dulbecco Telethon Institute Laboratory of Neurodegenerative Diseases, Centre for Integrative Biology (CIBIO), University of Trento, Italy; 3) 3Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milan, Italy.

Spinal and bulbar muscular atrophy (SBMA) is caused by a pathological expansion over 38 of a CAG repeat in the first exon of the androgen receptor (AR) gene on chromosome X, coding for a polyQ tract (La Spada et al, 1991). SBMA is an androgen-dependent disorder, with males with full disease manifestations, and females showing only mild symptoms even if homozygous for the mutation. While a correlation between expansion size of polyQ tract and disease severity has been reported, patients with the same number of CAG repeats have different age at onset and disease progression even if relatives. This study aims to find AR gene polymorphisms that can explain the genotype-phenotype correlation of spinal bulbar muscular atrophy (SBMA) in an Italian cohort of patients. In order to do that, we evaluated polymorphism (SNPs and trinucleotide repeats) of the human AR gene. The effect of these sequences on SBMA phenotype has not been studied yet. To characterize the effect of AR coding repeated sequences on SBMA phenotype, we genotyped AR polymorphisms in 132 molecularly defined Italian SBMA patients, referring to the Motor Neuron Clinic of the University of Padua and to Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milan. Age of onset was considered as the marker of disease severity. Among candidate AR polymorphisms, only a GGN sequence coding polyG is polymorphic in our cohort. Interestingly, the genotyping of polyG stretch highlights that patients belonging to different polyG allele classes showed different levels of correlation between polyQ lenght and disease's severity. Our study confirms that the size of SBMA-causing polyQ tract does not fully explain the disease phenotype and point to a polyG strech within AR exon 1 as a modifier in SBMA phenotype.

3065M

Abnormalities in neuronal architecture and synaptic activity impairment in mice heterozygous for different deletions of the Williams-Beuren syndrome locus. *C. Borralleras[1, 3], M. Bosch-Morató[2], B. Guivernau[2], M. Bravo-Garmendia[2], LA. Pérez-Jurado[1, 2, 3], FJ. Muñoz[2], V. Campuzano[1, 2, 3].* 1) Hospital del Mar Research Institute (IMIM), Barcelona, Barcelona, Spain; 2) Universitat Pompeu Fabra, Barcelona, Spain; 3) Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona, Spain.

Williams-Beuren syndrome (WBS, [MIM 194050]) is a neurodevelopmental disorder caused by a spontaneous 1.55-1.83 Mb deletion at chromosal band 7q11.23 containing between 26-28 genes, including 2-3 TFII-I family members. WBS is characterized by distinctive dysmorphic features, mild growth retardation, cardiovascular abnormalities, and a specific cognitive profile with mild intellectual disability, visuospatial deficits, hypersensitivity to sounds and hypersociability. Neurologic examination and brain imaging in patients have shown changes in brain function and brain anatomy. Haploinsufficiency at *GTF2I* has been shown to play a major role in the neurodevelopmental abnormalities of WBS. To clarify the involvement of *GTF2I* in neurocognitive features we characterized four WBS mouse models deleting *Gtf2i* with and without additional genes: PD (proximal deletion, from *Gtf2i* to *Limk1*), CD (complete deletion, from *Gtf2i* to *Fkbp6*), *ΔGtf2i* +/- and *ΔGtf2i* -/-. We examined the morphology of a transgenically labeled subset of cortical and hippocampal pyramidal neurons in YFP-H mice crossed with the different mutant lines. The number of YFP+ cells was decreased by up to 87% in WBS mouse models in various cortical regions. Lower number of YFP+ cells was present as early as P12-14 developmental stages. Spine density of apical dendrites in the CA1 area of the hippocampus was reduced in *ΔGtf2i* -/-, PD and CD mice by 14%, 15% and 17%, respectively. Moreover, spines in all animals were shorter, pointing to a more immature type of spine. Calcium imaging experiments in primary cultures of hippocampal neurons showed alterations in glutamatergic synaptic activity. As GTF2I has a role in PI3K/AKT pathway (involved in synaptic plasticity) and is a regulator of *Dlx* genes (involved in the differentiation of GABAergic projection neurons), we analyzed the expression of these genes in our mutant mice. Mutant mice showed deregulation of these genes, suggesting impairment in synaptic plasticity and inhibitory synaptic activity. These findings suggest that *GTF2I* plays an important role in neurodevelopmental abnormalities in WBS. Abnormalities in neuronal architecture and perturbations in synaptic activity could be associated with the cognitive impairment seen in WBS patients.

3066T

A Homozygous PIGN Missense Mutation in Soft Coated Wheaten Terriers with Paroxysmal Dyskinesia. *A. Kolicheski[1], G.S. Johnson[1], D. Gilliam[1], T. Mhlanga-Mutangadura[1], J.F. Taylor[2], R.D Schnabel[2], D.P. O'Brien[1].* 1) College of Veterinary Medicine, University of Missouri-Columbia, Columbia, MO; 2) College of Agriculture, Food and Natural Resources, University of Missouri, Columbia, MO.

The paroxysmal dyskinesia (PD) we studied is a previously undescribed Soft Coated Wheaten Terrier (SCWT) movement disorder characterized by multiple episodes of dyskinesia that sometimes last for only a few minutes and other times last for over an hour. Episodes begin in young adult SCWTs and include hyperextension or hyperflexion primarily of the pelvic limbs with difficulty moving forward and arrhythmic hopping. Between episodes the dogs move normally. We generated a whole genome sequence (WGS) with 21-fold average coverage and 6.9 million sequence variants (differences from the reference canine genome sequence) with DNA from a PD-affected SCWT and another WGS with 18-fold average coverage and 5.5 million sequence variants with DNA from a second PD-affected SCWT that had no known relationship to the first. We filtered the sequence variants to identify those that were homozygous in both case WGSs, predicted to alter the amino acid sequence of the gene product, and absent from 100 other canid WGSs. The only variant that met these criteria was a *PIGN*:c.398C>T transition that predicts a *PIGN*:p.T133I substitution. All 13 PD-affected SCWT (including the 2 with WGSs) were homozygous for the *PIGN*:c.398T allele. The genotyping of archived SCWT DNA samples of unknown phenotype identified 469 *PIGN*:c.398C homozygotes, 7 heterozygotes and no *PIGN*:c.398T homozygotes. Samples of 505 dogs of varied breeds were also genotyped and all were homozygous wild type. *PIGN* encodes an enzyme in the biosynthetic pathway for glycosylphosphatidylinositol anchors that attach many different proteins to cell surfaces. Mutations in human *PIGN*, cause multiple congenital anomalies-hypotonia-seizures syndrome 1 (MCASH1 [MIM 614080]), with characteristic neonatal hypotonia, facial dysmorphism, and psychomotor underdevelopment. A minority of MCASH1 patients have exhibited choreoathetosis which may be related to the PD in the SCWTs. Nonetheless, the canine *PIGN* mutation reported here causes disease phenotype that is very different than that of the human MCASH1 patients.

3067S

Congenital hypotonia: Two rare diseases in one family. *T. Falik-Zaccai[1,2], M. Baydany[1,2], L. Kalfon[1], N. Nasser Samra[1,2], Y. Shoval[1], D. Savitzki[3], S. Ivry[4], Z. Zonis[5], H. Mandel[6].* 1) Institute of Human Genetics, The Galilee Medical Center, Nahariya, Israel; 2) The Galilee Faculty of Medicine, Bar Ilan University, Safed, Israel; 3) Department of Pediatric Neurology, The Galilee Medical Center, Nahariya, Israel; 4) Department of Anesthesiology, The Galilee Medical Center, Nahariya, Israel; 5) Pediatric Intensive Care Unit, The Galilee Medical Center, Nahariya, Israel; 6) Meyer Children's Hospital, Metabolic Unit, Rambam Medical Center, Haifa, and Rappaport Faculty of Medicine, Technion, Haifa, Israel.

Background: Congenital hypotonia is a non specific physical finding, common to a heterogeneous group of diseases. We present an extended highly consanguineous family of Druze origin, with multiple affected children born with severe congenital hypotonia. We hypothesize that congenital hypotonia in this family is a monogenic autosomal recessive disorder. We aim to determine its causative gene and mutation, and to characterize its molecular and biochemical basis. Methods: Clinical investigations, homozygosity mapping of patients and their parents, linkage analyses and whole exome sequencing, were performed to find the novel sequence variation segregating with the disease in two nuclear families within the extended pedigree. mRNA and protein levels were further determined. Results: We have identified 4 novel sequence variations within two genes in the two nuclear families that present both with multiple affected children with congenital hypotonia. In the first family we have identified a novel nonsense mutation in the gene NGLY1. A recently reported gene with only 8 patients described worldwide. RNA levels were significantly reduced among the patients compared to controls in both fibroblasts and leukocytes. In the second nuclear family which is closely related to the first, we have identified compound heterozygosity for three novel RYR1 mutations in two affected sisters. Two mutations inherited from the father in cis and a third mutation inherited from the mother. We screened 100 healthy ethnically matched individuals from the same village and found 1:30 carriers for the mutation in NGLY1, 1:100 carriers' frequency for the allele presenting two mutations in cis, and 1:33 carriers' frequency for the third mutation in the gene RYR1. No healthy individual in this village was found to be homozygous or compound heterozygous for any of the mutations. Conclusions: We have identified a genetic isolate for two extremely rare forms of congenital hypotonia presenting to our surprise in one extended highly consanguineous family. Revealing the causative gene and mutation will enable us to provide credible genetic counseling and pre-natal diagnosis to the extended family, and genetic screening for this high risk population. The family presented here demonstrates the critical role of comprehensive clinical evaluation along with whole exome sequencing for accurate determination of the genetic basis of congenital hypotonia and other rare genetic disorders.

3068M

Identification of a novel autophagy-related gene mutation in a canine storage disease. *K. Kyöstilä[1,2,3,4], V. Jagannathan[5], P. Syrjä[1], T. Jokinen[6], G. Chandrasekar[7], J. Kere[2,4,7,8], E.H. Seppälä[1,2,3,4], D. Becker[5], M. Drögemüller[5], E. Dietschi[5], C. Drögemüller[5], J. Lang[9], F. Steffen[10], C. Rohdin[11], P. Wohlsein[12], D. Henke[13], A. Oevermann[13], H. Lohi[1,2,3,4], T. Leeb[5].* 1) Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland; 2) Research Programs Unit, Molecular Medicine, University of Helsinki, Helsinki, Finland; 3) Department of Medical Genetics, University of Helsinki, Helsinki, Finland; 4) Department of Molecular Genetics, Folkhälsan Institute of Genetics, Helsinki, Finland; 5) Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland; 6) Department of Equine and Small Animal Medicine, University of Helsinki, Helsinki, Finland; 7) Center for Biosciences, Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden; 8) Science for Life Laboratory, Karolinska Institutet, Solna, Sweden; 9) Department of Clinical Veterinary Medicine, Division of Clinical Radiology, Vetsuisse Faculty, University of Bern, Bern, Switzerland; 10) Neurology Service, Department of Small Animals, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland; 11) University Animal Hospital, Swedish University of Agricultural Sciences, Uppsala, Sweden; 12) Department of Pathology, University of Veterinary Medicine Hannover, Hannover, Germany; 13) Division of Neurological Sciences, Vetsuisse Faculty, University of Bern, Bern, Switzerland.

Purebred dogs suffer from different types of inherited neurodegenerative diseases similar to humans. As a large animal, the dog is physiologically closer to human than traditional model organisms and therefore a better replicate for human phenotypes. We have studied clinical, pathological and genetic aspects of a storage disease in the Lagotto Romagnolo dog that predominantly affects the nervous system. The affected dogs suffer primarily from progressive ataxia but show also other symptoms, such as episodic nystagmus and behavioral changes. The onset of clinical signs varies from 4 months to 4 years of age. The main histological findings in the central and peripheral nervous system are profound cytoplasmic vacuolization in neurons along with granular spheroid formation. Cytoplasmic vacuolization of several secretory epithelia is also present. To identify the genetic cause of the disease, we genotyped three pathologically confirmed cases and four control dogs using Illumina's canine 170K SNP arrays. In addition, we performed low-coverage whole genome sequencing (WGS) in one affected dog. Genotyping data was analyzed by performing homozygosity mapping and parametric linkage analysis. WGS data was filtered against the genomes of 46 dogs from different breeds. Combining of results from genome-wide analyses and the WGS data revealed a single non-synonymous variant in an autophagy-related gene on canine chromosome 20. Genotyping of this variant in a cohort of nearly 2,000 Lagotto Romagnolo dogs, including 17 clinically confirmed affected dogs, indicated a highly significant association with the disease ($p=3.8 \times 10^{-108}$). Furthermore, the variant was absent in 642 dogs from 40 other breeds. The gene has not been previously associated with inherited diseases in any species. The encoded protein is poorly characterized but proposed to function in a macroautophagy pathway, which has been implicated in neurodegenerative diseases. We are currently performing experiments to further delineate the function of the gene in the zebrafish model. In conclusion, our study identifies a novel mutation associated with a previously uncharacterized canine storage disorder. This animal model may be useful to further characterize the interplay between dysfunctional autophagy and neural degeneration, and to explore possible therapeutic options. Meanwhile, dog breeders will benefit from genetic testing.

3069T

Mutations in the neurofilament heavy chain gene (NEFH) trigger pathological aggregates in Charcot-Marie-Tooth disease. *A. Rebelo[1], A. Abrams[2], A. Horgha[3], M. Gonzalez[1], H. Houlden[3], E. Cottenie[3], M. Sweeney[3], M. Hanna[3], M. Auer-Grumbach[4], M. Reilly[3], S. Zuchner[1].* 1) HIHG, University of Miami, Miami, FL; 2) Neurology, University of Miami, Miami, FL; 3) University College London, London, England; 4) Medical University of Graz, Graz, Austria.

Charcot-Marie-Tooth disease (CMT) comprises a group of heterogeneous motor and sensory neuropathies caused by mutations in several genes. Abnormal accumulation of neurofilaments and disruption of neurofilament network is a common cause of motor neuron degeneration in CMT. In particular, mutations in the neurofilament light chain (NEFL) and heat shock protein, HSPB1, leads to neurofilaments pathological aggregates. Neurofilament aggregation is also the hallmark of neuronal dysfunctional associated with other neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS), Parkinson and Alzheimer. Applying exome sequencing, we have identified 2 novel dominantly inherited CMT mutations in the neurofilament heavy (NEFH) subunit. Both mutations caused frameshifts in the conserved c-terminal domain, resulting in the loss of the stop codon and consequently producing an extra 40 amino acid long extension at the c-terminal tail. NEFH mutations identified in the patients were also evaluated at the functional level, where we observed prominent abnormal aggregates in neuro2a cells transfected with the mutant protein. Therefore, we concluded that abnormal aggregation of the mutant NEFH suggests a possible toxic gain of function effect that leads to the progression of CMT in patients. Moreover, zebrafish eggs injected with mRNA encoding the mutant NEFH resulted in embryos with significantly decreased motor neuron lengths. Our studies support the hypothesis of neurofilament aggregation and dysfunction as a disease mechanism for axonopathies.

3070S

MUTATIONS IN THE TRICARBOXYLIC ACID CYCLE ENZYME, ACONITASE 2, CAUSE EITHER ISOLATED OR SYNDROMIC OPTIC NEUROPATHY WITH ENCEPHALOPATHY AND CEREBELLAR ATROPHY. *JM. Rozet[1], M. Metodiev[1], S. Gerber[1], L. Hubert[1], D. Chretien[1], X. Gerard[1], P. Amati-Bonneau[2], MC. Jacomoto[3], N. Boddaert[4], A. Kaminska[4], I. Desguerre[4], M. Rio[4], J. Kaplan[4], A. Munnich[1,4], A. Rotig[1], C. Besmond[1].* 1) Inserm UMR1163, UPD - Sorbonne Paris Cité, Imagine-Institute of genetic diseases, Paris, France; 2) Department of Biochemistry and Genetics, CHU d'Angers, Angers, France; 3) Department of Ophthalmology, Polyclinique du Maine, Laval, France; 4) Departments of Pediatrics, Radiology and Genetics, Hôpital Necker-Enfants Malades, Paris, France.

Inherited optic neuropathy has been ascribed to mutations in mitochondrial fusion/fission dynamic genes, nuclear and mitochondrial DNA-encoded respiratory enzyme genes or nuclear genes of poorly known mitochondrial function. On the other hand, enzymopathies of the tricarboxylic acid cycle (TCA) have been reported to cause severe encephalopathies or isolated retinitis pigmentosa, but no TCA-cycle enzyme deficiency has been hitherto reported in isolated optic neuropathy. Studying a series of five patients with optic atrophy, we found homozygous or compound heterozygous missense and frameshift mutations in the gene encoding mitochondrial aconitase (ACO2), a TCA-cycle enzyme, catalyzing interconversion of citrate into isocitrate. Retrospective studies using patient-derived cultured skin fibroblasts revealed various degrees of deficiency in ACO2 activity but also in ACO1 cytosolic activity. Our study shows that autosomal recessive ACO2 mutations can cause either isolated or syndromic optic neuropathy. This observation identifies ACO2 as the second gene responsible for non syndromic autosomal recessive optic neuropathies and provides evidence for a genetic overlap between isolated and syndromic forms, giving further support to the view that optic atrophy is a hallmark of defective mitochondrial energy supply.

3071M

UBE3B Deficiency in Kaufman oculocerebrofacial syndrome. *R. Yilmaz[1], L. Basel-Vanagaite[2,3,4], S. Tang[5], M.S. Reuter[6], N. Rahner[7], D.K. Grange[8], M. Mortenson[9], P. Koty[9], H. Feenstra[10], K.D. Farwell Gonzalez[5], H. Sticht[11], N. Boddaert[12], J. Desir[13], K. Anyane-Yeboa[14], C. Zweier[6], A. Reis[6], C. Kubisch[1], T. Jewett[9], W. Zeng[5], G. Borck[1].* 1) Institute of Human Genetics, University of Ulm, Ulm, Germany; 2) Raphael Recanati Genetic Institute and Felsenstein Medical Research Center, Rabin Medical Center, Beilinson Campus, 49100, Petah Tikva, Israel; 3) Sackler Faculty of Medicine, Tel Aviv University, 69978, Tel Aviv, Israel; 4) Pediatric Genetics, Schneider Children's Medical Center of Israel, 49202, Petah Tikva, Israel; 5) Ambry Genetics, Aliso Viejo, CA, 92656, USA; 6) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054, Erlangen, Germany; 7) Medical Faculty, Institute of Human Genetics, University of Dusseldorf, 40225, Dusseldorf, Germany; 8) Division of Genetics and Genomic Medicine, Department of Pediatrics, Washington University School of Medicine, St. Louis Children's Hospital, St. Louis, MO, 63110, USA; 9) Department of Pediatrics, Section on Medical Genetics, Wake Forest School of Medicine, Winston-Salem, NC, 27157, USA; 10) Roosevelt Hospital, New York, NY, 10019, USA; 11) Institute of Biochemistry, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054, Erlangen, Germany; 12) Department of Pediatric Radiology, Hôpital Necker, Enfants Malades and Medical Faculty, Université Paris Descartes, 75015, Paris, France; 13) Institut de Pathologie et de Génétique, 6041, Gosselies, Belgium; 14) Columbia University Medical Center, New York, NY, 10010, USA.

Blepharophimosis-mental retardation (BMR) syndromes are clinically and genetically heterogeneous group of disorders characterized by intellectual disability and typical facial features. Mutations in *KAT6B* and *MED12* were reported to cause autosomal dominant Say-Barber-Biesecker-Young-Simpson syndrome and X-linked Maat-Kievit-Brunner type BMR syndrome, respectively. We have previously identified biallelic *UBE3B* mutations by exome sequencing in patients affected by Kaufman oculocerebrofacial syndrome (KOS) (Basel-Vanagaite et al., 2012), a syndrome with clinical overlap with BMR syndromes. KOS is characterized by developmental delay, growth retardation with a small head circumference, congenital anomalies, low cholesterol levels and characteristic facial dysmorphisms. Here, we have screened 24 patients with clinical overlap to KOS by Sanger or exome sequencing. We have identified *UBE3B* mutations in six additional patients from five unrelated families. Two of the patients who were siblings had previously been reported as having Toriello-Carey syndrome and one patient had been reported as having a "new" syndrome by Buntinx and Majewski in 1990. In patients with no *UBE3B* mutation, we sequenced *KAT6B* and *MED12*. None of the patients carried mutations in either of the gene. As the clinical heterogeneity could also suggest the involvement of another gene, we performed exome sequencing in a sibling pair and our filters identified *PROM2* as candidate gene. We did not however find any rare or unreported variants in *PROM2* in the patients whom were screened by capillary sequencing. Although *UBE3B* is reported to be highly expressed in adult tissues, we found it has low- to medium- abundance expression based on PCR on various tissue cDNAs. Moreover, different cell lines such as U2OS, 293T, MCF-7 and H4 contain very low or undetectable UBE3B protein levels as shown by Western blot. We did not detect any difference in protein mobility on SDS-PAGE between phosphorylated and dephosphorylated forms of UBE3B. *oxi-1* is a *UBE3B* homolog in C. elegans and was shown to be involved in protein degradation under oxidative stress. However, in preliminary tests, we did not observe increased UBE3B protein levels upon H_2O_2 induced oxidative stress. Overexpression studies revealed that UBE3B is predominantly localized in the cytoplasm. We will also present the results of localization studies of *UBE3B* constructs containing KOS-associated missense variants.

3072T

Whole genome sequences from two individual dogs with neuronal ceroid lipofuscinosis contain novel truncating mutations: one in CLN8 and the other MFSD8. *J. Guo[1], M.L. Katz[2], H.A. Brown[3], T. Mhlanga-Mutangadura[1], J.F. Taylor[4], R.D. Schnabel[4], D.P. O'Brien[5], G.S. Johnson[1].* 1) Veterinary Pathobiology, University of Missouri-Columbia, columbia, MO; 2) Mason Eye Institute, University of Missouri School of Medicine, Columbia, MO; 3) Metz Petz Veterinary Clinic at Shawnee, Lima, OH, USA; 4) Division of Animal Science, University of Missouri College of Agriculture, Food and Natural Resources, Columbia, MO; 5) Department of Veterinary Medicine and Surgery, College of Veterinary Medicine, University of Missouri Columbia, MO.

We generated separate whole genome sequences (WGSs) with DNA from 2 different young adult female dogs that were diagnosed with neuronal ceroid lipofuscinosis (NCL) based on a clinical history of neurodegeneration and the presence of CNS neurons with autofluorescent cytoplasmic storage granules containing material with characteristic multilamellar ultrastructures. One of the dogs had both Australian Shepherd and Blue Heeler ancestry. She was euthanized at approximately 22 months of age due to intractable seizures. For about a year prior to euthanasia, she had shown signs of progressive neurodegeneration that included visual loss, cognitive decline and MRIs indicative of diffuse brain atrophy. Among the homozygous sequence variants in the WGS from this dog was a CLN8:c.585G>A transition that predicts a CLN8:p.W95* nonsense mutation. This variant was considered most likely to be causal because it was the only rare variant in a canine ortholog of a gene associated with human NCL and because it was predicted to encode a truncated protein missing the 93 C-terminal amino acids. By genotyping archived DNA samples from 1488 Australian Shepherds and 133 Blue Heelers at CLN8:c.585, we identified a pair of purebred Australian Shepherd siblings that were homozygous for CLN8:c.585A. These siblings had been previously diagnosed with NCL of unknown cause. The other dog was a purebred Chinese Crested. This dog was euthanized at approximately 19 months of age due to a 5-month history of progressive neurological decline that included blindness, anxiety, and cognitive impairment. Among the homozygous sequence variants in the WGS from this dog was MFSD8:c.843delT. This single base deletion is predicted to cause a frame shift and premature stop codon resulting in a truncated protein, MFSD8:p.F282Lfs13*, missing its 239 C-terminal amino acids. All but one of our archived DNA samples from 1477 Chinese Cresteds tested homozygous for the ancestral MFSD8:c.843T allele. The exception was a sample from a 10-year-old Chinese Crested that lived in Sweden and tested heterozygous for the c.843delT allele. MFSD8 mutations have been identified in human patients with a form of variant late infantile NCL known as CLN7. These examples indicate that the likely molecular genetic causes of canine Mendelian diseases can be identified by generating WGSs from individual affected dogs.

3073S

Mice with combined deficiencies of β-Hexosaminidase A and Sialidase Neu3 mimic the fundamental aspects of the neurological abnormalities of Tay-Sachs disease due to accumulation of ganglioside: New hope for Tay-Sachs patients and families. *V. Seyrantepe[1], S. Akyildiz Demir[1], Y. Calhan[1], Z. Timur[1], C. Marsching[2], R. Sandhoff[2], T. Miyagi[3].* 1) Dept Molecular Biology and Genetics, Izmir Institute of Technology, Urla Izmir, Turkey; 2) DKFZ Heidelberg Zelluläre und Molekulare Pathologie AG Lipidpathobiochemie Heidelberg Germany; 3) 6Division of Cancer Glycosylation Research, Institute of Molecular Biomembrane and Glycobiology, Tohoku Pharmaceutical University, Sendai, Japan.

Tay-Sachs disease is a severe lysosomal storage disorder caused by mutations in the HEXA gene coding for α subunit of lysosomal β-Hexosaminidase A enzyme, which converts GM2 to GM3 ganglioside. HexA-/- mice, depleted of β-Hexosaminidase A enzyme, remains asymptomatic to 1 year of age, owing to the ability of these mice to catabolise stored GM2 ganglioside via sialidase (s) removing sialic acid into glycolipid GA2 which further processed by β-Hexosaminidase B, thereby bypassing the HexA defect. Previously we showed that mice with targeted distruption of both HexA and Neu4 genes (HexA-/-Neu4-/-) have epileptic seizures and accumulating GM2 ganglioside in brain. To elucidate whether plasma membrane associated sialidase Neu3 can also contribute to GM2 ganglioside degradation, we generated a double deficient mouse model by crossing previously generated HexA-/- mouse model and Neu3-/- mouse model. HexA-/-Neu3-/- were health at birth but lost weight gradually and died at 4 months of age. Thin layer chromatography analysis of double deficient mice showed increased GM2 ganglioside level and altered ganglioside pattern in brain as well as liver, kidney and lung. Mass spectrometry analysis confirmed accumulation of GM2 and other gangliosides such as LacCer, GA2 and GM3. Immunohistochemical analysis using anti-GM2 antibodies indicated massive accumulation of GM2 in hippocamous and cortex of mice brain Slow movement, ataxia and tremor were among neurological abnormalities. The unexpected severe phenotype of HexA-/- mice appeared to be influence by the status of Neu3 gene. HexA-/-Neu3-/- mice mimic the fundamental aspects of the neurological abnormalities of Tay-Sachs disease. Our data suggests potential therapy of Tay-Sachs disease based on the upregulation of human sialidase Neu3 by drug treatment and give new hope to patients and families.

3074M

Mutation of NUP50 in a consanguineous family with intellectual disability. *JM. Capo-chichi, C. Nassif, ME. Samuels, FF. Hamdan, JL. Michaud.* Centre de Recherche du CHU Sainte-Justine, 3175 Chemin Côte Ste-Catherine Montréal, Québec, Canada.

We studied the case of two sisters from a consanguineous North-African family that presented with intellectual disability (ID). We first performed whole-genome SNP genotyping and homozygosity mapping on both affected siblings and identified 6 regions of shared homozygosity (HR) of ≥ 25 consecutives SNPs and extending over 1 Mb. The largest HRs identified were of ~ 10 Mb (chr22:37,650,774-47,938,048) and ~7 Mb (chr7:116,156,423-123,007,208). Next, we performed exome sequencing on one of the affected siblings and identified 55 rare homozygous amino acid-altering or splicing variants in the exome. Only one variant, a non-synonymous change in NUP50 (NM_007172:exon5:c.502C>T (p.R168W)) was found in an HR (the 10Mb one) shared by the two probands. Sanger sequencing showed that this variant segregates with a recessive mode of inheritance in the family. The 502C>T (p.R168W) variant was absent from all public SNP databases (1000 Genomes, Exome Variant Server (EVS), dbSNP138), and from our in-house exome dataset (>1000). The R168 residue in NUP50 is highly conserved among various species and the p.R168W mutation is predicted to be damaging by Polyphen-2, SIFT and Provean. NUP50 (or NPAP60) encodes for a component of the nuclear pore complex that plays a role in nucleocytoplasmic transport of proteins through direct interaction with importin-alpha. Nup50 is expressed in various mouse tissues, particularly in the brain with high signals in cerebellum and hippocampus. Nup50 has been shown to play a role in development as mice homozygous for a targeted null mutation in Nup50 die embryonically and present neural tube abnormalities and intrauterine growth retardation. Humans express two functional isoforms of NUP50 (NUP50-S and NUP50-L), both of which are affected by the p.R168W mutation. NUP50-S has been shown to stabilize the binding of importin-alpha to classical NLS-cargo, whereas NUP50-L rather promotes the release of importin-alpha from the NLS-cargo. The antagonizing roles of NUP50 isoforms suggest that changes in NUP50 levels might perturb the efficiency of nucleocytoplasmic transport in humans. Our findings suggest that the homozygous p.R168W mutation in NUP50 might be responsible for the ID phenotype in the studied family. Identification of additional similarly affected families with mutations NUP50 will be necessary to establish the role of NUP50 as a neurodevelopmental disease gene. Collaborations towards this effort are welcome.

3075T

Phenotype of 21 novel autosomal recessive cognitive disorders. *K. Kahrizi[1], A. Tzschach[2,3], H. Hu[2], Z. Fattahi[1], M. Hosseini[1], S. Abedini[1], T. Wienker[2], H.H. Ropers[2], H. Najmabadi[1].* 1) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; 2) Max Planck Institute for Molecular Genetics, Berlin, Germany; 3) Institut für Human Genetik Universitäts Klinikum Tübingen Calwerstr, Tübingen, Germany.

After releasing 52 novel intellectual disability genes by Najmabadi et al in 2011, most of researchers working on cognitive pathway in human and animal models were interested in phenotypically aspect of some of these genes to reach the function of changes in mouse, drosophila and human brain. From 52 novel candidate genes of Intellectual Disability (ID) that have been reported, 21 families had additional features that some of them could have been related to the effect of these genes on neuronal networking. Ten out of 21 syndromic families with microcehaly or neurological symptoms were selected and MRI was performed for one or two affected siblings in each family. In the previous study, homozygosity mapping, coupled by conventional sequencing or next generation sequencing (NGS) were used to define underlying genetic defects. From total of 10 families, eight families showed different types of abnormalities as periventricular leukoencephalopathy, reduced cerebral cortex, pachygyria, polymicrogyria, small corpus callosum and cerebellar hypoplasia on MRI including subjects with mutations in CAPN10, TAF2, CNKSR1, WDR45L, ERLIN2, PARP1 and SLC31A1 genes. Three syndromic families had normal brain architecture (ZBTB40, KIF7 and TMEM135).

3076S

An emerging role for the Rho-GEF Collybistin in neuropathological mTORC1-mediated protein translation. *A. Sertie[1], C.O.F. Machado[1], K. Griesi-Oliveira[2], C. Rosenberg[2], F. Kok[2,3], S. Martins[1], M.R. Passos Bueno[2].* 1) Hospital Israelita Albert Einstein, Centro de Pesquisa Experimental. Av. Albert Einstein, 627/701 - Morumbi - São Paulo, Brasil; 2) Centro de Estudos do Genoma Humano, Instituto de Biociências, Universidade de São Paulo, Brasil; 3) Faculdade de Medicina, Universidade de São Paulo, Brasil.

Protein synthesis regulation via the mammalian target of rapamycin (mTOR) signaling pathway plays key roles in neural development and function, and its dysregulation is involved in neurodevelopmental disorders associated with autism and intellectual disability. mTOR regulates assembly of the translation initiation machinery by interacting with the eukaryotic initiation factor eIF3 complex and controlling phosphorylation of key translational regulators. Collybistin (CB), a neuron-specific Rho-GEF responsible for X-linked mental retardation with epilepsy, also interacts with eIF3 and its binding partner gephyrin associates with mTOR. However, studies addressing whether CB also binds mTOR and affects mTORC1-signaling activity are not yet available. To test these hypotheses, we used two model systems: heterologous expression in HEK293T cells and induced pluripotent stem cells (iPSC)-derived neural progenitor cells (iNPCs) from a patient with a CB gene deletion and from normal subjects. Our results suggest that CB forms a complex with mTOR both in 293T cells and in control iNPCs. We also show that CB inhibits mTORC1-signaling pathway and protein synthesis. We suggest a model whereby CB forms a complex with mTOR and eIF3 and by sequestering these proteins down-regulates mTORC1 signaling and protein synthesis. These findings also suggest that disinhibited mTORC1 signaling mays also contribute to the pathological process in patients with loss-of-function mutations in CB. Elucidation of the signaling network regulating protein synthesis in neuronal cells is essential for understanding the pathological process of several neurological disorders and further studies on the involvement of CB in translation control will likely prove highly fruitful. Support: FAPESP, CNPq, Autismo & Realidade.

3077M

De novo dominant mutations in the kinesin motor protein KIF1A cause a severe static or progressive encephalopathy with cerebral and cerebellar atrophy. *S. Esmaeeli Nieh[1], M. Madou[1], B. Fregeau[1], A.J. Barkovich[2], J.G. Pappas[3], B. Hallinan[4], C.G. Spaeth[5], J.B. Strober[1], N. Smaoui[6], E.H. Sherr[1].* 1) Neurology, University of California, San Francisco, San Francisco, CA; 2) Department of Radiology, University of California, San Francisco CA; 3) 3Department of Pediatrics, Clinical Genetic Services, NYU School of Medicine, New York, NY; 4) Division of Pediatric Neurology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 5) Division of Human Genetics, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 6) GeneDx, Gaithersburg, MD.

The causes of developmental encephalopathies are complex and often remain poorly understood. Arriving at a diagnosis can have important implications for clinical management and prognosis. Clinically indicated whole exome sequencing was performed on 750 trios in which the proband had a range of neurologic conditions including encephalopathy and global developmental delay. Four individuals were identified with de novo missense mutations in KIF1A, the major plus-end directed microtubule motor in neuronal axons. Two patients had the same mutation (T99M) in the ATP binding domain; two patients had additional novel mutations (pR216C, E253K) in the motor domain. All four patients had microcephaly, cortical visual impairment, global hypotonia with spasticity and severely impaired development. Movement disorders and epilepsy were also observed. Three patients had evidence of progressive neurologic disease both clinically and radiologically, with supratentorial and cerebellar white matter and grey matter volume loss seen on sequential MRI scans. To predict the effect of KIF1A motor domain mutations, we performed protein-structure-based prediction using Schrödinger, and compared all de novo mutations with the previously reported recessive mutations as well as reported population polymorphisms. Results showed that dominant mutations were located within 5 Å of the ATP binding site recessive mutations were along the interface with tubulin. All 8 reported polymorphisms from the exome variant server database were located at neither of these sites. To address the functional consequences of these mutations, we are performing a gliding assay of tubulin over kinesin using the KIF1A motor-domain with wild type and both dominant and recessive mutant constructs. Recessive mutations in KIF1A were previously demonstrated to cause hereditary motor and sensory neuropathies. We describe a novel and consistent clinical phenotype resulting from de novo dominant highly penetrant mutations in the ATP and microtubule binding domains of KIF1A, the most severe form leading to a rapidly progressive early childhood onset neurodegenerative disease. Given the severity of these dominant mutations in comparison to silent recessive mutations, it is critical that we develop tools to predict clinical severity for family-based anticipatory guidance. Early data from prediction algorithms suggest that this is possible and ongoing in vitro and cell-based assays will address these directly.

3078T

Clinical and molecular characterization of progressive encephalopathies in children. *J.R. Helle[1], D. Misceo[1], T. Barøy[1], C. Ramane[1], A. Holmgren[1], T. Hughes[1], M.D. Vigeland[1], S. Thirukeswaran[1], S.S. Amundsen[1], T.H. Gamage[1], A. Torgersbråten[1], N. Skauli[1], I. Akkouh[1], E.O. Carlsen[1], M. Fannemel[1], P. Strømme[2], E. Frengen[1].* 1) Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway; 2) Women and Children's Division, Department of Clinical Neurosciences for Children, Oslo University Hospital, Ullevaal, Oslo, Norway.

BACKGROUND AND OBJECTIVES: Progressive encephalopathies (PE) in children consist of clinically and genetically heterogeneous disorders.They often affect brain morphology or signaling intensity visualized on cerebral MRI and lead to developmental arrest before subsequent regression of acquired skills. Metabolic defects or other neurodegenerative mechanisms result in atrophy of already existing neural cells. Neural tissues, such as spinal cord, peripheral nerves, and retina may also be involved. Intellectual disability, epilepsy, ataxia, spasticity and dystonia frequently accompany PE. Mortality is high. Approximately 20% of PE lack a diagnosis, and this gap in knowledge about etiology impairs genetic counseling and treatment. Once a causative gene is identified, clinical data suggesting time and focus of the incipient PE may also provide valuable insight on this gene's implication in normal brain functioning. Our objectives were to characterize mutations in a series of undiagnosed cases of PE and to perform functional analyses in novel disease genes. METHODS: We have collected more than 70 patients in 50 families presumed to have a genetic etiology, and where extensive medical work up had been negative. We ruled out CNS infection, trauma, vascular accidents and sequelae after asphyxia and prematurity. Cerebral MRI examinations revealed cortical atrophy, subcortical white matter changes, cerebellar degeneration, or basal ganglia abnormalities in the majority of the patients. Congenital anomalies, also those outside the nervous system, were evaluated using the London Medical Database for syndrome identification. After karyotyping, aCGH and analysis of candidate genes by MLPA and/or sequencing, we have so far performed Whole Exome Sequencing (WES) in 50 family trios and continued with data filtering, alignment, variant calling and selection based on population frequency, estimated severity of variants and inheritance pattern. RESULTS: Among the 20 WES trios finalized through the bioinformatics pipeline, a putative disease causing mutation has been identified in ten. Functional studies are currently being performed in two novel disease genes and in one novel disease entity affecting peroxisomal functioning. CONCLUSION: Clinical and molecular characterization, particularly via WES, provides additional etiologic and pathogenic criteria to classical clinical criteria in childhood PE taxonomy.

3079S

Low levels of CHIP in fibroblasts derived from patients with autosomal recessive cerebellar ataxia caused by mutations in *STUB1*. *S. Johansson[1], K. Heimdal[2], M. Sanchez-Guixé[1], I. Aukrust[3], J. Bollerslev[4], A.K. Erichsen[5], E. Gude[6], J.A. Koht[7], S. Erdahl[3], T. Fiskerstrand[1], B.I. Haukanes[3], L.G. Bjorkhaug[1], C.M.E. Tallaksen[8], P.M. Knappskog[3].* 1) Department of Clinical Science, University of Bergen, Bergen, Bergen, Norway; 2) Department of medical genetics, Oslo University hospital, Oslo, Norway,; 3) Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 4) Section of Specialized Endocrinology, Medical Clinic B, Oslo University Hospital; 5) Department of ophthalmology, Oslo University Hospital; 6) Department of Cardiology, Oslo University Hospital Rikshospitalet, Oslo, Norway; 7) Department of Neurology, Vestre Viken Hospital, Drammen; 8) Department of Neurology, Oslo University Hospital, Norway.

A subset of hereditary cerebellar ataxias is inherited as autosomal recessive traits (ARCAs). Recently, reports have linked mutations in genes involved in ubiquitination (*RNF216, OTUD4, STUB1*) to ARCA with hypogonadism. We used homozygozity mapping and exome sequencing, to identify a homozygous missense variant (c.194A>G, p.Asn65Ser) in STUB1 in three consanguineous siblings with ARCA and cognitive impairment, and a missense change (c.82G>A, p.Glu28Lys) which was inherited in trans with a nonsense mutation (c.430A>T, p.Lys144Ter) in another patient. *STUB1* encodes CHIP (C-terminus of Heat Shock protein 70 - interacting protein), a dual function protein with a role in ubiquitination as a co-chaperone with heat shock proteins, and as an E3 ligase. We show that the p.Asn65Ser substitution impairs CHIPs ability to ubiquitinate HSC70 in vitro, despite being able to self-ubiquitinate. These results are consistent with previous in vitro studies highlighting this as a critical residue for the interaction between CHIP and its co-chaperones. Furthermore, we show that the levels of CHIP are strongly reduced in vivo in patients' fibroblasts compared to controls. These results suggest that STUB1 mutations might cause disease by impacting not only the E3 ligase function, but also its protein interaction properties and protein amount. Whether the clinical heterogeneity seen in *STUB1* ARCA can be related to the location of the mutations remains to be understood, but interestingly, all siblings with the p.Asn65Ser substitution showed a marked appearance of accelerated aging not previously described in *STUB1* related ARCA, none display hormonal aberrations / clinical hypogonadism while some affected family members had diabetes, alopecia, uveitis and ulcerative colitis, further refining the spectrum of *STUB1* related disease.

3080M

Truncating mutations in the negative feed-back regulator of interferon 1 signalling, *USP18* gene causes pseudo-TORCH syndrome. *G.M.S. Mancini[1], M.E.C.M. Meuwissen[1], R. Schot[1], S. Tinschert[5], L. van Unen[1], D. Heijsman[2], W.F.J. van IJcken[3], J.M. Kros[4], R. Willemsen[1], F.W. Verheijen[1].* 1) Clinical Genetics, Erasmus University Medical Center, Rotterdam, Netherlands; 2) Bioinformatics, Erasmus University Medical Center, Rotterdam, Netherlands; 3) Biomics core, Erasmus University Medical Center, Rotterdam, Netherlands; 4) Pathology, Erasmus University Medical Center, Rotterdam, Netherlands; 5) Medical genetics, University of Innsbruck, Austria.

Inappropriate stimulation of the type 1 interferon (IFN1) response pathway has been proposed as the disease mechanism in autoimmune disorders such as systemic lupus erythematosus. Also disorders resembling congenital viral infections such as pseudo-TORCH and Aicardi-Goutières syndrome have been ascribed to mutations in receptors and enzymes that lead to cytosolic accumulation of nucleic acids as triggers, indirectly linking these disorders to the IFN1 pathway. The general term “type 1 interferonopathies” is now used to indicate these disorders. We report on two unrelated families with five affected individuals with a pseudo-TORCH-like phenotype, presenting in the perinatal period with severe brain hemorrhage, destruction of ependymal layer and early demise.In the first consanguineous family including 3 affected sibs we used linkage analysis, whole exome sequencing and Sanger sequencing of the non-covered exons in the linkage area and identified a homozygous truncating mutation in *USP18* causing total absence of *USP18* transcript in cultured fibroblasts. In the second family including two sibs the same truncating mutation in heterozygous form was found. However, no transcript was amplified from cells of the two patients, suggesting that a second pathogenic mutation is present in areas not covered by Sanger sequencing. *Usp18* knockout mice have cerebral hemorrhage and hydrocephalus with ependymal necrosis. Microscopy of patient brain tissues shows severe ependymal abnormalities, similar to the mouse. *USP18* encodes UBP43, a protein with a dual independent function in cleavage of ubiquitin-like protein 15 (ISG15) conjugates, through its carboxypeptidase domain, and, through its SH2 domain, a function as negative feed back regulator of the intracellular interferon-dependent JAK-STAT signalling pathway. *USP18* mutations therefore result in a sustained, deleterious and deregulated interferon response. *USP18* mutations are a novel genetic cause of pseudo-TORCH syndrome and severe cerebral hemorrhage. While the other interferonopathies are linked to inappropriate stimulation of the IFN1 response pathways, this is the first example of defective negative feed-back regulation of the IFN1 pathway, which might have consequences for the design of future therapies.

3081T

The Expanding Phenotype of TRPV4 Related Neuropathies With Notable Intrafamilial Variability. *L. Medne[1], C. Bonnemann[2], S. Scherer[3, 4], R. Finkel[5], X. Ortiz-Gonzalez[1], A. Glanzman[6], T. Estilow[7], A. Moll[7], R. Leshner[8], Y. Wang[9], T. Winder[9], S. Yum[1, 4].* 1) Division of Neurology, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) NINDS, NIH, Bethesda, MD; 3) Division of Neurology, The Hospital of the University of Pennsylvania; 4) School of Medicine, The University of Pennsylvania; 5) Nemours Children's Hospital, Orlando, FL; 6) Department of Physical Therapy, The Children's Hospital of Philadelphia, Philadelphia, PA; 7) Department of Occupational Therapy, The Children's Hospital of Philadelphia, Philadelphia, PA; 8) Division of Neurology, Rady Children's Hospital, San Diego, CA; 9) Prevention Genetics Laboratory, Marshfield, WI.

TRPV4 gene mutations are known to cause 5 skeletal dysplasias, 1 arthropathy and 3 forms of neuropathies: congenital distal spinal muscular atrophy (CDSMA), scapuloperoneal spinal muscular atrophy (SPSMA) and Charcot-Marie-Tooth disease type 2C (CMT2C). We report 5 patients with TRPV4 related neuropathies, three of whom have positive family histories. Three patients share the common p.Arg269His mutation but have varied phenotypes. Patients 1 and 2 presented with arthrogryposis multiplex congenital (AMC), hip dysplasia, scoliosis, distal>proximal weakness in lower extremities (LE) with limited walking, and weakness of proximal upper extremities (UE). Patient 3 with the p.Arg269His has CMT2 with pure axonal neuropathy, which she inherited from her subclinically affected mother with evidence of mild chronic denervation on EMG. Her maternal uncle and cousin, also with p.Arg269His, had AMC, motor axonal neuropathy and severe course: uncle lost ambulation in his 40s; her cousin never walked. Patient 4 has AMC with limited walking since 2.5 years, distal>proximal weakness in LE, mild proximal UE weakness and prominent scapulae. He has 2 mutations: p.Val620Ile - seen in CMT2C and brachyolmia type 3, but not in CDSMA; and p.Arg151Trp seen with low frequency in dbSNP. Patient 5 is a 3 y/o with gait abnormalities and axonal neuropathy with a similarly affected father and grandmother. His p.Arg186Gln is a known CMT2C mutation. All five probands reported here had clear motor axonal neuropathies on nerve conduction and EMG studies. All had muscle fasciculations noted during their muscle ultrasound examinations that showed an overall neurogenic pattern of echogenicity The broad phenotypic spectrum of TRPV4 neuropathies represents both a diagnostic and clinical management dilemmas. There is no definitive genotype : phenotype correlation, particularly as demonstrated by the presence of both CMT2C and CDSMA phenotypes in the same family (patient 3) reported here. All three patients with congenital arthrogryposis presentation also had proximal upper extremity weakness and patients 2 and 4 had scapular winging, showing an overlap between CDSMA and SPSMA phenotypes rather than completely distinct clinical entities. TRPV4 gene mutations should be considered in all cases of AMC presentation with evidence of axonal neuropathy.

3082S

WWOX and severe early onset epileptic encephalopathies: description of two additional patients and new clinical insights. *C. Philippe[1, 2], C. Mignot[3], L. Pasquier[4], I. lambert[5], L. Allou[2], J.L. Guéant[2], P. Jonveaux[1, 2].* 1) Laboratory of Medical Genetics, Centre Hospitalier Universitaire, Vandoeuvre les Nancy, France; 2) Laboratoire INSERM-U954-NGERE-Nutrition Génétique et Exposition aux Risques Environnementaux, Faculté de Médecine, Université de Lorraine, Vandoeuvre les Nancy, France; 3) Département de Génétique, Centre de Référence "Déficiences Intellectuelles de Causes Rares"; APHP; Groupe Hospitalier Pitié Slapêtrière, Paris, France; 4) Service de Génétique Clinique, Centre de Référence CLAD Ouest, CHU Rennes, Rennes, France; 5) Unité de Génétique Clinique, Médecine Néonatale, Maternité Régionale Universitaire, Rue du Dr Heydenreich, Nancy, France.

WWOX on chromosome 16q was initially identified as a putative tumor suppressor gene considering that it is frequently altered by deletions or translocations in many neoplastic lesions. The 46-kDa WWOX protein contains two N-terminal WW domains and a central short-chain dehydrogenase/reductase domain. It has been implicated in many biological processes including tumor suppression, metabolic disorders, immune defects, bone tumors, neurodegenerative diseases, and early onset severe epileptic encephalopathies. The description of individuals carrying germline WWOX mutations or deletions is recent. The first germline rearrangement of WWOX in a heterozygous state was reported in a child with ambiguous genitalia (White et al. Eur J Hum Genet 2012). Recently, WWOX was implicated in autosomal recessive spinocerebellar ataxia-12 (SCAR12) in two consanguineous families with homozygous missense mutations (Mallaret et al., Brain 2014). In another consanguineous family, a homozygous early stop codon was identified in an Egyptian girl affected by a severe lethal neurologic phenotype resulting in death at age 16 months (Abdel-Salam, Orphanet J Rare Dis 2014). Early onset severe epileptic encephalopathies (EOEE) are genetically heterogeneous disorders. By comparative genomic hybridization, we initially identified a heterozygous composite genotype consisting of two large deletions in WWOX in a female with EOEE. We subsequently screened the WWOX locus for point mutations in patients with EOEE and heterozygous for an inherited copy number variation affecting the WWOX locus. We characterized a second patient positive for WWOX alterations affecting both alleles. Our study confirms that biallelic WWOX mutations are a (not so rare?) cause of autosomal recessive early onset epileptic encephalopathies with clinical heterogeneity depending on the mutation(s) type(s). WWOX-related severe encephalopathy is characterized by 1) normal pregnancy and delivery or mild intrauterine growth retardation 2) severe developmental delay, poor spontaneous motility and absent eye contact from birth, little if any acquisitions thereafter 3) early-onset pharmacoresistant epilepsy starting in the first two months of life 4) compromised survival during the first years. This study also demonstrates the need for combining data from array comparative genomic hybridisation and next generation sequencing for the identification of genes responsible for rare autosomal recessive disorders.

3083M

Mechanisms leading to brain malformations in tubulinopathies. *M.I. Rees[1,2], T.D. Cusion[1], S-K. Chung[1,2], R. Oegema[3], A. Paciorkowski[4], W.B. Dobyns[5], D. Doherty[5], D.T. Pilz[2,6].* 1) Neurology & Molecular Neuroscience Research, College of Medicine, Swansea University, UK; 2) Wales Epilepsy Research Network, College of Medicine, Swansea University, UK; 3) Department of Clinical Genetics and Expertise Centre for Neurodevelopmental Disorders, Erasmus University Medical Center, Rotterdam, The Netherlands; 4) Departments of Pediatrics and Biomedical Genetics, University of Rochester Medical Center, Rochester, NY, USA; 5) Center for Integrative Brain Research, Seattle Children's Hospital and Department of Pediatrics, University of Washington, Seattle, USA; 6) Institute of Medical Genetics, University Hospital Wales, Cardiff University, UK.

The tubulinopathy spectrum of overlapping brain malformations, including lissencephaly, atypical polymicrogyria and mildly disorganized gyral patterning, arise from tubulin gene variations affecting neuronal proliferation, migration and post-migrational organisation during cerebral cortex development. We ascertained a series of children with these cortical malformations, and/or characteristic basal ganglia and cerebellar malformations, and screened them for gene-variations in four human brain-expressed tubulin genes (*TUBA1A, TUBB2B, TUBB2A* and *TUBB3*). Although many variants in these genes have been reported by us and other groups, the in vitro functional consequences for the putatively causative mutations have only partially revealed cellular mechanisms. Here we present data where 3 different mechanisms are becoming apparent. These include intracellular tubulin disorganisation as detected by immunocytochemistry and functional domain specificity; neurite-length defects as detected in outgrowth assays; and possible cell-cycle delays in the G2/M phase as indicated by flow cytometry. This varied collection of mechanisms may underlie the clinical and structural heterogeneity observed in individuals with tubulinopathies, and provides contextual detail before the design of in vivo models for interventional research.

3084T

A novel pathogenic mechanism in Hereditary Spastic Paraplegia. *R. Schüle[1,2,3], A. Rebelo[1], M. Bonin[4], E. Battaloglu[5], G. Woehlke[6], L. Schöls[2,3], S. Zuchner[1], A. Caballero Oteyza[2,3].* 1) Hussman Institute for Human Genomics, Miami, FL; 2) Hertie Institute for Clinical Brain Research, Tubingen, Germany; 3) German Center for Neurodegenerative Disease, Tubingen, Germany; 4) Institute of Medical Genetics and Applied Genomics, Tübingen, Germany; 5) Bogazici University, Department of Molecular Biology and Genetics, Istanbul, Turkey; 6) Department of Physics E22 (Biophysics), Technical University Munich, Garching, Germany.

Hereditary Spastic Paraplegias (HSP) are characterized by a length dependent axonopathy of corticospinal tract motor neurons. With more than 70 genetic subtypes described, HSPs are among the most heterogeneous Mendelian disorders. Known HSP genes are involved in numerous cellular processes including membrane shaping, intracellular membrane transport, metabolism of complex lipids and mitochondrial function. Here we describe the discovery and functional characterization of a novel autosomal recessive HSP disease gene. The gene was identified via whole exome sequencing in a family with spastic ataxia and confirmed in a second family with a similar phenotype. Missense mutations observed in our family lead to reduced protein stability and dramatic subcellular misslocalization. The novel disease gene has been poorly studied so far. Using mass spectrometry we identified binding partners of this novel HSP-associated protein. Interaction was later confirmed using immunoprecipitation with two known RNA binding proteins: Polyadenylate-binding protein 1 (PABP-1) and Insulin-like growth factor 2 mRNA-binding protein 1 (IMP-1). PABP-1 and IMP-1 are known to be part of a multisubunit autoregulatory ribonucleoprotein complex that is involved in regulation of mRNA translation and nonsense mediated decay. To confirm the interaction of the novel HSP protein with RNA we performed RNA immunoprecipitation and were indeed able to precipitate mRNA using antibodies directed against the novel HSP protein. Further studies to reveal the identity of the bound mRNA and further functional consequences of pathogenic mutations are underway. To our knowledge this is the first report linking HSP to RNA metabolism. Disturbances of RNA metabolism with formation of RNA stress granules and potentially more widespread alteration of RNA metabolism are however increasingly recognized as a key pathogenic factors of motor neuron death in Amyotrophic Lateral Sclerosis, a related motor neuron disease.

3085S

DNM3; a genetic modifier of LRRK2 parkinsonism. *J. Trinh[1], E. Gustavsson[1], I. Guella[1], C. Vilarino-Guell[1], C. Thompson[1], H. Han[1], S. Bortnick[1], H. Sherman[1], D.M. Evans[1], M. Toft[2], M. Petersen[3], J. Aasly[4], F. Hentati[5], M.J. Farrer[1].* 1) Djavad Mowafaghian Centre for Brain Health, Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada, V6T 2B5; 2) Department of Neurology, Oslo University Hospital, Oslo, Norway; 3) Department of Occupational Medicine and Public Health, The Faroese Hospital System, Tórshavn, Faroe Islands; 4) Department of Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway; 5) Service de Neurologie, Institut National Mongi Ben Hamida de Neurologie, La Rabta, Tunis, Tunisia.

Background: *LRRK2* p.G2019S accounts for 30-40% Parkinson disease (PD) in North African Arab Berbers, 18-30% in Ashkenazi Jews and 1-3% in Caucasians. Disease penetrance of *LRRK2* p.G2019S is variable, may result in early-onset parkinsonism or elderly asymptomatic carriers, and we hypothesize additional genetic factors are responsible.

Methods: Genome-wide STR linkage analysis for age-of-onset (AAO) of disease was evaluated in multi-incident *LRRK2* p.G2019S families (126 affected and 71 asymptomatic carriers). Case-only haplotype association of AAO was performed with Plink, Snipsnip and Beagle in 266 *LRRK2* p.G2019S Arab-Berber patients. Analysis of gene expression and transcript isoforms in striatum was evaluated using Taqman RT-PCR and Ion Torrent PGM. Sanger and next-generation sequencing were performed for *DNM3* (n=400 PD and controls), with subsequent Sequenom iPLEX MassArray genotyping of coding substitutions in a Caucasian case-control series (n= 1999).

Results: Linkage of *LRRK2* p.G2019S AAO gave a LOD=2.9 on chromosome 1, and suggested regions on 11, 21 and X (LOD>2.0). A significant haplotype association for AAO was only observed on chromosome 1 ($p<0.00001$), spanning *DNM3*. The SNPs responsible appear to regulate gene transcription. Subsequent sequencing in Caucasian subjects showed a 2-fold increase in *DNM3* substitutions in PD compared to control subjects, of which *DNM3* p.M664V segregates with parkinsonism within 5 families, albeit with reduced penetrance.

Conclusions: *DNM3* is a penetrance modifier of *LRRK2* parkinsonism, as shown by linkage, association and sequencing analyses, and highlights how a homogeneous population may yield genetic modifiers of Mendelian disease. Importantly, chromosome 1 linkage for AAO was previously reported in Ashkenazi Jewish *LRRK2* p.G2019S carriers, whereas co-immunoprecipitation/co-localization has shown a biological interaction between LRRK2 and DNM1-3. Findings suggest the regulation of protein sorting and trafficking in neurons is central to the etiology of PD.

3086M

Novel de novo sequence variation in *HNRNPU* gene is associated with generalized epilepsy responsive to ketogenic diet. *R. Veith[1], L. Carey[1], D. Helbling[2], M. Tschannen[2], J. Wendt-Andrae[2], M. Waknitz[2], G. Scharer[1,2].* 1) Pediatrics-Genetics, Children's Hospital of Wisconsin, Milwaukee, WI, United States; 2) HMGC, DNL-Seq Laboratory, Dept. of Pediatrics, Medical College of Wisconsin, Milwaukee, WI, United States.

Infantile and childhood onset seizure disorders and epileptic encephalopathies (EE) are a heterogeneous group of disorders for which there is established evidence of genetic contribution. Novel or de-novo sequence changes play an important role in seizure disorders and have been described in several known and new genes through the use of whole exome sequencing (WES). Our patient is a 27 month old Caucasian female with onset of generalized seizures between 10-12 months of age, which initially appeared to be single febrile events. Mild developmental regression was noted following the second and third seizure. At 12 months the patient experienced her first cluster (afebrile) seizures, was admitted with report of “normal” MRI and EEG (with diffuse slowing), and at that time was started on Phenobarbital. Several antiepileptic drugs were administered, but did not control seizures adequately. During the next 12 months she developed a severe EE and at peak experienced 20-25 seizures per day. The patient empirically began a ketogenic diet at 22 months of age and experienced significant relief of seizures within a month of initiation. This has dramatically improved her clinical course and she now has only 3-4 seizures per day; and has gone for up to 10 days without seizures. Extensive genetic evaluation (including comprehensive epilepsy gene panel), and metabolic work-up did not reveal a diagnosis. WES (SureSelect® v4) was pursued and identified a novel deleterious sequence variant in HNRNPU (p.ARG770TER). This de-novo variant was not detected in both parents or the healthy sister. The variant was also not identified in 10,000 reference exomes or control chromosomes. HNRNPU has been recently published in association with EE via point mutation and microdeletions. Individuals with deletions of chromosome 1q44 have been documented with significant seizure disorder. It has been suspected that an EE gene resides in this region. HNRNPU would be a likely candidate. The known predisposition genes for the EE work in a complex network and HNRNPU directly interacts with other well documented EE genes including PNKP. Interestingly, none of the previously documented cases of HNRNPU associated EE patients have achieved adequate seizure control or significant improvement with traditional pharmacologic management. Based on our patient’s symptomatic relief, ketogenic diet may be a reasonable alternative management for epileptic encephalopathy resulting from HNRNPU deficiency.

3087T

Modifiers of age at onset in spinocerebellar ataxia type 2: a preliminary study in a Brazilian population. *F.dosS. Pereira[1,10], T.L. Monte[4], L.D. Locks-Coelho[1], A.S.P. Silva[1], A.L. Silva[1], O. Barsottini[6], J.L. Pedroso[6], F.R. Vargas[7,8], M.A.F.D. Lima[7], H.V.D. Linden Jr[9], R.M. Castilhos[1,5], M.L. Saraiva-Pereira[1,2,5,10], L.B. Jardim[1,3,4,5,10] on behalf of Rede Neurogenetica.* 1) Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil; 2) Department of Biochemistry, Universidade Federal do Rio Grande do Sul, Brazil; 3) Department of Internal Medicine, Universidade Federal do Rio Grande do Sul, Brazil; 4) Post-Graduation Program in Medical Sciences, Universidade Federal do Rio Grande do Sul, Brazil; 5) Post-Graduation Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul, Brazil; 6) Department of General Neurology and Ataxias- Discipline of Clinical Neurology of UNIFESP - Escola Paulista de Medicina, Universidade Federal de São Paulo, Brazil; 7) Laboratory of Epidemiology of Congenital Malformations, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil; 8) Departament of Genetic and Molecular Biology, Universidade Federal do Estado do Rio de Janeiro, Brazil; 9) Rehabilitation Centre Dr. Henrique Santillo, Goiânia, Goiás, Brazil; 10) Instituto Nacional de Genética Médica Populacional, Brazil.

The spinocerebellar ataxia type 2 (SCA2) is caused by a CAGn expansion at *ATXN2* gene, and ATXN2 expansions account for 50% of the variability in age at onset (AO). Former reports pointed to CAGn variations and polymorphisms at other genes as responsible for the remaining AO variance. Aims: to address a confirmation study of other polyglutamine tracts and of a mitochondrial DNA polymorphism as modifiers of AO in SCA2 patients. Methods: symptomatic individuals with a molecular SCA2 diagnosis were recruited from Brazil. Capillary electrophoresis was performed to detect CAG lengths at SCA1, SCA2, SCA3/MJD, SCA6 and RAI1 associated genes; the mitochondrial complex I gene polymorphism (10398G>A) was determined by PCR followed by capillary electrophoresis analysis. Pearson correlations with AO were tested against each CAGn for all individuals; the 10398G>A polymorphism of one person per maternal lineage was analysed by t test; all followed by a step-wise linear regression. Results: 83 individuals (33 families and 42 maternal lineages) were studied. Mean (range) AO and CAGn at normal and expanded ATXN2 alleles were 32.45 (3-76) years and 22 (22-33) and 42 (34-67) repeats. At first, AO correlated with the large alleles at *ATXN2* with the short allele of *ATXN2*, and with short allele at *RAI1*. 10398G>A polymorphism was not associated with AO. On step-wise regression, only the correlation with *ATXN2* expanded allele was maintained ($r = -0.78$; $r2 = 0.61$; $p < 0.0001$). Discussion: This data confirms the link between the expanded allele of the ATXN2 and antecipation of AO of symptoms and reveals conflicting results to those previously reported for both the mitochoondrial polymorphism 10398G>A and for the small allele of the *RAI1* gene. Preliminar data should be confirmed by an outlier sampling strategy, in the future.

3088S

Defining the presence of GTF2IRD1 in epigenetic complexes as a means to understand features of Williams-Beuren syndrome. *P. Carmona-Mora[1], F. Tomasetig[1], C.P. Canales[1], A. Alshawaf[2], M. Dottori[2], E.C. Hardeman[1], S.J. Palmer[1].* 1) Cellular and Genetic Medicine Unit, School of Medical Sciences, UNSW Australia, Sydney, New South Wales 2052, Australia; 2) Centre for Neural Engineering, The University of Melbourne, Melbourne, VIC 3010, Australia.

GTF2IRD1 is a member of the *GTF2I* gene family, which lies on chromosome 7 within a 1.8 Mb region that is prone to duplications and deletions leading to a series of physical and neurobehavioural abnormalities. Hemizygous deletions cause Williams-Beuren syndrome (WBS) and duplications cause WBS duplication syndrome. Genotype-phenotype correlations of patients with atypical deletions and knockout mouse models show *GTF2IRD1* as responsible for the craniofacial abnormalities, mental retardation, visuospatial construction deficits and hypersociability of WBS. In order to understand the basis of the phenotypes related to GTF2IRD1, we have studied its molecular function *in vitro*. We evaluated its expression in human ES cell-derived neurons, to correlate it with specific neuronal types and stages of differentiation. Furthermore, a detailed analysis in cell lines of endogenous GTF2IRD1 shows a specific nuclear speckled pattern of expression, colocalizing with markers of gene silencing complexes. To define GTF2IRD1 functional relationships, we used a yeast two-hybrid approach to isolate interaction partners and validated many interactions in mammalian cells. Most of the proteins are involved in chromatin modification and transcriptional regulation in agreement with the nuclear localization data. In addition, several of these partners, such as ZMYM2, ZMYM3, ZC4H2 and DCAF6, were shown to occupy the same nuclear compartment as GTF2IRD1 at the endogenous level. Some of these proteins are involved in histone deacetylase (HDAC) complexes related to neuron-specific gene regulation and others are directly linked to mental retardation through mutations in their encoding gene. We have demonstrated the presence of GTF2IRD1 in chromatin modifier complexes and identified direct associations with HDACs, allowing us to expand its interactional network and propose functional links with specific transcription factors and epigenetic silencing complexes. Within the same context, we also explored GTF2IRD1 capability to bind nucleic acids and tested the dependence of relevant protein-protein interactions upon such binding. Our data indicates that GTF2IRD1 forms complexes with DNA binding and chromatin modifying proteins that direct gene regulation through epigenetic mechanisms. For the first time, we provide an empirical account of the molecular basis of GTF2IRD1 function and permit the construction of testable hypotheses regarding its role in WBS and 7q11.23 duplication syndrome.

3089M

Generation of a comprehensive panel of patient-derived pluripotent stem cells to dissect oligodendrocyte dysfunction in the pediatric myelin disorder Pelizaeus-Merzbacher Disease. *Z. Nevin[1], R. Karl[1], G. Hobson[2], P. Tesar[1,3].* 1) Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH; 2) Nemours Biomedical Research, Alfred I. duPont Hospital for Children, Wilmington, DE; 3) New York Stem Cell Foundation, New York, NY.

Patient-derived induced pluripotent stem cells (iPSCs) have become integral tools for parsing the molecular biology of complex human genetic disorders. Pelizaeus-Merzbacher Disease (PMD [MIM 312080]) is a pediatric leukodystrophy affecting central nervous system myelin that results in severe motor impairment, intellectual delay, and premature death and has no available treatment. Though PMD is a single-gene disorder, it exhibits a spectrum of phenotypes and clinical severities attributed to over 200 different mutations in the essential myelin gene *PLP1* (MIM 300401), which encodes for two protein isoforms, proteolipid protein 1 (PLP1) and DM20. Protein misfolding and endoplasmic reticulum stress have been implicated in the pathogenesis of PMD in transgenic animal models overexpressing PLP1, but this may not hold true for the many various point mutations and gene deletions identified in patients. Detailed investigations of PLP1's normal function and the etiology of PMD are further complicated by a lack of access to primary human oligodendrocytes, the only cell type in which PLP1 protein is known to be expressed. To address this need, we have generated a large panel of patient-derived iPSCs from children with unique duplication, triplication, deletion, and point mutations in *PLP1* that encompass the genotypic and phenotypic variation seen in patients. Using new methods optimized in our lab, we have derived pure populations of oligodendrocytes and oligodendrocyte progenitors from these iPSCs in order to dissect individual patient phenotypes at a cellular level. Access to this unique resource is now allowing us to identify the cellular pathways and processes disrupted by particular *PLP1* mutations, which in turn will inform the pursuit of patient-specific chemical and genetic therapies.

3090T

The NINDS Repository collection of patient-derived fibroblasts and induced pluripotent stem cells for neurodegenerative disease research. *C.A. Pérez[1,4], S. Heil[1,4], J. Santana[1,4], C. Grandizio[2,4], S. Gandre-Babbe[2,4], K. Hodges[3,4], M. Sutherland[5], R.A. Corriveau[5], C. Tarn[1,4].* 1) NINDS Repository; 2) Stem Cell Biobank; 3) Cell Culture Laboratory; 4) Coriell Institute for Medical Research, Camden, NJ; 5) National Institute of Neurological Disorders and Stroke - NIH, Bethesda, MD.

Induced pluripotent stem cells (iPSCs) reprogrammed from patient-derived primary fibroblasts have become an increasingly utilized and needed resource for the study of human disease and have proven especially valuable in studying neurodegenerative disorders for which disease models are difficult to establish. The National Institute of Neurological Disorders and Stroke (NINDS) Repository is a public resource established in 2002 aiming to provide a centralized and open collection of biological samples (DNA, lymphoblastoid cell lines, fibroblasts, iPSCs, biofluids such as plasma, serum, cerebrospinal fluids, and urine) and associated de-identified clinical data from a diverse population of patients and normal controls. Since 2011, the NINDS Repository has added to its web-based catalog (http://ccr.coriell.org/NINDS) close to 50 iPSC and 165 fibroblast lines. Most iPSC lines are contributed by Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS) or Huntington's disease (HD) investigators from NINDS-sponsored Stem Cell Consortia. By making available to the research community neurodegenerative disease iPSCs and fibroblasts, the NINDS Repository continues to fulfill the NINDS mission of reducing the burden of neurological disease - a burden borne by every age group, by every segment of society, by people all over the world. To ensure the quality of these valuable resources, all iPSCs and fibroblasts submitted to the NINDS Repository by iPSC Consortia and other investigators undergo rigorous quality assessments (viability, pluripotency, karyotyping, differentiation status, gene expression analysis, sterility) prior to distribution by the NINDS repository. The results are summarized in a Certificate of Analysis and/or displayed on the web-based catalog along with the recommended culturing protocol. The NINDS Repository fibroblast and iPSC collections include mostly cell lines bearing specific genetic mutations associated with PD, ALS, HD, frontotemporal degeneration, or Alzheimer's disease, as well as samples derived from neurologically normal controls. For certain affected individuals, the parental fibroblast, corresponding iPSC line, and whole blood DNA are available. The NINDS Repository serves as a unique and effective centralized resource where these iPSCs, fibroblasts and their de-identified phenotypic data, are available to basic and applied research investigators worldwide.

3091S

Human iPSC model of the Ras/MAPK pathway role in neurodevelopmental disorders. *E. Yeh[1], Z.Y. Wu[1], C. Tom[1], B. Adviento[1], K.A. Rauen[2], L.A. Weiss[1].* 1) Department of Psychiatry, UCSF, San Francisco, CA; 2) Department of Pediatrics, UC Davis, Sacramento, CA.

Germline mutations in the primordial Ras/MAPK pathway cause a class of human genetic syndromes termed RASopathies characterized by overactivation of this pathway. Cardiofaciocutaneous syndrome (CFC) is a rare RASopathy characterized by heart defects, a distinctive facial appearance, and ectodermal abnormalities. Neurological abnormalities include low IQ, learning disability, and speech delay. Autistic traits are present in 54-64% of the CFC patients, the highest percentage among RASopathies. To better understand the role of the Ras/MAPK signaling pathway in neurodevelopment, we established induced pluripotent stem cell (iPSC) lines from skin fibroblasts from 3 CFC patients carrying the most common mutation ($BRAF^{Q257R}$, present in 29% of the patients) and 3 sex- and age-matched controls. Next, we submitted the iPSC lines to neural differentiation and generated neural progenitor cells (NPC). Finally, we induced NPC to differentiate into neurons for characterization. $BRAF^{Q257R}$ NPC showed higher proliferation rates compared to wild-type (WT) NPC, as evidenced by higher percentage of KI67 positive cells (74.9±6.6% vs 50.7±7.5%, p=0.028) and shorter doubling time (41.9±7.8 vs 85.6±20.8 h, p=0.038). Seven days after induction, $BRAF^{Q257R}$ and WT cultures expressed similar levels of the neuron-specific markers Beta III tubulin, MAP2 and Tau protein. However, image analysis revealed that $BRAF^{Q257R}$ neurons had more neurites (3.5±0.1 vs 2.9±0.1 neurites/cell, p=0.0047) and longer neurites (156.2±17.6 vs 79.8±9.5 μm, p=0.0096). The early maturation in $BRAF^{Q257R}$ neurons was further evidenced by a higher number of cells positive for TBR1, a layer 5/6 cortical marker (44.07±5.8% vs 16.33±5.3%, p=0.0105). The higher NPC proliferation rate we observe is particularly interesting as >50% of CFC patients have macrocephaly. Previous literature shows that *Braf* null mice have shorter neurites and kinase-activated *Braf* mice have longer neurites, which are consistent with our observations. In addition, increased numbers of neurites and longer neurites are indicators of neuronal differentiation and maturation, which was supported by increased presence of a cortical layer 5/6 marker. Thus, we showed for the first time in human cells that a CFC mutation in *BRAF* is associated with early neuronal maturation. This study will help us obtain insight not only into the pathophysiology of CFC, but also into the role of Ras/MAPK pathway in common disorders, like Autism Spectrum Disorders.

3092M

New insights into the biological role of VAX2 in human, in health and disease. *G. Alfano[1], A.Z Shah[1], N.H. Waseem[1], A.R. Webster[1,2], S.S. Bhattacharya[1]*. 1) Institute of Ophthalmology, UCL, London EC1V 9EL, UK; 2) Moorfields Eye Hospital, London EC1V 2PD, UK.

Purpose: VAX2 (MIM 604295) is a transcription factor expressed in the ventral region of the prospective neural retina in vertebrates and is required for ventral eye specification. Its genetic inactivation in mouse leads to an incompletely penetrant ocular coloboma. Based on the similarities between the mutant mouse phenotype and the clinical features of isolated coloboma in human, VAX2 represents a good candidate gene for this human condition. Thus far, however, no mutations affecting VAX2 in coloboma patients have been reported. Interestingly, recently it has been shown that Vax2 in mouse plays a key role in ensuring appropriate gene expression in cone photoreceptor cells. Despite a wide range of studies in vertebrates its function in human is still poorly understood. This study was undertaken to better elucidate the role of VAX2 in human in health and disease. Methods: VAX2 transcripts were obtained by RT-PCR on cDNA from human tissues and cultured cells. These transcripts were sub-cloned into mammalian expression vectors for in vitro analysis. Variant proteins were generated by site-directed mutagenesis. A cohort of cone-rod dystrophy patients was screened for variant identification using Sanger sequencing. Results: It was observed that VAX2 is highly expressed in neuronal tissues and cultured cell lines. Two spliced variants were detected: isoform-1 (NM_012476), and a novel transcript (isoform-2), mainly expressed in retina and predicted to encode a shorter protein of 150 amino acids. In vitro exogenous expression of tagged protein isoforms showed differential localization potentially underlying distinct biological roles. Isoform-1 localizes to the nucleus while isoform-2 is widespread throughout the cell. The involvement of Vax2 in cone specific gene expression suggested screening of patients with cone-rod dystrophy. Sequence analysis revealed a heterozygous change c.416T>G (p.Leu139Arg) in one patient. Characterization of this variant showed damaging effects on protein localisation and is therefore suggestive of a functional defect, with potential involvement in the disease phenotype. Conclusions: Our data suggest that in addition to being a transcription factor VAX2 could be involved in other molecular pathways due to the short isoform in the human. Moreover, VAX2 may be involved in cone photoreceptor cell biology. In summary, our findings raise new questions about the VAX2 gene and its biological role in health and disease.

3093T

Digital gene expression differences in the OVT73 ovine model of Huntington's Disease. *S.J. Reid[1], R. Brauning[2], P. Maclean[3], S. Patassini[1], R.R. Handley[1], P. Tsai[4], H.J. Waldvogel[4], J.F. Gusella[5], M.E. MacDonald[5], R.L.M. Faull[1], R.G. Snell[1], and.the H.D.S.C.R.G[6]*. 1) Centre for Brain Research, The University of Auckland, Auckland, Auckland, New Zealand; 2) AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand; 3) AgResearch, Ruakura Research Centre, Hamilton, New Zealand; 4) School of Biological Sciences, The University of Auckland, Auckland; 5) Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston; 6) Huntington's Disease Sheep Collaborative Research Group (10.3233/jhd-130067).

In order to better understand the early molecular mechanisms of Huntington's Disease, we examine the OVT73 transgenic sheep for progressive histopathological and molecular changes. These animals express at modest levels the full length human huntingtin protein with 73 glutamine repeats [1], demonstrate reduced striatum levels of $GABA_A$ α1 receptor, cortical huntingtin aggregates [2] and an early circadian abnormality [3]. The sheep are healthy and visibly indistinguishable from control South Australian Merino. They present an opportunity to better understand and utilise the pre-symptomatic phase of Huntington's Disease.

To identify genome-wide gene expression differences, Illumina RNA-seq data (2.3×10^9 50bp PE reads) was collected from ribosomal depleted total RNA isolated from anterior striatum tissue of twelve 5-year old OVT73 sheep (6 transgenic, 6 control). For assembly of a *de-novo* transcriptome, further reads (1.3×10^9 100bp PE) were collected from brain and liver total RNA from fetal and neonate sheep. Differentially expressed transcripts in OVT73 were identified using both genome based approaches and the assembled transcriptome. A selection of these annotated transcript findings will be presented, including OXTR and AVPR1A which may relate to the circadian abnormality seen in the OVT73 animals.

1.Jacobsen, J.C., et al., An ovine transgenic Huntington's disease model. Hum Mol Genet, 2010. 19(10): p. 1873-82.

2.HDSCRG, et al., Further Molecular Characterisation of the OVT73 Transgenic Sheep Model of Huntington's Disease Identifies Cortical Aggregates. J Huntingtons Dis, 2013. 2(3): p. 279-295.

3.Morton, A.J., et al., Early and progressive circadian abnormalities in Huntington's disease sheep are unmasked by social environment. Hum Mol Genet, 2014.

3094S

De Novo mutations in TEAD1 and OCEL1 in non-X linked Aicardi Syndrome. *I. Schrauwen[1,2,3], S. Szelinger[1,2], A.L. Siniard[1,2], J.J. Corneveaux[1,2], A. Kurdoglu[1,2], R. Richholt[1,2], M. De Both[1,2], I. Malenica[1,2], S. Swaminathan[1,2], S. Rangasamy[1,2], N. Kulkarni[4], S. Bernes[4,5], J. Buchhalter[6], K. Ramsey[1,2], D.W. Craig[1,2], V. Narayanan[1,2], M.J. Huentelman[1,2]*. 1) Dorrance Center for Rare Childhood Disorders, Translational Genomics Research Institute, Phoenix, AZ; 2) Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ; 3) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; 4) Phoenix Children's Hospital, Phoenix, AZ; 5) MIHs integrated program, Phoenix, AZ; 6) Alberta Children's Hospital , Calgary, Alberta, Canada.

Aicardi syndrome (AIC) is a congenital neurodevelopmental disorder characterized by infantile spasms, agenesis of the corpus callosum and chorioretinal lacunae. There is a wide range in severity of these classic symptoms however, with chorioretinal lacunae as the most constant feature of the disease. Although cytogenetic studies have linked AIC to the X-chromosome, no gene has been identified to date, and little is known about the pathogenic mechanisms underlying the disease. Recently, XY males with the disease have been described as well, suggesting the presence of genetic heterogeneity. In this study, we investigated 10 female children diagnosed with AIC, and their parents, by exome/genome sequencing. In two cases, we identified de novo mutations in the autosomal genes TEAD1 and OCEL1, both expressed and important in the retina and brain. The mutation in TEAD1 is a nonsense mutation (p.Trp206Ter; CADD score = 40), and the mutation in OCEL1 is a non-synonymous mutation (p.Ala167Thr; CADD score = 15.3), which is predicted damaging according to MutationTaster, LTR, Polyphen2, and affects a highly conserved residue (PhyloP = 2.1). A missense mutation in TEAD1 has previously been associated with Sveinsson's chorioretinal atrophy, which is characterized by chorioretinal degeneration. In a second part of this study, we performed RNA-sequencing on RNA extracted from blood from 9 of these cases, their parents and unrelated controls. We found that altered expression of genes associated with neuronal synaptic plasticity is an important underlying potential pathogenic mechanism shared among cases, in addition to several seizure-associated genes and retinal developmental genes. Last, we found 1 case with skewed X-inactivation (XCI), supporting the idea that non-random XCI might be important in AIC, as previously reported. In conclusion, this is the first report of genes mutated in AIC, and proposal of putative pathogenic mechanisms underlying AIC. Our data suggest that AIC is a genetically heterogeneous disease and is not entirely restricted to the X-chromosome. This raises the important possibility that mutations in the genes detailed here may be present in male patients with similar phenotypic presentations. Investigation of the pathological mechanisms that lead to disease (retinal and neuronal) through OCEL1 and TEAD1 dysfunction might reveal novel therapeutic avenues for treatment and symptomatic management of the disease.

3095M

Absence of ER cation channel *TMEM38B*/TRIC-B causes recessive osteogenesis imperfecta by dysregulation of collagen post-translational modification. *W.A. Cabral[1], E.N. Makareeva[2], M. Ishikawa[3], A.M. Barnes[1], M.A. Weis[4], F. Lacbawan[5], D.R. Eyre[4], Y. Yamada[3], S. Leikin[2], J.C. Marini[1].* 1) Bone & Extracellular Matrix Branch, NICHD, NIH, Bethesda, MD, USA; 2) Section on Physical Biochemistry, NICHD, NIH, Bethesda, MD, USA; 3) Molecular Biology Section, NIDCR, NIH, Bethesda, MD, USA; 4) Department of Orthopaedics and Sports Medicine, University of Washington, Seattle, WA, USA; 5) Department of Medical Genetics, Children's National Medical Center, Washington DC, USA.

A novel form of recessive osteogenesis imperfecta (OI) has been reported in Israeli and Saudi Bedouins, caused by homozygous deletion of exon 4 and surrounding intronic sequence of *TMEM38B*. *TMEM38B* encodes TRIC-B, an integral ER membrane protein proposed to facilitate Ca^{++} release from intracellular stores. However, the molecular mechanisms through which this founder mutation causes OI are unknown. We identified a 20 month-old girl with moderately severe OI who is the offspring of consanguineous parents from Saudi Arabia, and is homozygous for the *TMEM38B* founder mutation. The proband presented with growth deficiency, osteopenia, recurring long bone fractures and blue sclerae. Proband *TMEM38B* transcripts are 25% of normal control level and include six spliced forms. Although one minor transcript is in-frame, absence of TRIC-B protein was confirmed by immunoblot. Consequently, proband cells have decreased intracellular and ER luminal Ca^{++} concentrations, which were more rapidly depleted upon ATP-induced Ca^{++} efflux from the ER. The lower Ca^{++} level does not involve changes in expression or stability of SERCA2b or IP3R, the ion channels involved in ER Ca^{++} loading and release, respectively. Furthermore, depletion of Ca^{++} is associated with ER stress, activation of ATF4 translation, and dysregulation of type I collagen synthesis. Although proband cells have normal PDI levels, the 20-minute delay in procollagen chain assembly into heterotrimeric molecules is consistent with sequestration of PDI by calreticulin in low Ca^{++} conditions. Abnormal modification of proband collagen is evident by increased electrophoretic migration of alpha chains on SDS-Urea PAGE, consistent with a 30% reduction in collagen lysine hydroxylation, despite increased LH1 expression. The presence of lower stability species (30-40% of secreted collagen) on DSC, decreased procollagen processing and increased intracellular accumulation of BiP, indicate abnormal procollagen conformation. However, extracellular matrix deposited by proband fibroblasts contained only collagen with normal thermal stability, and was associated with 30% reduction of crosslinked collagen in matrix. These data support a role for TRIC-B in maintaining intracellular Ca^{++}, and demonstrate that absence of TRIC-B causes OI by dysregulation of multiple Ca^{++}-regulated collagen-specific chaperones and modifying enzymes in the ER. Therefore, the mechanism of TRIC-B absence falls within the collagen-related paradigm of OI.

3096T

Exome sequencing identifies locus heterogeneity in multiple epiphyseal dysplasia. *K. Balasubramanian[1], M.J. Bamshad[2,3], D.A. Nickerson[3], R.S. Lachman[5], D.H. Cohn[1,2,4], University of Washington Center for Mendelian Genomics.* 1) Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, Los Angeles, CA; 2) Department of Pediatrics, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) Department of Orthopaedic Surgery and Orthopaedic Hospital Research Center, University of California, Los Angeles, Los Angeles, CA; 5) International Skeletal Dysplasia Registry, University of California, Los Angeles, Los Angeles, CA.

Multiple Epiphyseal Dysplasia (MED) is a relatively common skeletal dysplasia associated with mild short stature and early-onset osteoarthritis. Mutations in five genes have been associated with dominant forms of the disease: cartilage oligomeric matrix protein (*COMP*), matrilin-3 (*MATN3*), and the three type IX procollagen genes (*COL9A1*, *COL9A2*, and *COL9A3*). All of these genes are selectively expressed in cartilage and encode structural molecules of the cartilage extracellular matrix. A clinically distinct recessive form of MED results from mutations in a widely expressed gene, Solute Carrier Family 26 Member 2 (*SLC26A2*), which encodes a protein required for post-translational sulfation. In about 15% of MED cases the molecular basis of disease remains unknown, implying further genetic heterogeneity.

To identify additional disease loci, exome sequencing of eight MED families was conducted. This cohort included four sporadic cases hypothesized to represent dominant MED, two recurrent cases thought to represent recessive MED, and two large families with clear dominant inheritance. In three cases, mutations in the known genes were identified. A novel de novo c.949C>T mutation in *COMP* that results in the protein change p.Asp317Asn in the calmodulin-like repeats was identified in one sporadic case. A recurrent c.518G>T mutation resulting in the p.Ala173Asp amino acid change and a novel c.326A>T mutation resulting in the p.Ile109Lys protein change in *MATN3*, both of which target the Von Willebrand Factor-A domain in which all previous *MATN3* MED mutations have been found, were identified in the two large families with dominant MED.

The remaining five cases were negative for mutations in the known genes, but heterozygosity for two novel de novo variants was identified in one case: a c.871G>T missense variant resulting in a p.Val291Phe protein change in *PTH2R*, a G protein-coupled receptor expressed in the endochondral growth plate, and a c.734delT frameshift variant in *PRICKLE1*, a protein involved in the WNT5A noncanonical planar cell polarity pathway. Screening of 32 additional MED cases did not detect mutations in either *PTH2R* or *PRICKLE1*, suggesting that mutations in one of these genes may be a rare cause of MED. Causative variants were not found in the remaining 4 cases, demonstrating further locus heterogeneity in MED.

3097S

Genetic evaluation of inherited arthropathies. *G.SL. Bhavani[1], H. Shah[2], A. Shukla[1], S. Danda[3], K. Gowrishankar[4], S.R. Phadke[5], S. Nampoothri[6], M. Bhat[7], M. Kabra[8], N. Gupta[8], R.D. Puri[9], M.L. Kulkarni[10], A. Gupta[11], S. Guha[12], P. Ranganathan[13], K. Mohanasundaram[14], S.V. Hariharan[15], A.B. Dalal[16], K.M. Girisha[1].* 1) Department of Medical Genetics, Kasturba Medical College, Manipal University, Manipal, India; 2) Pediatric Orthopedics services, Department of Orthopedics, Kasturba Medical College, Manipal University, Manipal, India; 3) Clinical Genetics Unit, Christian Medical College, Vellore, India; 4) Department of Medical Genetics, Childs Trust Medical Resaerch Foundation, Kanchi Kamakoti childs Trust Hospital Chennai, India; 5) Department of Medical Genetics, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India; 6) Department of Pediatric Genetics, Amrita Institute of Medical Sciences and Research Center, AIMS Ponekkara, Cochin, India; 7) Centre for Human Genetics, Bangalore, India; 8) Division of Genetics, Department of Pediatrics, AIIMS, New Delhi, India; 9) Centre of Medical Genetics, Sir Ganga Ram Hospital, New Delhi, India; 10) Department of Pediatrics, JJM Medical College, Davanagere, India; 11) Department of Pediatrics, PGIMER, Chandigarh, India; 12) Vivekananda Institute of Medical Sciences, Kolkata, India; 13) Department of Medical Genetics, Nizam's Institute of Medical Sciences, Hyderabad, India; 14) Department of Rheumatology, Madras Medical College, Chennai, India; 15) Department of Pediatrics, SAT Hospital, Government Medical College, Trivandrum, India; 16) Diagnostics Division, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India.

A small subset of arthropathies is monogenic. We aimed to analyze the mutations underlying all the conditions with arthropathy listed under the different groups in the Nosology and Classification of Genetic Skeletal Disorders (2010). The list includes progressive pseudorheumatoid dysplasia; chronic infantile neurologic cutaneous articular syndrome; sterile multifocal osteomyelitis, periostitis, and pustulosis; familial hyperphosphatemic tumoral calcinosis; infantile systemic hyalinosis; familial digital arthropathy with brachydactyly; multicentric osteolysis, nodulosis and arthropathy; multicentric carpal-tarsal osteolysis with and without nephropathy; stickler syndrome type 1, 2, 3 and type 4. The clinical and molecular spectrum of these syndromes is being studied. We recruited 43 families with progressive pseudorheumatoid dysplasia; 11 families with multicentric osteolysis nodulosis and arthropathy; 2 families with familial hyperphosphatemic tumoral calcinosis; one family each with chronic infantile neurologic cutaneous articular syndrome and infantile systemic hyalinosis. We identified 13 mutations (6 novel: c.49-1G>A, p.S177*, p.C209Mfs*21, p.C247Lfs*31, p.I260Nfs*17, p.Q269Nfs*44 and 5 known: p.C52*, p.C78Y, p.G226V, p.R230Lfs*4, p.C247Lfs*31, p.C337Y) in *WISP3* gene. In *MMP2* gene we identified 6 mutations: 5 are novel (p.S304Pfs*115 and p.N430Tfs*68, p.R101C and p. G410V, p.E231*) and one known (p.R101H). These are the preliminary results of our cohort of patients with inherited arthropathies. We wish to extend this study by collaborating with national and international investigators and hope to explore the genetic bases of Mendelian arthropathies.

3098M

XYLT1 mutations impact on cellular proteoglycan biosynthesis leading to Desbuquois dysplasia type 2. *C. Huber[1], C. Bui[1], B. Tuysuz[2], Y. Alanay[3], C. Bole-Feysot[4], J. Leroy[5], G. Mortier[6], P. Nitschke[7], A. Munnich[1], V. Cormier-Daire[1].* 1) Department of Genetics, INSERM U781, Université Paris Descartes- Sorbonne Paris Cité, Institut Imagine, Hôpital Necker Enfants Malades (AP-HP), Paris, 75015, France; 2) Istanbul University, Cerrahpasa Medical Faculty, Department of Pediatric Genetics, Istanbul, 34098, Turkey; 3) Acibadem University, School of Medicine, Department of Pediatrics, Pediatric Genetics Unit, Istanbul, 34457, Turkey; 4) Plateforme de Génomique, Fondation IMAGINE, Paris, 75015, France; 5) Greenwood Genetic Center, Greenwood, South Carolina 29646, USA; 6) Department of Medical Genetics, Antwerp University Hospital and University of Antwerp, Edegem, 2650, Belgium; 7) Plateforme de Bioinformatique, Université Paris Descartes, Paris, 75015, France.

Desbuquois dysplasia (DBQD) is a severe chondrodysplasia belonging to the multiple dislocation group and characterized by short stature, joint laxity and advanced carpal ossification. Based on the presence of additional hand anomalies, we have previously distinguished DBQD type 1 and identified CANT1 (calcium activated nucleotidase 1) mutations as responsible for DBQD type 1. We report here the identification of 5 distinct homozygous xylosyltransferase 1 (XYLT1) mutations in 7 DBQD type 2 cases from 6 consanguineous families. Among the 5 mutations, 4 were expected to result in loss of function and a drastic reduction of XYLT1 cDNA level was demonstrated in 2 cultured individual fibroblasts. No significant clinical or radiological differences could be found with the remaining 14 DBQD type 2 cases with unknown molecular bases. However, long term follow up of XYLT1 mutated individuals emphasizes the severity of the short stature (< -6 SD) contrasting with obesity, lower limb and foot deformities requiring often repeated surgeries and intellectual disability (5/7). Interestingly, respiratory distress was present at birth in 4/7 cases and spontaneously resolved in the first years of life but thorax narrowness persisted in the eldest children. Since xylosyltransferase 1 (XT-I) catalyzes the very first step in proteoglycan (PG) biosynthesis, we further demonstrated in the two individual fibroblasts a significant reduction of cellular PG content, as well as an important reduction of secreted and large chondroitin sulphate PG production. A recent study published by Mis et al., describe the phenotype of the pug mouse mutant carrying a homozygous missense mutation in Xylt1 and characterized by a disproportionate dwarfism. They observed only 20% Xylt1 activity leading to a reduction of PG processing sufficient to disrupt premature chondrocyte maturation. Finally, our findings of XYLT1 mutations in DBQD type 2 further support a common physiological basis involving PG synthesis in the multiple dislocation group of disorders. This observation sheds light on the key role of the XT-I during the ossification process.

3099T

LOX, among the LOX family proteins, plays a key role in osteoblast differentiation. *Y. Kim, M. Salahuddin.* Dept Biochem, Wonkwang Univ Sch Med, Iksan, Jeonbuk, South Korea.

The lysyl oxidase (LOX) family is an emerging family of amine oxidases that play an essential role in the formation of collagen fibrils in the extracellular matrix. So far, five *LOX* family genes have been identified in humans, encoding LOX and LOX-like proteins (LOXL, LOXL2, LOXL3, and LOXL4). During osteoblastogenesis, LOX regulates the formation of collagen fibrils that function as a three-dimensional scaffold for mineralization. To assess the functional roles of the LOX family members in osteoblastogenesis, we investigated the temporal expression of this gene family as a function of phenotypic development during osteoblastic differentiation of primarily cultured mouse calvaria cells. Among the *LOX* family members, only *LOX* showed prominent expression during osteoblast differentiation. During osteoblast differentiation of the mouse calvaria cells, *LOX* showed the highest level of expression on day 9 in both RT-PCR and Western blot analyses. *COL1A2* that encodes the 2-chain of human collagen type I showed a similar expression pattern to that of LOX, reaching the highest level on day 9. The total amine oxidase activity of the differentiating osteoblast cells also reached the greatest level on day 9. Presence of BAPN, a well-known specific inhibitor of LOX-derived amine oxidase activity, significantly inhibited both mineral nodule formation and expression of osteoblastic marker genes during the differentiation. These findings suggest that the LOX-mediated organization of collagen fibrils in extracellular matrices of differentiating osteoblasts is an important regulator of osteoblastogenesis. Further studies on the functional regulation of LOX in osteoblastogenesis will provide valuable information for better understanding the molecular mechanisms associated with aberrant formation of bone matrix.

3100S

Identification of mutations in patients with osteogenesis imperfecta from Russia. *D.D. Nadyrshina[1], R.I. Khusainova[2], E.K. Khusnutdinova[1,2].* 1) Biological Department, Bashkir State University, Ufa, RE., Russian Federation; 2) Laboratory of human molecular genetics, Institute of Biochemistry and Genetics, Ufa, Russian Federation.

Osteogenesis imperfecta (OI) is a clinically and genetically heterogeneous brittle bone disorder. Whereas dominant OI is mostly due to heterozygous mutations in either *COL1A1* or *COL1A2* encoding type I procollagen, recessive OI is caused by mutations in genes encoding proteins involved in type I procollagen processing or chaperoning. The aim of our study was to identify mutations in *COL1A1, COL1A2, CRTAP, LEPRE1, PPIB* and *SERPINF1* genes in Russian OI patients. We examined 78 patients with OI and 100 healthy controls corresponding by age, gender, ethnicity and place of residence. We sequenced the coding and exon-flanking regions of *COL1A1, COL1A2, CRTAP, LEPRE1, PPIB* and *SERPINF1* genes. We identified five distinct mutations, undescribed before. For the first time previously unreported nonsense mutation c.967G>T (p.Gly323X) in *COL1A1* gene was identified in heterozygous state in two Russian patients, the novel frameshift mutation c.3541insC (p.Gly1181ArgfsX38) in *COL1A1* gene was observed in Yakut patient, the splicing mutation c.1724+4G>A in *LEPRE1* gene was identified in two patients from Tatar population. And novel compound heterozygous mutations (c.913C>G (p.Leu305Val) of *SERPINF1* gene and c.641T>C (p.Val214Ala) of *CRTAP* gene) was observed in three patients from Bashkir population. We also detected three previously described nonsense mutations in five Russian patients: c.1081C>T (p.Arg361X), c.1243C>T (p.Arg415X) and c.2869C>T (p.Gln957X), two frameshift mutations in two Tatar patients: c.579delT (p.Gly194ValfsX71) and c.2444delG (p.Gly815AlafsX293) and one splicing mutation in one Yakut patient: c.4005+1G>T. One frameshift mutation (c.579delT (p.Gly194ValfsX71), three nonsense mutations (c.967G>T (p.Gly323X), c.1081C>T (p.Arg361X), c.2869C>T (p.Gln957X)) and one splice mutation (c.4005+1G>T) were identified in patients with OI type I. Two mutations (c.2444delG (p.Gly815AlafsX293) and c.3540_3541insC (p.Gly1181AlafsX38)) occurred in sporadic cases of OI type I, whereas c.1243C>T (p.Arg415X) mutation - of OI type III. The mutation c.1724+4G>A were identified in patients with OI type VIII. In conclusion, the present study revealed 8 mutations in *COL1A1* gene, 1 mutation in *SERPINF1* gene, 1 mutation in *LEPRE1*, 1 mutation in *CRTAP* gene, five of them was not observed before and no mutations in *COL1A2* and *PPIB* genes in Russian patients with OI. Future research will focus on other genes responsible for OI development in Russian patients.

3101M

Two Distinct Mutations in *IFITM5* Causing Different Forms of Osteogenesis Imperfecta Using Reciprocal Mechanisms. *A. Reich[1], AM. Barnes[1], CR. Farber[2, 3], P. Becerra[4], F. Rauch[5], WA. Cabral[1], A. Bae[1], FH. Glorieux[5], TL. Clemens[6], JC. Marini[1].* 1) NICHD, NIH, Bethesda, MD; 2) Center for Public Health Genomics, Univ Virginia, Charlottesville, VA; 3) Depts Public Health Sciences, Biochemistry & Mol Genet, UVA, Charlottesville, VA; 4) Section on Protein Structure and Function, LRCMB, NEI, NIH, Bethesda, MD; 5) Shriners Hosp for Children and McGill Univ, Montreal, Canada; 6) Dept Orthopaedic Surgery, Johns Hopkins Sch Med, Baltimore, MD.

Osteogenesis imperfecta (OI) type V is caused by a recurrent dominant mutation (c.-14C>T) in *IFITM5*, which encodes BRIL, a transmembrane ifitm-like protein most strongly expressed in osteoblasts, whose expression coordinates with mineralization. Patients with type V OI have distinctive clinical manifestations with overactive bone mineralization and include mesh-like lamellation on bone histology. Recessive null mutations in *SERPINF1* cause OI type VI, which impairs mineralization. Type VI OI patients have absence of serum PEDF, elevated alkaline phosphatase (ALPL) as children, and bone histology with broad unmineralized osteoid and fish-scale pattern. We identified a 25-year-old woman with severe OI, who had multiple features of type VI OI. Her dermal fibroblasts (FB) and cultured osteoblasts (OB) displayed minimal secretion of PEDF, but her serum PEDF was normal. *SERPINF1* sequences were normal despite bone histomorphometry typical of type VI OI, and elevated childhood serum ALPL. Exome sequencing yielded a de novo mutation in IFITM5 in one allele of the proband, causing a p.S40L substitution in the BRIL intracellular domain, despite her not having type V OI characteristics. In OB with the *IFITM5* p.S40L mutation, *IFITM5* expression was normal, as was BRIL protein level on western blot and in permeabilized OB by microscopy. Notably, *SERPINF1* expression was decreased in p.S40L OB, and PEDF was barely detectable in conditioned media of S40L cells. OB with the S40L mutation also had decreased expression of ALPL and osteocalcin (OCN), as seen in primary PEDF defects. Conversely, OB from type V OI, with 5 residues added to the N-terminus of BRIL, have increased *SERPINF1* expression and PEDF secretion during differentiation, and increased ALPL and OCN expression. The *IFITM5* mutations also have opposite effects on OB mineralization in culture - OB with the 5'-terminal mutation have increased mineralization during differentiation, while mineralization was decreased in S40L OB. OB from both mutations share a collagen related defect, with decreased expression and secretion of COL1A1. Together, these data show that 2 distinct mutations in *IFITM5* generated different forms of OI with distinctive phenotypes using reciprocal mechanisms; the type V OI and p.S40L BRIL are gain- and loss-of-function mutations, respectively. Furthermore, BRIL and PEDF have a relationship that connects the genes for types V and VI OI and their roles in bone mineralization.

3102T

Identification and Genetic Characterization of a Mysterious Crippling Disorder of Arai Village, J&K, India. *S. Sharma[1], A. Mahajan[2], P. Kumar[3], M.K. Dhar[2], S. Razdan[4], K. Thangaraj[5], C. Wise[6], S. Ikegawa[7], K.K. Pandita[4], E. Rai[1].* 1) School of Biotechnology, Shri Mata Vaishno Devi University, Katra, J&, India; 2) School of Biotechnology, University of Jammu, Jammu, J&, India; 3) Human Genetic Research cum Counselling Centre, University of Jammu, Jammu, J&, India; 4) Acharya Shri Chander College of Medical Sciences, Jammu, J&, India; 5) Centre for Cellular and Molecular Biology, Hyderabad, A.P, India; 6) Texas Scottish Rite Hospital for Children, Dallas, Texas, USA; 7) Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences,Tokyo, Japan.

Clinical features of rare disorders are very often poorly understood due to their low prevalence. Quite a few times, these rare disorders remain uncharacterized, or patients are misdiagnosed and get poor medical attention. A mysterious skeletal disorder that remained unidentified for decades and rendered many people physically challenged and disabled for life has been reported in an isolated remote village 'Arai' of Poonch district of Jammu and Kashmir state, India. The typical phenotypic characteristics shared among the affected included: pain, kyphoscoliosis, fatigability, muscular weakness, progressive restriction of joint movement, stiffness and swelling at several joints, including the proximal and distal interphalangeal joints. This village is located deep in mountains, and the population residing in the region is highly consanguineous. Combining multiple lines of evidence (familial histories and genetic data, clinical features, radiological and biochemical data and phenotypic features), we identified the disorder as a recessive hereditary skeletal disease "Progressive Pseudorheumatoid Arthropathy of Childhood" (PPAC) also known as "Spondyloepiphyseal Dysplasia Tarda with Progressive Arthropathy" (SEDT-PA). Typically a very rare disorder [for example, with an estimated frequency one per million in UK], in our survey, of this village with a total population of approximately 5000 individuals, we observed about 70 affected people. Mutations in one reported gene "** WISP3**" are known to be responsible for causing of PPAC. Therefore, we sequenced coding and UTR regions of this gene of two multigenerational extended families from the village. We found a novel splice site mutation at position c.643+1G>A that perfectly segregated with the disease in one pedigree and is most plausibly the cause of disease in this family. Interestingly, the affected in second family also shared identical clinical characteristics and disorder showed similar recessive inheritance. However, *WISP3* was negative for co-segregating mutations in the second family suggesting the existence of a different candidate gene for SEDT-PA. Therefore, we are using Next Generation (exome) Sequencing as a promising option and pursuing it. This genetic characterization will aid in molecular diagnosis and genetic counseling, critically required to curb the disorder and to prevent its appearance in future generations in the population.

3103S

Novel mutation in *RUNX2/CBF-alpha-1* with alanine tract expansion from Japanese cleidocranial dysplasia patient. *A. Shibata[1, 2], J. Machida[1, 3], S. Yamaguchi[1, 4], M. Kimura[1, 5], T. Tatematu[1, 2], H. Miyachi[1], Y. Higashi[2], A. Nakayama[6], K. Shimozato[1], Y. Tokita[1, 2].* 1) Oral and Maxillofacial Surgery, school of Dentistry, Aichi-Gakuin University, Nagoya, Japan; 2) Department of Perinatology, Institute of Developmental Research, Aichi Human Service Center, Kasugai, Japan; 3) Department of Oral and Maxillofacial Surgery, Toyota Memorial Hospital, Toyota, Japan; 4) Department of Dentistry and Oral surgery, Aichi Children's Health and Medical Center, Obu, Japan; 5) Department of Dentistry Oral and Maxillofacial Surgery, Ogaki Municipal Hospital, Ogaki, Japan; 6) Department of Embryology, Institute of Developmental Research, Aichi Human Service Center, Kasugai, Japan.

Cleidocranial dysplasia (CCD [MIM 119600]) is an autosomal-dominant skeletal disease with high penetrance and variable expressivity caused by heterozygous mutations in the osteoblast-specific transcription factor known as runt related transcription factor 2 or core-binding factor subunit alpha-1 (*RUNX2/CBF-alpha-1*). Molecular analysis of the *RUNX2* gene in a 29-year-old woman with a diagnosis of CCD demonstrated a heterozygous in-frame 9-bp duplication between codon 181 and 189 (c.181_189dupGGCGGCTGC), corresponding to 3-alanine expansion in the polyalnine tract (17Ala>20Ala). To date, only one case has been reported with alanine tract expansion of Runx2 (27Ala) as a cause of CCD. To investigate the role of this short alanine expansion in the CCD pathologies, we investigated the transcription activity of the 20Ala and 27Ala tract expansion variants of RUNX2 with the p6OSE2-luciferase reporter gene, which contains six tandem copies of osteoblast-specific element-2, the RUNX2-binding promoter sequence. Then we studied the protein aggregates formation in the gene-transfected cells, since it has been demonstrated that alanine tract stretch of RUNX2 with 27Ala variants induces cytoplasmic protein aggregation. Although we detected minor effect of the alanine stretches on the transactivation function at 24 hours after gene transfection, we observed large amorphous aggregates of the 20Ala and 27Ala variants of RUNX2 in the cytoplasm of transfected cultured COS7 cells at 48 hours after gene transfection. These findings indicate that the 20Ala allele, as well as 27Ala variant, is a biologically relevant polymorphism that influences osteoblast differentiation. Furthermore, our results suggested that the lengths of alanine expansion correlated with aggregation severity, while there is no relation between the alanine expansion lengths and clinical severity.

3104M

Rapid Turnover Skeletal Disease Caused By A Multiple-Exon Duplication of *TNFRSF11A* Encoding RANK. *S. Mumm[1], F. Collins[2,3], S. Duan[1], M. Huskey[1], D. Sillence[3], M. Whyte[4].* 1) Bone/Mineral Div, Box 8301, Washington Univ Sch Med, St Louis, MO; 2) The Children's Hospital at Westmead, Australia; 3) Sydney University Medical School, Australia; 4) Shriners Hospital for Children, St. Louis MO.

Mendelian disorders of constitutive RANK activation are caused by in-frame 12, 15, 18, or 27-bp duplications in exon 1 of *TNFRSF11A* encoding the signal peptide of RANK. The similar rapid bone turnover disorder juvenile Paget's disease (JPD) is due to deactivating mutations in *TNFRSF11B* encoding OPG. We report a rapid turnover skeletal disease in a mother and fetus from a novel RANK defect. The 38-year-old proposita is the child of nonconsanguineous parents. Family history was negative for fractures or deafness. She wore hearing aids for deafness at age 15 months. Brittle teeth were noted at age 3 years. Secondary tooth loss began at age 9 years. At age 10 years, she was non-dysmorphic and without bone pain or fractures. Bone scintigraphy showed increased uptake in the long bones, and CT revealed hypoplastic ossicles and semicircular canals and no cochleas. A skeletal survey revealed thickened cortices of tubular bones, but a normal skull. Biopsy of a tibia and ileum showed unremarkable lamellar bone, slightly increased vascularity, numerous osteoblasts, and normal marrow. Serum alkaline and acid phosphatase were elevated. An atypical form of JPD was diagnosed. At age 14 years, she had normal stature, slight knock-knee deformity, and fractured major limb bones. She received calcitonin and then a bisphosphonate. At age 16 years, biopsy of the maxilla and jaw showed areas resembling cementifying fibroma, irregular woven bone, and marked osteoblastic and osteoclastic activity in a highly cellular fibrovascular stroma resembling osteolytic Paget's disease. Teeth were eroded and replaced by osteocementum and fibrous tissue. She was edentulous by age 27 years. At age 30 years, she had a prominent forehead and nasal bridge, deep set eyes, and anterior tibial bowing. Then, ultrasound of her first pregnancy at 18 weeks showed abnormal fetal bones with angulated femurs. Sequencing of coding exons and adjacent mRNA splice sites of the RANK and OPG genes of the mother and abortus was negative. However, in both, microarray-based copy number analysis showed amplification of exons 4-9 of the RANK gene. Confirmatory qPCR showed 3 copies of exons 4-9 in the mother, and >3 in the fetus. An in-frame tandem duplication of exons 4-9 creates a RANK fusion protein of one extracellular RANKL-binding domain combined with double intracellular activation domains. Thus, we have identified a new genetic basis within the family of disorders featuring constitutive RANK activation.

3105T

Asynchronous remodelling is a driver of failed regeneration in Duchenne Muscular Dystrophy. *S. Dadgar, Z. Wang, H. Johnston, K. Nagaraju, A. Hill, T. Partridge, E. Hoffman.* Children's National Medical Center- George Washington University, Washington, DC.

The molecular mechanisms underlying the progressive fibrosis and failure of regeneration in Duchenne muscular dystrophy have remained elusive. Analysis of 166 patient muscle biopsy mRNA profiles (test and validation sets) identified a 56 member network centered on TGFβ associated with severe pathology (fibrosis and failed regeneration). Superimposing this pathology-related network on a 27 time point murine normal muscle regeneration series showed stage-specific regulation of each network member during normal staged regeneration, but at distinct time points (temporal parsing into subnetworks). From this, we developed an asynchronous remodeling model for fibrosis and failed regeneration. This model predicted that the normal 2 week regeneration cycle of muscle is disrupted through inappropriate crosstalk between neighboring cells in different time points of the 2 week-cycle. To test this hypothesis, we developed an experimental model of focal asynchronous bouts of muscle regeneration, with laser capture microscopy of marked tissue regions (first bout, second bout, and in between [crosstalk] areas). mRNA profiling and immunohistochemical validations showed that the crosstalk areas in between staged bouts of regeneration became inappropriately fixed in the developmental time point by which the initial bouts were separated. This led to a chronic inflammatory state and mitochondrial insufficiency in bouts separated by 4 days, and a chronic profibrotic state in bouts separated by 10 days. Molecular networks associated with these localized areas of pro-inflammatory states were suppressed by treatment with glucocorticoids and VBP15. In summary, synchronous regeneration of muscle is successful, whereas neighboring asynchronous bouts create inappropriate crosstalk between cells in different stages of the regenerative process that results in failed regeneration and the pro-fibrotic and pro-inflammatory state. Our data provide a unifying model for failed regeneration and pathological fibrosis in muscular dystrophies that is likely generalizable to chronic inflammatory states in other regenerative tissues. This model also provides a novel mechanism of action for glucocorticoids in many of these disorders: they serve to re-synchronize remodeling, much as diurnal cortisol fluctuations do in most animals.

3106S

Novel mutations in *GNAI3* associated with Auriculocondylar Syndrome strengthen a common dominant negative effect. *V.L.R. Tavares[1], C.T. Gordon[2], R.M. Zechi-Ceide[3], N.M. Kokitsu-Nakata[3], N. Voisin[2], T.Y. Tan[4], A.A. Heggie[5], S. Vendramini-Pittoli[3], E.J. Propst[6], B.C. Papsin[6], T.T. Torres[7], H. Buermans[8], L.P. Capelo[1,9], J.T. den Dunnen[8], M.L. Guion-Almeida[3], S. Lyonnet[2,10], J. Amiel[2,10], M.R. Passos-Bueno[1].* 1) Centro de Pesquisas do Genoma Humano e Células Tronco, Departamento de Genetica e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brasil; 2) INSERM U1163, Université Paris Descartes-Sorbonne Paris Cité, Institut Imagine, Paris, France; 3) Department of Clinical Genetics, Hospital of Rehabilitation of Craniofacial Anomalies, University of São Paulo (HRCA-USP), Bauru, SP, Brasil; 4) Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Royal Children's Hospital, and Department of Paediatrics, University of Melbourne, Melbourne, Australia; 5) Department of Plastic and Maxillofacial Surgery, Royal Children's Hospital, Melbourne, Australia; 6) Department of Otolaryngology - Head & Neck Surgery, The Hospital for Sick Children, Toronto, Canada; 7) Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brasil; 8) Leiden Genome Technology Center, Leiden University Medical Center, Leiden, Nederland; 9) Instituto de Ciência e Tecnologia, Universidade Federal de São Paulo, São José dos Campos, Brasil; 10) Département de Génétique, Hôpital Necker-Enfants Malades AP-HP, Paris, France.

Auriculocondylar syndrome is a rare craniofacial disorder comprising core features of micrognathia, condyle dysplasia and question mark ear. Causative variants have been identified in *PLCB4*, *GNAI3* and *EDN1*, which are predicted to function within the EDN1-EDNRA pathway during early pharyngeal arch patterning. To date, two *GNAI3* variants in three families have been reported. Here we report three novel *GNAI3* variants, one segregating with affected members in a family previously linked to 1p21.1-q23.3 and two *de novo* variants in simplex cases. Two variants occur in known functional motifs, the G1 and G4 boxes, and the third variant is one amino acid outside of the G1 box. Structural modeling shows that all five altered GNAI3 residues identified to date cluster in a region involved in GDP/GTP binding. We hypothesize that all *GNAI3* variants lead to dominant negative effects.

3107M

Mouse Model with Mutant Type I Collagen C-propeptide Cleavage Site has Brittle Bones and Increased Osteoblast Mineralization. *A.M. Barnes[1], J.E. Perosky[2], M.H. Rajpar[1], K.M. Kozloff[2], J.C. Marini[1].* 1) BEMB, NICHD/NIH, Bethesda, MD; 2) Orthopedic Surgery and Biomedical Engineering, University of Michigan, Ann Arbor, MI.

Classical dominant osteogenesis imperfecta (OI) is caused by mutations in type I collagen. Mutations in the C-propeptide cleavage site of both *COL1A1* and *COL1A2* were shown to cause high bone mass OI, characterized by bone hypermineralization and normal to increased DXA Z-scores. We generated a mouse with a heterozygous C-propeptide cleavage site defect (high bone mass, HBM) in which the Ala-Asp residues of the proα1(I) cleavage site were mutated to Thr-Asn to prevent cleavage by BMP1. We utilized this mouse to investigate the role of proα1(I) C-propeptide processing in bone formation. Collagen from HBM calvarial osteoblasts (OB) had normal biochemistry and the collagen was incorporated normally into trimers. Conditioned media from HBM fibroblasts showed decreased processing of the C-propeptide. HBM OB deposited only about 50% of WT matrix. An *in vitro* mineralization assay showed that HBM calvarial OB had a 15% increase in mineralization ($p=0.006$), ± BMP2. Dermal fibril diameters were smaller and more homogeneous in HBM than WT, with a loss of large fibrils. HBM mice are growth deficient, remaining smaller than WT from 4-16 weeks. At 2 months, male HBM mice are significantly smaller in weight (77%) and length (92%) and have shorter femurs (92%). Femoral areal BMD in HBM mice is decreased 25% ($p<0.001$); vertebral BMD is normal. All 2 month HBM mice have pelvic deformities, and 40% have kyphosis. On μCT, HBM femora have a thinner cortex with decreased cortical area; the distal femoral metaphysis is notably thinner. Cortical TMD is modestly reduced, while trabecular TMD, trabecular bone volume and trabecular number and thickness are decreased. Four point bending revealed an extremely brittle phenotype; post-yield displacement is only ~10% of WT ($p<0.001$). HBM femoral stiffness, yield load, and ultimate load are also significantly decreased. The bone phenotype of HBM mice is similar to *Bmp1*$^{-/-}$/*Tll1*$^{-/-}$ mice (Muir, et al, *HMG*, 2014, 23:3085-3101), in which cleavage of multiple procollagens as well as bone-related factors is defective. *Bmp1*$^{-/-}$/*Tll1*$^{-/-}$ mice have small size, thin cortices, reduced failure energy and maximum load and a dramatic decrease in post-yield displacement. The HBM mouse demonstrates that the essential elements of the broader BMP1 enzyme deficiency are reproduced by a substrate defect in type I C-propeptide cleavage. These data show the importance of the proα1(I) C-propeptide to both collagenous and mineral properties of bone.

3108T

Spliceosomopathies: an emerging link between the spliceosome and disorders of craniofacial and skeletal development. *D.C. Lynch[1], A.M. Innes[1, 2], R.L. Lamont[1, 2, 3], J.S. Parboosingh[1, 2, 3], F.P. Bernier[1, 2].* 1) Medical Genetics, University of Calgary, Calgary, Canada; 2) Alberta Children's Hospital Research Institute, Calgary, Canada; 3) Genetic Laboratory Services, Alberta Health Services, Calgary, Canada.

Recently a group of craniofacial and skeletal disorders caused by dominant mutations in spliceosomal genes has emerged. The first of these to be identified is mandibulofacial dysostosis type Guion-Almeida (MFDGA), caused by haploinsufficiency of the U5 snRNP component EFTUD2. We recently identified haploinsufficiency of SF3B4 as the cause of Nager syndrome, an acrofacial dysostosis (AFD). Mutation of SF3B4 was subsequently reported in AFD Rodriquez type (AFDR). SF3B4 is part of the U2 snRNP of the major spliceosome. Herein we also present heterozygous mutations in a highly conserved regulatory exon of SNRPB as the cause of cerebrocosto-mandibular syndrome (CCMS). Our experimental data show that these mutations disturb highly conserved exonic splicing silencer sequences crucial to the regulation of the gene's expression. SNRPB encodes a component of the Sm complex of the spliceosome, which is present in all snRNPs. Pierre Robin sequence and ear anomalies are present in all four syndromes. However, malar hypoplasia is present only in Nager, MFDGA and AFDR. Microcephaly is typically present only in MFDGA. Other disease-specific features include posterior rib gaps in CCMS, tracheoesophageal fistula in MFDGA, pre-axial anomalies in Nager and severe limb defects in AFDR. Together, these findings invite questions on the sensitivity of craniofacial and in particular pharyngeal arch development to spliceosomal defects, and on the role of the spliceosome in the regulation of development. We propose a comprehensive review of the clinical and molecular features of this emerging group of disorders with a particular emphasis on the role of SF3B4, EFTUD2 and SNRPB in disorders of craniofacial and skeletal development.

3109S

Novel COL1A2 Gene Mutation in Czech Osteogenesis Imperfecta Patient: Case report. *S. Mazurova[1], L. Hruskova[1], I. Marik[2], P. Martasek[1], I. Mazura[1].* 1) Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University in Prague, Ke Karlovu 2, 12100 Prague 2, Czech Republic; 2) Ambulant Centre for Defects of Locomotor Apparatus, Olšanská 7, 13000 Prague 3.

Osteogenesis imperfecta /OI/ is an inherited disorder of connective tissue. About 90% of affected patients with OI type I-IV show causative mutation in one of the two genes, COL1A1 and COL1A2, coding 2 alpha 1 chains and 1 alpha 2 chain of collagen type I protein. Clinical effect of these mutations results in mild to lethal phenotype regard on type of osteogenesis imperfecta and position of the nucleotide change. We continue with analyses of the Czech patients with OI type I-IV and COL1A1 and COL1A2 genes some of which have already been described earlier. This paper is focused on one mutation of COL1A2 gene and correlation of genotype-phenotype parameters in Czech osteogenesis imperfecta patient. Mutation observed in the COL1A2 gene is localized in exon 40 and we describe clinical symptoms of patient correlating with Gly814Trp substitution. We used molecular genetic analysis of genomic DNA extracted from peripheral blood leucocyte. DNA quality and quantity were confirmed by spectrophotometry and horizontal electrophoresis and subsequently target amplification and two-directional sequencing was performed. We demonstrate a case of 25 year old woman affected by OI type III which is clinically followed from 3 years and 4 months of life. Last clinical examination was performed in 17 years of life. This woman was born from second gravidity of unaffected parents and perinatal fractures of both low extremities were described. This work was supported by following grants: grant no.: SVV-2014-260034 supported by Charles University Prague, grant no.: UNCE 204011 supported by Ministry of Education, Youth and Sport, Czech Republic, grant no.: PRVOUK P24/1LF/3 supported by Ministry of Education, Youth and Sport, Czech Republic.

3110M

Expending the phenotypic spectrum of *PDE4D* and *PRKAR1A* mutations : from acrodysostosis to acroscyphodysplasia. *C. Michot[1], C. Le Goff[1], G. Baujat[1], K. Piquant[1], E. Blair[2], O. Boute[3], B. Gilbert-Dussardier[4], A. Goldenberg[5], A. Henderson[6], B. Isidor[7], H. Kayserili[8], E. Kinning[9], P.O. Simsek-Kiper[10], Q. Quelin[11], R. Savarirayan[12], M. Simon[13], M. Splitt[6], J. Verhagen[13], A. Verloes[14], M. Le Merrer[1], A. Munnich[1], V. Cormier-Daire[1].* 1) Department of Genetics, Necker Hospital, INSERM U781, Université Paris Descartes, Institut IMAGINE, Paris, France; 2) Department of Clinical Genetics, Churchill Hospital, Oxford OX3 7LE, UK; 3) Department of Clinical Genetics, C.H.R.U. de Lille, Hôpital Jeanne de Flandre, Lille, France; 4) Department of Genetics, C.H.U. La Milétrie, Poitiers, France; 5) Department of Clinical Genetics, C.H.U. de Rouen, Rouen, France; 6) Northern Genetics Service, The Institute of Genetic Medicine, Newcastle upon Tyne, UK; 7) Department of Medical Genetics, C.H.U. de Nantes, Nantes, France; 8) Department of Medical Genetics, Istanbul University, Istanbul Faculty of Medicine, Istanbul, Turkey; 9) The Ferguson-Smith Centre for Clinical Genetics Royal Hospital for Sick Children, Glasgow, UK; 10) Pediatric Genetics Unit, Department of Pediatrics, Hacettepe University Faculty of Medicine, Ankara, Turkey; 11) Department of Medical Genetics, C.H.U. de Nantes, Rennes, France; 12) Royal Children's Hospital Genetics Clinic, Royal Children's Hospital, Victoria, Australia; 13) Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands; 14) Department of Medical Genetics, INSERM U676, Hôpital Robert Debré, Paris, France.

Acrodysostosis (MIM101800) is a dominantly inherited condition associating 1) a skeletal dysplasia characterized by short stature, facial dysostosis and severe brachydactyly with cone-shaped epiphyses 2) resistance to mutiple hormones 3) moderate to mild intellectual disability. Acroscyphodysplasia (MIM250215) is rare and characterized by lower femoral and upper tibial epiphyses embedded in cup-shaped, large metaphyses, associated to severe growth retardation and brachydactyly. We and others have identified *PDE4D* or *PRKAR1A* mutations in acrodysostosis cases. More recently, *PDE4D* mutation has been reported in one case of acroscyphodysplasia. The aim of our study was to review the clinical and molecular findings of 27 acrodysostosis and 5 acroscyphodysplasia cases to further establish genotype-phenotype correlation. Among the 27 unrelated acrodysostosis cases, we identified heterozygous de novo *PRKAR1A* mutations in 9 and heterozygous de novo *PDE4D* mutations in 10 cases. Neither *PDE4D* nor *PRKAR1A* mutations were found in 8 patients. These 8 patients presented with features highly evocative of acrodysostosis (short hands and cone-shaped epiphyses), but none had the canonical facial features and 2 presented with a less severe and atypical brachydactyly (not affecting the five digits). In the 5 cases of acroscyphodysplasia, we identified *PDE4D* mutations in 2. For 3 negative cases, the review of medical records revealed early and severe infection, which has already been described in some cases of acroscyphodysplasia. Splitting our series of acrodysostosis cases based on the disease causing gene confirmed genotype-phenotype correlations. Hormone resistance was consistently observed in the patients carrying *PRKAR1A* mutations while no hormone resistance was observed in at least 6 patients with *PDE4D* mutations. Moreover, all patients with *PDE4D* mutations shared characteristic facial features (midface hypoplasia with the canonical nasal hypoplasia) and some degree of intellectual disability. Finally, our findings of *PDE4D* mutations in 2 cases of acroscyphodysplasia support also that *PDE4D* is responsible for a severe skeletal dysplasia, with major epiphyseo-metaphyseal changes.

3111T

Transcriptional Dysregulation Associated with Somatic Neurofibromin Deficiency in Tibial Pseudoarthrosis with Neurofibromatosis Type 1.

J.J. Rios[1], N. Paria[1], T-J. Cho[2], IH. Choi[2], N. Kamiya[1], K. Kayembe[3], R. Mao[4], RL. Margraf[4], G. Obermosser[3], I. Oxendine[1], DW. Sant[4], MH. Song[6], DA. Stevenson[5], DH. Viskochil[5], CA. Wise[1], HKW. Kim[1]. 1) Research, Texas Scottish Rite Hospital for Children, Dallas, TX; 2) Division of Pediatric Orthopaedics, Seoul National University Children's Hospital, Seoul, Republic of Korea; 3) Baylor Institute for Immunology Research, Dallas, Texas; 4) ARUP Laboratories, Salt Lake City, Utah; 5) Department of Pediatrics, University of Utah, Salt Lake City, Utah; 6) Department of Orthopaedic Surgery, Jeju National University Hospital, Jeju, Republic of Korea.

NF1 haploinsufficiency causes neurofibromatosis type 1 (NF1) and results in constitutive Ras activation, predisposing individuals to neurofibroma and malignant peripheral nerve sheath tumor (MPNST) formation. Children with NF1 are also predisposed to tibial dysplasia with high risk of fracture, pseudoarthrosis and chronic nonunion. Previous studies described somatic loss-of-heterozygosity (LOH) at the NF1 locus in tibial pseudoarthrosis tissue. This study sought to confirm and expand this observation by characterizing the somatic mutation spectrum and transcriptional dysregulation in pseudoarthroses from 16 individuals with NF1. Genomic microarray analysis and whole-exome sequencing identified somatic variants in the NF1 gene in 12 samples, and unlike neurofibromas and MPNSTs, no other recurring somatic variants were identified. Somatic NF1 variants included a single large deletion, six individuals with sequence variants and five individuals with LOH encompassing the entire long-arm of chromosome 17. We did not identify somatic variants for four individuals. Whole-transcriptome sequencing (RNA-seq) was performed using cells cultured from pseudoarthrosis tissue from five individuals, and somatic variants were confirmed in the mosaic cell cultures. RNAseq analysis identified significant upregulation of a tumor-promoting transcriptional pattern in pseudoarthrosis samples, as compared to iliac crest cells from two affected individuals and control samples from three individuals without NF1. Upregulated genes included EGFR (2.5-fold, p=4.8e-4), KITLG/SCF (3.4-fold, p=1.9e-4) and EREG (55.7-fold, p=2.49e-11). Because NF1 haploinsufficiency results in activation of the ERK signaling pathway, we investigated whether pseudoarthrosis cells with neurofibromin deficiency had higher levels of ERK activation compared to haploinsufficient iliac crest cells. Flow cytometry identified no difference in the levels of ERK activation, suggesting upregulation is associated with neurofibromin deficiency rather than quantitative differences in pathway activation. Despite the transcriptional dysregulation, NF1 pseudoarthroses lacked other obvious oncogenic variants, and tumor formation from pseudoarthrosis tissue was not observed nor previously reported in the literature. Results from this study suggest receptor tyrosine kinase inhibitors should be investigated as a non-surgical treatment for tibial pseudoarthrosis to promote bone union in individuals with NF1.

3112S

Compound heterozygosity for a frameshift mutation and an upstream deletion that reduces expression of *SERPINH1* in siblings with a moderate form of osteogenesis imperfecta.

U. Schwarze[1], T. Cundy[2], Y.J. Liu[1], P.L. Hofman[3], J. Schleit[1], M.L. Murray[1,4], P.H. Byers[1,4]. 1) Department of Pathology, University of Washington, Seattle, WA; 2) Department of Medicine, Faculty of Medical & Health Sciences, University of Auckland, New Zealand; 3) Liggins Institute, Faculty of Medical & Health Sciences, University of Auckland, New Zealand; 4) Department of Medicine (Medical Genetics), University of Washington, Seattle, WA.

Osteogenesis imperfecta (OI) is a heritable disorder characterized primarily by bone fragility and increased fracture risk, ranging in severity from mild to perinatal lethal. Currently mutations in 17 genes have been implicated in this disorder; more than 90% of cases of OI are caused by heterozygous mutations in *COL1A1* or *COL1A2*. Most of the remainder are recessively inherited and result from mutations in genes that encode proteins involved in the synthesis of type I collagen. *SERPINH1* encodes the collagen chaperone HSP47 that binds to arginine rich sequences in the type I procollagen trimers and provides the final steps in the folding and stabilization of the triple helical domain. Loss of both alleles in mice results in early embryonic lethality. Homozygous missense mutations in dachshunds and in one child resulted in a moderately severe form of OI with continuing fracture in the face of bisphosphonate treatment in the child. In both instances the identified missense mutations resulted in substitution of different interacting leucine residues by proline. We identified an additional family with non-consanguineous unaffected parents who have two children with moderate short stature, low bone density, fractures, and no evidence of dentinogenesis imperfecta. Both children are heterozygous for two mutations (allele 1; allele 2); one of each was derived from each parent. Initial study identified only a frameshift mutation in the last exon that did not lead to nonsense-mediated mRNA decay (c.1233dupT, p.Asp412*). Analysis of the expression of the two alleles in cultured cells indicated that the other allele was expressed, but was only about half as abundant as the frameshift containing allele. High density CGH identified a 5.3kb deletion upstream from the translation start site that removed a region of DNase sensitivity identified in the ENCODE data set. This allele was inherited from the father, and the mRNA in his cells was also expressed at a low level, which confirmed that this domain has a regulatory function for *SERPINH1*. This class of mutation is currently rarely found in individuals with genetically determined disorders and would not be detected by whole exome sequence analysis. While whole genome sequence analysis could identify it, targeted gene studies may be the more efficient way of identification, especially in families such as this one in which only one mutant allele is found, and appropriate cells are available to measure expression levels.

3113M

Rare TBX6 null mutations in congenital scoliosis. *N. Wu[1], Z. Wu[1], X. Ming[2], J. Xiao[2], X. Chen[4], J. Liu[1], S. Liu[1], Y. Ming[5], L. Jin[2,3], X. Zhang[6], G. Qiu[1], F. Zhang[2,3].* 1) Department of Orthopedic Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, Beijing, China; 2) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China; 3) Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China; 4) Beijing Municipal Key Laboratory of Child Development and Nutriomics, Capital Institute of Pediatrics, Beijing, China; 5) PET-CT Center, Cancer Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China; 6) McKusick-Zhang Center for Genetic Medicine and State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China.

Introduction: Congenital scoliosis (CS) is a common disorder with an estimated incidence of around 0.5-1 in 1000 live births. Genetic mutations have been implicated in CS pathogenesis, but the genetic basis of sporadic CS remains unclear. Human 16p11.2 deletions are rare but recurrent mutations. The CS conditions have been recently observed in a small portion of 16p11.2 deletion carriers, suggesting a potential involvement of 16p11.2 in CS. Methods: Subject Information. We enrolled 161 unrelated sporadic CS patients of Han Chinese from PUMCH between October 2010 and November 2012. Their CS conditions were confirmed by radiological imaging. The patients with known syndromes were excluded. A total of 166 unrelated healthy participants were used as population controls. High-density Oligonucleotide Comparative Genomic Hybridization Microarray. A high-density genome-wide copy number variation (CNV) analysis was performed in 20 trios. The genomic DNA was extracted from peripheral blood leukocytes. Quantitative Polymerase Chain Reaction (qPCR) Analysis. qPCR analysis was conducted to screen for 16p11.2/TBX6 deletion in additional 141 sporadic CS patients. Any 16p11.2 deletion candidate suggested by qPCR was further confirmed by the CGH microarrays. DNA Sequencing of the TBX6 gene. The entire TBX6 gene and its approximately 1-kb upstream region were amplified using long-range PCR in both CS patients and healthy controls. The PCR products were analyzed by Sanger DNA sequencing. Statistical analysis. The Fisher's exact test (two-sided) was used to investigate the potential significant difference in TBX6 mutation/variant frequencies between the CS patients and controls. The differences with $P<0.05$ were considered as statistically significant. Results: We identified recurrent deletions in proximal 16p11.2 in two patients (2/20). Furthermore, both deletions were de novo. Notably, ten more sporadic CS patients were found to have this recurrent 16p11.2 deletion. However, we did not observe any 16p11.2 deletion in 166 Han Chinese controls (12/161 vs. 0/166; $P<0.0002$). The DNA sequencing of the TBX6 gene revealed additional one nonsense and four frameshift mutations of TBX6 in these group of CS patients. Conclusions: Up to our knowledge, we reported the largest genetic study of sporadic CS, suggesting that the TBX6 null mutation is responsible for CS (approximately 10% of sporadic CS). Our findings will facilitate molecular diagnostics of CS.

3114T

Alteration of conserved alternative splicing in AMELX causes enamel defects. *J. Kim[1], K.-J. Kim[1], E.-J. Lee[1], K.-E. Lee[1], C.-Y. Yun[2], E.-S. Cho[2].* 1) Pediatric Dentistry, Seoul National University School of Dentistry, Seoul, South Korea; 2) Cluster for Craniofacial Development and Regeneration Research, Institute of Oral Biosciences, School of Dentistry, Chonbuk National University, Jeonju, Korea.

Tooth enamel is the most highly mineralized tissue in vertebrates, and its mineralization should be well-controlled under strict and harmonious enamel crystal formation and elongation to achieve exceptional hardness and a compact microstructure. Enamel matrix calcification occurs with several matrix proteins, such as amelogenin, enamelin and ameloblastin. Among them, amelogenin is the most abundant enamel matrix protein and multiple isoforms resulting from extensive but well-conserved alternative splicing and post-secretional processing have been identified. In this report, we recruited a family with unique enamel defect and identified a silent mutation in exon 4 of the AMELX gene. We show that the mutation caused the inclusion of exon 4, which is almost always skipped, in the mRNA transcript. We further show by generating and characterizing transgenic animal model that the alteration of the ratio and quantity of the developmentally conserved alternative splicing repertoire of AMELX caused defects in enamel matrix mineralization. This work was supported by grants from the Bio & Medical Technology Development Program (2013037491), the Science Research Center grant to Bone Metabolism Research Center (2008-0062614), and Basic Science Research Program (2013R1A2A1A01007642) through the National Research Foundation of Korea (NRF).

3115S

Loss of function mutations reveal that *DGAT1* is essential for gastrointestinal homoeostasis, lipid absorption and triglyceride deposition in humans and cows, but not rodents. *K. Lehnert[1,2], H. Ward[2], S.D. Berry[2], A. Ankersmit-Udy[2], A. Burrett[2], E.M. Beattie[2], N.L. Thomas[2], B. Harris[2], C.A. Ford[2], S.R. Browning[4], P. Rawson[7], G.A. Verkerk[5], Y. van der Does[6], L.F. Adams[2], S.R. Davis[2], T.W. Jordan[7], A.K.H. MacGibbon[6], R.J. Spelman[3], R.G. Snell[1,2].* 1) University of Auckland, Auckland, New Zealand; 2) ViaLactia Biosciences (NZ) Ltd., Auckland, New Zealand; 3) LIC, Hamilton, New Zealand; 4) Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; 5) DairyNZ Ltd, Hamilton, New Zealand; 6) Fonterra Research Centre, Palmerston North, New Zealand; 7) Centre for Biodiscovery, Victoria University of Wellington, Wellington, New Zealand.

Diacylglycerol O-acyltransferase 1 (DGAT1) catalyzes the transfer of fatty acid-CoA to diacylglycerol in the terminal step in triglyceride synthesis. Dgat1-/- mice exhibit increased longevity, decreased body fat, resistance to diet-induced obesity and a failure to lactate. The absence of detectable negative effects outside the mammary gland in Dgat1-deficient mice, and the enzyme's role in human storage lipid synthesis, have made pharmacological inhibition of DGAT1 an attractive approach to the management of obesity. Interestingly, a family has been reported with a contrary phenotype of growth retardation and diarrhoea due to homozygosity for a splice site mutation in *DGAT1*, abolishing the production of active enzyme [Haas et al., 2012, PMID: 23114594]. In an endeavour to select for altered milk production in dairy cows we have screened 2.5 million cows and identified an extremely unusual animal producing milk with a 40% reduction of saturated fat (4 SD below breed mean). To identify the genetic basis of this extreme phenotype, we bred a three-generation pedigree from the unique founder. Linkage mapping and genome sequencing revealed a *de novo* heterozygous A>C transversion in a previously unidentified, conserved splice enhancer motif in exon 16 of the DGAT1 gene [g.8078>C; p.M435L]. The mutation caused skipping of exon 16 in 90-95% of mutant mRNA molecules, and the mutant enzyme was unable to transfer fatty acids from CoA to diglyceride. Growth, development, weight, and life span of heterozygous cows were normal. However, calves homozygous for the mutation developed severe, non-bloody diarrhoea within 2-3 days after being apparently born healthy and with normal birth weights. The animals failed to thrive despite normal food intake, and serum triglyceride and cholesterol levels were extremely low. Gastrointestinal function normalized on isoenergetic low-fat diets. Bi-weekly intravenous supplementation of essential and unsaturated lipids was required to restore growth, albeit at slower rates compared to unrelated calves of similar weights. Our discovery of the mutation in animals supports the identification of the causal mutation in the family above and a potential role for DGAT1 in malabsorption and diarrhoeal disorders. Finally, our observations underscore concerns for DGAT1 inhibition as a strategy to combat obesity in humans, and suggest the use of a large animal model to understand the role of *DGAT1* in the intestine.

3116M

Hennekam syndrome can be caused by FAT4 mutations and be allelic to Van Maldergem syndrome. *M. Alders[1], L. Al-Gazali[2], I. Cordeiro[3], B. Dallapiccola[4], L. Garavelli[5], B. Tuysuz[6], F. Salehi[1], M. Haagmans[1], O. Mook[1], C. Majoie[1], M. Mannens[1], R. Hennekam[1].* 1) Academic Medical Center, Amsterdam, The Netherlands; 2) United Arab Emirates University, Al-Ain, United Arab Emirates; 3) Hospital Santa Maria. Lisboa, Portugal; 4) Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy; 5) Arcispedale S. Maria Nuova, Reggio Emilia, Italy; 6) Cerrahpasa Medical School, Istanbul University, Istanbul, Turkey.

The Hennekam lymphangiectasia-lymphedema syndrome is a genetically heterogeneous disorder. It can be caused by mutations in CCBE1 which are found in approximately 25% of cases. We used homozygosity mapping and whole exome sequencing in a family with multiple affected individuals in whom no CCBE1 mutation had been detected, and identified a homozygous mutation in the FAT4 gene. Subsequent targeted mutation analysis of FAT4 in a cohort of 24 CCBE1 mutation negative Hennekam syndrome patients identified homozygous or compound heterozygous mutations in 4 additional families. Mutations in FAT4 have been previously associated with Van Maldergem syndrome. Detailed clinical comparison between van Maldergem syndrome and Hennekam syndrome patients shows that there is a substantial overlap in phenotype, especially in facial appearance. We conclude that Hennekam syndrome can be caused by mutations in FAT4 and be allelic to Van Maldergem syndrome.

3117T

Novel variant of *TNNI2* causes an atypical Distal Arthrogryposis syndrome. *C.T. Marvin[1], J.X. Chong[1], K.J. Buckingham[1], K.M. Shively[1], A.E. Beck[1,2], H.I.S. Gildersleeve[1], M.J. McMillin[1], D.A. Nickerson[3], J. Shendure[3], R.E. Stevenson[4], M.J. Bamshad[1,2,3], University of Washington Center for Mendelian Genomics.* 1) Dept. of Pediatrics, University of Washington, Seattle, WA, USA; 2) Seattle Children's Hospital, Seattle, WA, USA; 3) Dept. of Genome Sciences, University of Washington, Seattle, WA, USA; 4) J.C. Self Research Institute of Human Genetics, Greenwood Genetic Center, Greenwood, SC, USA.

Distal arthrogryposis (DA) type 2B, or Sheldon-Hall syndrome, can be caused by pathogenic variants in any one of several genes that encode proteins of the contractile apparatus of skeletal muscles. Causal variants in four genes (*MYH3*, *TNNI2*, *TPM2* and *TNNT3*) each explain approximately 10% of families, so the genetic basis remains unknown in most circumstances. We were referred a single affected male who presented as a newborn with a history of reduced fetal movement accompanied by polyhydramnios, severe contractures of the upper and lower limbs, multiple dysmorphic facial characteristics, hyperelastic skin, multiple rib and long-bone fractures, and hypotonia born to apparently unaffected parents. Despite an exhaustive clinical evaluation, no specific diagnosis could be confirmed. However, evaluation by a group of expert dysmorphologists considered atypical DA2B to be one possible diagnosis. We screened exons of each gene known to underlie DA2B including exon 8 of *TNNI2*, exon 10 of *TNNT3,* and all coding exons of *MYH3* and *TPM2* without finding a compelling candidate variant. Subsequently, we performed exome sequencing of the proband and his parents and identified a *de novo* missense variant (c.337C>T, predicted p.Arg113Trp) in exon 7 of *TNNI2*. To date, variants causing DA2B have been reported only in exon 8 of *TNNI2*. Additionally, we identified a rare (MAF 0.024%) in-frame deletion (c.272_274del, predicted p.Glu92del) in exon 6 of *TNNI2*. This variant was also found in the mother of the proband who in retrospect was noted to also have somewhat hyperelastic skin. No candidate functional variants were found in genes potentially associated with hyperelastic skin. These results suggest that the atypical DA phenotype and in particular, the hyperelastic skin, may be caused by the occurrence of more than one disruptive variant in *TNNI2*, either on the same allele or affecting both alleles. Exome sequencing of additional atypical cases of DA2B is underway to determine whether causal variants exist in regions other than those encoded by exon 8 and whether the phenotype in some individuals is caused by multiple disruptive variants in *TNNI2*.

3118S

MTHFR and CBS: A risk factor for Down syndrome. *A. kaur, A. kaur.* Department of Human Genetics, Guru Nanak Dev University,., Amritsar, Punjab, India.

The mutations in metabolic pathway has been observed to be associated with higher birth frequency of Down syndrome. The most common mutations include MTHFR C677T (1p36.3) and CBS 844ins68 (21q22.3). MTHFR helps in converting homocystein to methionine while cystathionine beta synthase (CBS) helps in converting homocystein and serine to cystathionine, thus regulating the homocystein concentration. Mutations in MTHFR and CBS gene impair the activity of enzymes, leading to increased homocystein concentration, which has been reported to be the possible risk factor for Down syndrome. The present study investigated 126 mothers of DS children and 58 mothers having normal children for CBS 844ins68 polymorphism whereas 110 case mothers and 111 control mothers for MTHFR C677T polymorphism. Peripheral blood from mothers was collected in EDTA-coated vials for DNA extraction after taking informed consent. Genotyping was done using polymerase chain reaction. Products were electrophoresed in 2% agarose gel using ethidium bromide. In CBS 844ins68, presence of 68bp insertion produced 242bp band while wild type produced 174bp band. Among cases, 88.9% (112) were having -/- genotype and 11.11% (14) were having +/- genotype. However, +/+ genotype was not observed among any case or control mothers. The chi square value (χ^2- 0.065, p-0.79) suggested non-significant association between birth of DS child and presence of CBS 844ins68 polymorphism. In case of MTHFR C677T polymorphism, homozygous wild genotype (CC) was observed in 78.2%, heterozygous mutant (CT) in 20% and homozygous mutant (TT) in 1.8% mothers of DS children. On the other hand, control mothers had 80.2% CC and 19.8% CT genotype. The chi-square value (χ^2- 2.064) revealed non-significant association between cases and controls. We are working on larger sample size to understand the complex mechanism behind the presence of maternal polymorphisms and maternal risk of having DS child.

3119M

Investigation of a missense in NOTCH4 in autosomal dominant scleroderma. *C.J. Cardinale[1], D. Li[1], R.M. Chiavacci[1], L. Tian[1], S.J. Burnham[2], H. Hakonarson[1].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

We encountered an 8-year-old girl presenting with systemic sclerosis (SSc), i.e., scleroderma, and a positive family history. This patient had an affected maternal aunt and a maternal grandfather who died of SSc. The pedigree suggests an autosomal dominant inheritance model with incomplete pentrance. We performed whole-exome sequencing on the proband and aunt as well as the unaffected maternal grandmother. We examined missense, nonsense, splice-altering, and coding indels matching the inheritance model. Results were filtered to exclude synonymous variants, variants with minor allele frequency (MAF) greater than 0.5%, and variants previously identified in controls by our in-house exome variant database. Relevant candidates were taken forward for manual curation. A p.Met1415Ile substitution in *NOTCH4*, previously implicated in the pathways by which TGF-β induces pulmonary fibrosis—one of the most severe clinical manifestations of SSc—was identified as the most likely candidate. We have cloned an expressed an AcGFP1-tagged clone of full-length NOTCH4 in HEK 293T cells and examined the impact of the mutation through immunoprecipitation and luciferase reporter assays.

3120T

Molecular Spectrum of Mutations in CFTR gene: First Report from the Aegean region of Turkey and definition of three novel mutations. *A. Aykut[1], H. Onay[1], I.M. Tekin[1], F. Gulen[2], F. Ozkinay[1].* 1) Ege University Faculty of Medicine, Department Medical Genetics, Izmir,Turkey; 2) Ege University Faculty of Medicine, Department of Pediatric Allergy and Respiratory Diseases, Izmir,Turkey.

Cystic fibrosis (CF) is an autosomal recessive disease caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Up to date, more than 1900 different mutations in the CFTR gene have been reported. The aim of this study was to evaluate the spectrum of CFTR gene mutations in the Aegean region of Turkey. In this study, a total of 251 mutated CF alleles were identified in 194 individuals referred to our center for CFTR molecular analysis between January 2005 and May 2014. Among 251 mutant CFTR alleles, 55 different mutations, 3 of them being novel, were detected. The most common five mutations were F508del (32.8%), I148T (8%), 2183AA>G (7%), N1303K (4.7%) and G542X (4.7%). The novel mutations identified were p. S10N, c.2338delG and p.R1283G. The frequencies of CFTR mutations found in our region were slightly different from the frequencies found in general Turkish population.

3121S

Molecular Analysis of Dystrophic Epidermolysis Bullosa in Iran. *H. Vahidnezhad[1,2,3], L. Youssefian[2,3], M. Barzgar[4], S. Sotoudeh[2], Q. Li[3], N. Mozaffari[4], A. Isaian[2], M. Daneshpazhoh[2], M. Tabrizi[2], S. Zeinali[1], J. Uitto[3].* 1) Biotechnology Research Center, Department of Molecular Medicine, Pasteur Institute of Iran, Tehran, Iran; 2) Tehran University of Medical Sciences, Tehran, Tehran, Iran; 3) Department of Dermatology and Cutaneous Biology, Thomas Jefferson University, Philadelphia, Pennsylvania, USA; 4) Shahid Beheshti University of Medical Sciences, Tehran, Tehran, Iran;.

Epidermolysis bullosa (EB) is a clinically and genetically heterogeneous group of genodermatoses characterized by trauma induced mucocutaneous blistering and chronic epithelial fragility caused by mutations in at least 18 genes. The international consensus classification defines four types of EB based on the ultrastructural level of tissue separation: EB simplex (EBS), junctional EB (JEB), dystrophic EB (DEB) , and the Kindler syndrome. Numerous studies, mainly performed in European and US families with EB, have revealed a number of characteristic epidemiological and genetic features, which form the basis for current diagnostic and counseling strategies. However, there is currently no comprehensive study about the molecular epidemiology of EB in Iranian populations. In this study, we registered 250 EB families and here we report the evaluation of 150 EB families with ~ 200 affected individuals in Iran. Clinical examination and immunofluorescence mapping showed that EBS, JEB, and DEB represented 13, 15, and 66% of the cohort, respectively. Due to high rate of consanguineous marriages this ratio is different from EB patients in Europe and the USA in which EBS, JEB, and DEB were reported to represent 67, 27, and 33%, respectively. We have assessed a cohort of 67 DEB families for pathogenic sequence alterations in the *COL7A1* gene. Our results from 13 families show recurrent and novel mutations, including p.Gly2040Val, p.Arg2069Cys, p.Arg2063Trp , p.Gly2031Ser, c.6265delC, c.4958-4959delAA and IVS84+2 T>G (c.6714+2 T>G). Our data indicate the need for population-specific diagnostic and management approaches. Identification of specific mutations in the candidate genes and elucidation of the consequences of such mutations can provide a basis for the development of novel therapeutic approaches, including gene therapy, protein replacement, or cell based therapies, some of which are in early clinical trials for EB. Finally, DNA-based prenatal testing and preimplantation genetic diagnosis can be applied to families at risk for recurrence of this, currently intractable disease.

3122M

Genome-wide association meta-analysis of 6,365 subjects replicate EHF-APIP and identifies new modifier loci of lung disease severity in Cystic Fibrosis. *H. Corvol[1], G.R. Cutting[2], M.L. Drumm[3], M.R. Knowles[4], J.M. Rommens[5] on behalf of the International CF Gene Modifier Consortium.* 1) AP-HP, Hôpital Trousseau, Pediatric Pulmonary Department; Inserm U938, Paris, France; 2) Institute of Genetic Medicine, Johns Hopkins University, Baltimore MD; 3) Dept of Pediatrics, Case Western Reserve University, Cleveland, OH; 4) Dept of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC; 5) Program in Genetics & Genome Biology, SickKids Research Institute and Univ of Toronto, Toronto, ON, Canada.

Lung disease is responsible for 80% of mortality in cystic fibrosis (CF). Genetic modifiers play a prominent role in variation in lung function measurements (H^2~0.54). The North American Cystic Fibrosis (CF) Modifier Consortium had previously reported genome-wide significant association with lung function measures for SNPs positioned between *APIP* and *EHF* on chromosome 11p13 (GWAS1; *n* = 3,444; Wright *et al, Nat Genetics*, 2011). We now report replication of the *EHF-APIP* locus and identification of new modifier loci for lung disease severity in CF. Genome-wide SNP typing was performed on a second cohort of CF subjects from North America and the French CF Gene Modifier Consortium (GWAS2; *n*=2,921). Imputation using an admixed 1000 Genomes reference population generated a common set of 8,520,458 SNPs that were screened for association using linear mixed effects modeling with meta-analysis across patient groups and genotype platforms. Support for the chromosome 11p13 region was evident in a meta-analysis of GWAS1 and GWAS 2 (*P* = 4.8 x 10-9; GWAS1+2) and in the subgroup homozygous for the common CF mutation (*P* = 1.9 x 10-10; GWAS1+2 p.Phe508del p.Phe508del). Regions on chromosomes 5, 6 and X that previously indicated suggestive evidence for associations (GWAS1) met the genome-wide significance criterion (*P* = 6.8 x 10-12, 1.7 x 10-9 and 1.8 x 10-9, respectively; GWAS1+2) in the meta-analysis. Two new loci showed significant evidence for association, chromosome 3 (*P* = 3.3 x 10-11; GWAS1+2) and chromosome 17 (*P* = 1.8 x 10-7, GWAS2; and *P* = 2.4 x 10-9, GWAS2 p.Phe508del/p.Phe508del). The intervals of the six identified loci that contribute to lung function in CF patients suggest a series of biologically compelling genes that include *EHF* and *APIP*, the solute carrier *SLC9A3,* the mucins *MUC4* and *MUC20*, class II *MHC* genes, the lncRNA CASC17 and an angiotensin receptor *AGTR2* that is proximal to another solute carrier *SLC6A14.* Pleiotropic effects are evident as *SLC9A3* and *SLC6A14* have been observed to be modifiers of a different CF phenotype, neonatal intestinal obstruction. Together, these loci provide new molecular targets for therapeutic intervention of the major cause of mortality in CF.

3123T

A Homozygous NIPAL4 Mutation In A Case With Ichthyosis And Deafness. *E. Arslan Ates[1], H. Onay[2], I. Ertam[3], B. Senturk[3], F. Ozkinay[2].* 1) Medical Genetics, Sevket Yilmaz Training and Research Hospital, Bursa, Turkey; 2) Medical Genetics, Ege University Faculty of Medicine, Izmir, Turkey; 3) Dermatology, Ege University Faculty of Medicine, Izmir, Turkey.

Hereditary ichthyosis is a a group of keratinization disorder characterized by skin scaling and dryness. It is classified as syndromic and nonsyndromic ichthyosis depending on the presence of systemic findings. Keratitits-Ichthyosis-Deafness (KID, OMIM ID#148210) syndrome is one of the syndromic Ichthyosis disorders. Its cardinal features are ichthyosis, sensorineural deafness and ophthalmological findings. Heterozygous Connexin 26 (GJB2) gene mutations have been described as the underlying defects of KID. Autosomal Recessive Congenital Ichyhosis (ARCI) is a nonsyndromic ichthyosis form, without any systemic involvement. TGM1, NIPAL4, ALOX12B, ALOXE3, CYP4F22, ABCA12, PNPLA1 and LIPN genes have been described associated with ARCI. Here, we present a 6 year-old girl with ichthyosis and deafness. She was born to consanguineous parents. Scaling and redness of skin were recognised at birth. On physical examination she had typical skin findings of ichthyosis. Histopathological evaluation of skin biopsy was compatible with vulgar icthyosis. She had sensoryneural deafness. Ophthalmologic examination was normal. Because she was considered to have KID Syndrome, CX26 gene was sequenced and no mutation was found. Subsequently, TGM1 gene which have been known to be the most common cause of ARCI has been sequnced but no mutation was found in this gene as well. The second most common cause of ARCI are the defects found in NIPAL4 gene. Therefore NIPAL 4 gene was sequenced as the third gene and a homozygous c.527C>A (p.A176D) mutation was detected. To date there have been no reported ARCI case having deafness and NIPAL4 mutations in the literature. Althoug there may be a coincidence in this patient, NIPAL4 mutation analysis should be considered in patients with KID syndrome or in patients having both ichthyosis and deafness.

3124S

The *ALK1* IVS3-35A>G polymorphism is associated with arteriovenous malformations in hereditary hemorrhagic telangiectasia patients with *ENG* mutations, but not in patients with *ALK1* mutations. *L. Pawlikowska[1,2], J. Nelson[1], D.E. Guo[1], C.E. McCulloch[3], M.T. Lawton[4], W.L. Young[1,4], H. Kim[1,2,3], M.E. Faughnan[5,6], Brain Vascular Malformation Consortium HHT Investigator Group.* 1) Center for Cerebrovascular Research, Dept of Anesthesia and Perioperative Care, University of California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 3) Dept of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; 4) Dept of Neurological Surgery, University of California, San Francisco, San Francisco, CA; 5) Division of Respirology, Department of Medicine, University of Toronto, Toronto, Canada; 6) Division of Respirology, Dept of Medicine and Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada.

Background. Hereditary hemorrhagic telangiectasia (HHT) is caused by mutations in several TGFβ signaling pathway genes. Most HHT patients have a mutation in *ENG* or *ALK1* (*ACVRL1*) and have diverse phenotypes including chronic bleeding, hemorrhage, stroke, heart failure and other complications related to skin and mucosal telangiectases and organ arteriovenous malformations (AVM). The phenotypic heterogeneity of HHT suggests a possible role for genetic modifier effects. The common polymorphisms *ALK1* IVS3-35A>G and *ENG* 207G>A have been reported to be associated with sporadic brain AVM (BAVM); the *ALK1* IVS3-35A>G association has been replicated in 2 independent cohorts. We hypothesized that these polymorphisms are also associated with AVM in HHT patients. **Methods.** We genotyped *ALK1* IVS3-35A>G and *ENG* 207G>A in 716 HHT patients (95.9% Caucasian) recruited by the Brain Vascular Malformation Consortium. We evaluated association of genotype with AVM, defined as presence of any BAVM, liver AVM (LAVM) or pulmonary AVM (PAVM), and separately with each AVM type. Logistic regression analyses were adjusted for age, gender and family structure and further stratified by HHT mutation status (*ALK1* or *ENG*). **Results.** Among HHT patients with complete clinical data, 71% had at least one AVM, 50% had PAVM, 20% had BAVM and 20% had LAVM. Among 436 (61%) patients with mutation status available, 48% had *ENG* mutations and 43% had *ALK1* mutations. Among all 716 HHT patients, neither polymorphism was significantly associated with AVM, although *ALK1* IVS3-35A>G showed a trend toward association with PAVM (OR=1.48, 95%CI=0.90-2.22, p=0.062). When stratifying by HHT mutation, *ALK1* IVS3-35A>G was significantly associated with AVM among *ENG* mutation carriers (OR=2.66, 95%CI=1.15-6.13, p=0.022), but not among *ALK1* mutation carriers (OR=0.79, CI=0.38-1.63, p=0.52). *ALK1* IVS3-35A>G was also significantly associated with PAVM (OR=2.45, p=0.016) and LAVM (p<0.05, all 42 LAVM-positive patients carried the A risk allele) among *ENG* mutation carriers. There were no significant associations between *ENG* 207G>A and AVM, but a trend toward association with BAVM (OR=3.09, p=0.16) was observed among *ALK1* mutation carriers. **Conclusions.** *ALK1* IVS3-35A>G was associated with AVM in HHT patients bearing *ENG* but not *ALK1* mutations. These results suggest that common polymorphisms in TGFβ pathway genes other than the mutated HHT gene act as genetic modifiers and contribute to HHT phenotypic heterogeneity.

3125M

Neutral Lipid Storage Disease with Myopathy: disease modeling using patients' hiPSc. *D.A. Coviello[1], S. Missaglia[2], M. Castagnetta[1], E.M. Pennisi[3], M. Mogni[1], D. Tavian[2,4].* 1) Laboratory of Human Genetics, E.O. Ospedali Galliera, Genoa, GE, Italy; 2) Laboratory of Cellular Biochemistry and Molecular Biology, CRIBENS, Catholic University of the Sacred Heart, Milan, MI, Italy; 3) UOC Neurologia, A.C.O. San Filippo Neri, Rome, RM, Italy; 4) Psychology Department, Catholic University of the Sacred Heart, Milan, MI, Italy.

Mutations in the PNPLA2 gene cause the onset of Neutral Lipid Storage Disease with Myopathy (NLSD-M), a rare autosomal recessive disorder characterized by an abnormal accumulation of triacylglycerol into cytoplasmic lipid droplets (LDs). In most tissues the LDs are cellular organelles for the triacylglycerol storage. LDs metabolic functions are mediated by proteins bound to their surface. In particular, the lipase that catalyzes the removal of the first acyl chain from triacylglycerol is the adipose triglyceride lipase (ATGL), also known as patatin-like phospholipase domain-containing protein 2 (PNPLA2). To our best knowledge, twenty six different PNPLA2 mutations have been described in thirty two NLSD-M patients. NLSD-M patients are mainly affected by progressive myopathy, cardiomyopathy and hepatomegaly. However, their clinical severity appears to be highly variable. Other clinical symptoms may include diabetes, chronic pancreatitis and short stature. NLSD-M has, at present, no specific therapy. We have previously reported clinical and genetic findings of some NLSD-M patients obtaining dermal biopsies from them. Here we report the development of hiPSc (human induced pluripotent stem cell) from patients' fibroblasts harboring different PNPLA2 mutations. The first patient was found to be homozygous for a deletion at nucleotide 542 (c.542delAC). This deletion caused a premature stop codon at position 212. The molecular analysis of patient 2 showed a homozygous missense mutation, c.662G>C (p.R221P). Initial hiPSc colony selection was based on morphologic evaluation and on detection of pluripotency surface markers (SSEA-4 and TRA-1-81). HiPSc also expressed undifferentiated ES cell markers (NANOG, SOX2 and OCT4). Moreover, embryoid bodies (EBs) have been generated from NLSD-M-iPSCs to assess the pluripotent properties of these cells. Karyotypic analysis of hiPSc lines indicated a normal complement of chromosomes. Immunohystochemical evaluations of LDs on hiPSc revealed that they recapitulate pathological hallmark of the disease. We propose use of inherently patients- and disease specific hiPSc to study the pathogenetic mechanisms leading to NLSD-M and as a potential model for therapeutic evaluation. Communications to: daniela.tavian@unicatt.it.

3126T

British Ectopia Lentis (EL) patients with novel *ADAMTSL4* mutations: 2 homozygotes, 1 compound heterozygote & 1 compound heterozygote involving a splice site. *J.A. Aragon-Martin[1], Y.B.A. Wan[1], L. Collins[1], P. Mazadza[1], G. Arno[2], A. Chandra[3], D.G. Charteris[4], A. Saggar[5], A.H. Child[1].* 1) Cardiovascular Sciences Research Centre, St George's University of London, London, England, United Kingdom; 2) Institute of Ophthalmology, University College London, London, United Kingdom; 3) Vitreoretinal Unit, Royal Victorian Eye and Ear Hospital, Melbourne, Australia; 4) Vitreoretinal Surgery, Moorfields Eye Hospital, London, United Kingdom; 5) Clinical Developmental Sciences, St George's University of London, London, United Kingdom.

Background: Ectopia Lentis (EL) is clinically & genetically heterogeneous (autosomal dominant - MIM 129600; autosomal recessive - MIM 225100). The dominant form can arise through mutations in *FBN1* (Comeglio et al. 2007), at the milder end of the type-1 fibrillinopathy spectrum. The recessive form has been shown to be caused by *ADAMTSL4* mutations (Aragon et al. 2010). EL is characterized by the disruption of the zonular fibers. In time, the lens moves out of place causing lens subluxation.

Methods: Seventy-five UK patients with isolated EL were bidirectionally Sanger sequenced for mutations in the coding exons including intron/exon boundaries of *ADAMTSL4* gene. The patients did not fulfill the Ghent criteria for Marfan syndrome & 73% were *FBN1* negative. When possible, compound heterozygotes were confirmed by family screening. Mutations were based on the transcript NM_019032. Genome databases (1000Genome & EVS) & in-silico tools (GVGD, Mutation T@ster, Polyphene-2, SIFT, BDGP & ESE_finder) were used to help with the analysis of the mutations.

Results: Mutations in *ADAMTSL4* were found in 51/75 (68%) samples. Genetic Carriers - 19/75 (25%). Compound Heterozygotes - 12/75 (16%). Homozygotes - 20/75 (27%). Novel mutations: 2 homozygotes - [**c.963dupG/p.T322Dfs*10**] & [**c.2574_2580dupTGGGACG/p.G861Wfs*41**]; 1 compound heterozygote - [**c.888delG/p.R297Dfs*46** & **c.2594G>A/p.R865H**]; & 1 compound heterozygote involving a splice site - [**c.2270dupG/p.G758Wfs*59** & **c.2383-1G>C/splice site**]. The mutation p.R865H has been found so far as compound heterozygote in 4 patients; & once in an unaffected genetic carrier, parent of a compound heterozygote proband in our cohort.

Conclusions: This study supports the evidence that homozygous & compound heterozygous mutations in *ADAMTSL4* cause autosomal recessive EL, therefore demonstrating that *ADAMTSL4* could play a key role in the development &/or integrity of the zonular fibers. The identification of a causative mutation in *ADAMTSL4* will exclude Marfan syndrome in EL patients & guide clinical management, of particular relevance in young children affected with EL. In our EL population we screen first for *FBN1* mutations in marfanoid patients with autosomal dominant pedigrees, & first for *ADAMTSL4* in apparently autosomal recessive pedigrees. **Corresponding author**: jaragon@sgul.ac.uk.

3127S

Xeroderma Pigmentosum diagnosed in adulthood; atypical clinical presentation associated with a novel genetic defect in XPC gene. *C. Badenas[1,2], M. Meneses[1], M. Chavez-Bourgeois[1], S. Villablanca[1], P. Aguilera[1,2], A. Bennassar[1], Ll. Alos[1], S. Puig[1,2,3], J. Malvehy[1,2], C. Carrera[1,2].* 1) Hospital Clinic, Barcelona, Barcelona, Spain; 2) CIBER de Enfermedades Raras, Instituto de Salud Carlos III, Barcelona, Spain; 3) Universitat de Barcelona, Spain.

Importance: A case of Xeroderma Pigmentosum (XP) with a novel mutation in the XPC gene was diagnosed in a patient aged 42 after diagnosing multiple primary melanomas but none cutaneous carcinoma. XP is a rare genodermatosis caused by abnormal DNA-repair. XP complementation group C (XPC) is the most frequent type in Mediterranean countries. Early marked photoaging and non-melanoma skin cancers are the common clinical presentation from childhood. Observations: A phototype III otherwise healthy Caucasian male patient was referred to our Unit presenting sporadic melanoma. Initially another 4 melanomas were detected. Molecular studies did not identify mutations in CDKN2A, CDK4 or MITF genes. Two mutations in the XPC gene were detected: a c.2287delC (p.Leu763Cysfs*4) frameshift and a c.2212A>G (p.Thr738Ala) missense mutations. After digital follow-up, another 5 additional primary melanomas were diagnosed. Conclusions and Relevance: The p.Thr738Ala missense mutation has not been previously described. It may be the case that missense mutations in the XPC gene allow a partial functionality that could explain this unusual late onset XP. Atypical clinical presentation of XPC could be misdiagnosed as marked photoaging when genetic aberrations allow partial DNA repair capacity. Digital follow-up is the best strategy to ensure melanoma-early detection in these high-risk settings.

3128M

Mutation spectrum of the *ABCA4* Gene in Greek patients with Stargardt disease: Identification of two novel mutations and evidence of three prevalent mutated alleles. *S. KAMAKARI[1,2], G. KOUTSODONTIS[1], P. STAMATIOU[2], T. PANAGIOTOGLOU[3], I. DATSERIS[1], M. TSILIMBARIS[3].* 1) OPHTHALMIC GENETICS UNIT, OMMA OPHTHALMOLOGICAL INSTITUTE OF ATHENS, ATHENS, GREECE; 2) SCHOOL OF MEDICINE, UNIVERSITY OF ATHENS, GREECE; 3) DEPARTMENT OF OPHTHALMOLOGY, SCHOOL OF MEDICINE, IRAKLION, GREECE.

Purpose: Mutations in the *ABCA4*, photoreceptor-specific ATP-binding cassette transporter 4 gene (MIM 601691) have been associated with autosomal recessive Stargardt disease (STGD; MIM 248200) characterized by central vision impairment. This is the first systematic study investigating the frequency and pattern of disease-associated mutations among Greek STGD patients, never previously genetically characterized for sequence variations in the *ABCA4* gene. Methods: 34 unrelated and 2 related STGD patients were analyzed using the combined methodology of ABCR400 microarray analysis and direct sequencing of 4 selected or all 50 exons and flanking intronic regions of the *ABCA4* gene. Results: Disease-associated mutations, including two novel splice defects, c.4352+1G>A and c.5714+1G>C in introns 29 and 40, respectively, were detected in 28/34 unrelated patients analyzed (82.35%). Both mutant alleles were found in 16/28 cases (57.14%), whereas in 12/28 (42.85%) only one allele was identified. The major pathogenic allele c.5714+5G>A accounted for 28.57% of the mutant alleles. Other frequently mutant alleles were c.1622T>C/p.(Leu541Pro) alone or as a complex c.1622T>C, c.3113C>T /p.(Leu541Pro), p.(Ala1038Val) allele and c.5882G>A/p.(Gly1961Glu), each accounting for 23.8% of the mutated alleles. Notably, a young patient with unaffected parents whose paternal aunts were unexpectedly affected, were found homozygous for the p.(Leu541Pro)/p.(Ala1038Val) and compound heterozygous for the p.(Leu541Pro)/p.(Ala1038Val) and p.(Gly1961Glu) alleles, respectively thus resolving the atypical inheritance pattern and different disease severity. Conclusions: This first report of the *ABCA4* mutation spectrum underlying STGD disease in Greece further elucidates the distribution of *ABCA4* mutations in European populations. The detection rate by the combined methodology was 82.35%. We identified two novel, potentially pathogenic *ABCA4* splice mutations and 3 prevalent disease-causing alleles. Further evaluation of the genotype-phenotype correlations will advance our knowledge of STGD 's etiology in Greek patients.

3129S

Nlz1 is required for cilia formation in zebrafish embryogenesis. *S. Dutta[1], S. Sriskanda[1], E. Boobalan[1], R. Alur[1], M. Gunay-Aygun[2], A. Elkahloun[3], M. Asai-Coakwell[4], O. Lehmann[4], E. Valente[5], A. Micalizzi[5], J. Lander[6], M. Romani[5], S. Ware[7], N. Katsanis[8], B. Brooks[1].* 1) 1.Unit on Pediatric, Developmental & Genetic Ophthalmology, Ophthalmic Genetics and Visual Function Branch, National Eye Institute; 2) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; 3) 3.Microarray Core, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; 4) Department of Ophthalmology, University of Alberta, Edmonton, AL, Canada; 5) IRCCS Casa Sollievo della Sofferenza, Mendel Laboratory San Giovanni Rotondo, Italy; 6) Cincinnati Children's Hospital, Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA; 7) Indiana University School of Medicine, Wells Center for Pediatric Research, Indianapolis, IN; 8) Department of Cell Biology, Center for Human Disease Modeling, Duke University, Durham, NC.

The formation of cilia is a fundamental developmental process affecting diverse functions such as intracellular signaling, mechanotransduction, tissue morphogenesis and body patterning. Defects in ciliogenesis result in a diverse spectrum of human diseases, known as ciliopathies. However, the precise mechanisms of ciliogensis during vertebrate development are not fully understood. In this report we describe a novel role of the Nlz1 protein in ciliogenesis. We demonstrate morpholino-mediated knockdown of nlz1 in zebrafish caused abnormal specification of the cells of Kupffer's vesicle (KV, similar to the mammalian node); a severe reduction of the number of cilia in KV, the pronephros, and the neural floorplate; as well as a spectrum of later phenotypes reminiscent of human ciliopathies. nlz1 is expressed in KV and in the nucleus of dividing cells. In vitro and in vivo data indicate that Nlz1 acts downstream of the ciliary "master transcription factor", Foxj1a and, Nlz1, is upregulated by canonical Wnt signaling. Mutation screening of NLZ1 in humans with a variety of ciliopathies and left-right patterning defects are underway. Together, our data suggest a novel role of nlz1 in ciliogenesis and the morphogenesis of multiple tissues.

3130S

A primary ciliopathy protein plays an extra-ciliary role in neurodevelopmental disease. *N. Nuangchamnong[1], C.S. Carter[2,3], Q. Zhang[2,3], T. Vogel[4], V.C. Sheffield[2,3].* 1) Obstetrics and Gynecology, University of Iowa, Iowa City, IA; 2) Pediatrics, University of Iowa, Iowa City, IA; 3) Howard Hughes Medical Institute, University of Iowa, Iowa City, IA; 4) Neurosurgery, University of Cincinnati, Cincinnati, OH.

Hydrocephalus affects 1 to 3 of every 1000 live births and is a significant cause of perinatal and pediatric morbidity and mortality despite current surgical interventions. The majority of congenital cases are idiopathic, but hydrocephalus is increasingly considered to be a neurodevelopmental disorder. Bardet-Biedl Syndrome (BBS) is an autosomal recessive human ciliopathy in which neurological defects are common findings, and in which ventriculomegaly occurs at a higher incidence than in the general population. Primary cilia are now recognized for playing a critical role in development of the central nervous system. We have previously shown that a mouse model with a mutated component of the BBSome complex (BBS1) has aberrant cell signaling at primary cilia that causes decreased survival and proliferation of a specific class of neural progenitor cells and leads to mild communicating neonatal hydrocephalus that can be partially rescued with lithium. Further exploration of the molecular mechanisms underlying hydrocephalus has the potential to make noninvasive pharmacologic treatments possible. In this study, we use a hydrocephalic mouse model that lacks BBS3, a GTPase that is necessary for localization of the BBSome to the ciliary gate, but that, unlike BBS1, is not a component of the BBSome complex. We compare the development of hydrocephalus in the BBS3 knockout mouse to wild type littermate controls and to mice with a mutated component of the BBSome complex (such as BBS1). We find late fetal onset of hydrocephalus in the BBS3 knockout mouse model with ventriculomegaly as early as embryonic day 17 (E17). We also find decreased apoptosis along the cerebellar ventricular zones and increased apoptosis at the midbrain tectum overlying the Sylvian aqueduct of BBS3 knockout mice at E17. Similar to the BBS1 mouse model, there is increased apoptosis and decreased proliferation of cells in the subventricular zone of the lateral ventricles in the neonatal period. Unlike the BBS1 model, however, there is fetal rather than neonatal onset of ventricular dilation and postnatal progression to severe disease that is consistent with a non-communicating (obstructive) type of hydrocephalus. The knockout BBS3 mouse is a useful model of obstructive congenital hydrocephalus, as BBS3 may have an extra-ciliary function that prevents progression of hydrocephalus from mild to severe.

3131S

Congenital heart disease associated to PCD. *J. Wallmeier[1], H. Olbrich[1], P. Pennekamp[1], C. Edelbusch[1], N.T. Loges[1], C. Werner[1], D. Kececioglu[2], B. Stiller[3], H. Omran[1].* 1) Department of Pediatrics, University Hospital Muenster, Muenster, Germany. , Muenster, NRW, Germany; 2) Klinik fuer Angeborene Herzfehler; Herz- und Diabeteszentrum NRW; Georgstraβe 11; 32545 Bad Oeynhausen,Germany; 3) Department of Pediatric Cardiology, University Hospital Freiburg, Freiburg, Germany.

Primary ciliary dyskinesia (PCD) is a rare genetic disorder, that is characterized by recurrent infections of the upper and lower respiratory airways, sinusitis and otitis media, and is often associated with reduced fertility of males. In 50% of PCD-patients laterality defects occur, mostly *situs inversus totalis* (Kartagener's syndrome). About 6% of all patients show heterotaxy. There are certain organ abnormalities that are associated with heterotaxy such as asplenia and polysplenia. An important consequence associated with heterotaxy are complex cardiac defects. Dysfunction of motile cilia and flagella explains the complex PCD phenotype involving various organ systems. Respiratory epithelia covered with cilia move extracellular mucus towards the throat with a coordinated beating (mucociliary clearance). Loss of this mucociliary clearance causes recurrent respiratory infections. Dysfunction of nodal cilia during early embryogenesis causes randomization of left/right body asymmetry. So far 28 genes have been identified to cause PCD either coding for axonemal proteins or ciliary assembly factors. One of the first identified and most often mutated genes *DNAH5* causes an outer dynein arm defect leading to cilia with just residual ciliary beating. We sequenced all coding exons of *DNAH5*, by sanger sequencing including the exon-/intron boundaries in a cohort of 101 patients. All patients were followed up in cardiologic centres, suffered from heterotaxy and/or congenital heart defects, and had not been further evaluated for PCD. We refer to heterotaxy as the instance of *situs inversus totalis*, which means a mirror image of the common situs (*situs solitus*) as well as the instance of *situs ambiguous*, an abnormal arrangement of the thoracic or abdominal organs. Interestingly, two loss-of-function mutations in *DNAH5* (deletion-/stop-mutation) were identified. The results of these genetic analyses established the diagnosis of PCD. Even though PCD is a rare disorder 1% of our cohort with heterotaxy showed a mutation in one of the most common PCD causing genes. Because *DNAH5* only accounts for ~25% of PCD cases possibly some more patients of the analysed heterotaxia cohort will suffer from PCD. Therefore we like to emphasize the importance of the differential diagnosis of PCD in patients with CHD and respiratory disease.

† Trainee Award Finalist

3132S

PKHD1 mutations are associated with the whole spectrum of ductal plate malformations. *J.B. Courcet[1,2], A. Minello[3], F. Prieur[4], L. Morisse[5], J.M. Phelip[6], A. Beurdeley[7], D. Meynard[8], D. Massenet[8], F. Lacassin[9], Y. Duffourd[2,10], N. Gigot[2,10], J. St-Onge[2,10], P. Hillon[3], C. Vanlemmens[11], C. Mousson[12], J.P Cerceuil[13], B. Guiu[13], J. Thevenon[1,2], C. Thauvin-Robinet[1,2], E. Jacquemin[14], J.B. Rivière[2,10], L. Michel-Calemard[15], L. Faivre[1,2].* 1) Service de pédiatrie 1 et génétique médicale, Centre Hospitalo-universitaire, Dijon, France; 2) Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement (TRANSLAD), Centre Hospitalier Universitaire de Dijon, Dijon, France; 3) Service d'hépato-gastro-entérologie, Centre Hospitalo-Universitaire, Dijon, France; 4) Service De Génétique Clinique Chromosomique et Moléculaire, Pole De Biologie, Centre Hospitalo-Universitaire De Saint-Etienne, Saint-Priest-En-Jarez, France; 5) Service de médecine polyvalente, Hôpital de SIA, Uvea, France; 6) Service d'Hépato-gastro-entérologie, Centre Hospitalo-Universitaire, Sant-Etienne, France; 7) Service de chirurgie, Hôpital de SIA, Uvea, France; 8) Laboratoire de biologie médicale, Hôpital de SIA, Uvea, France; 9) Service de Médecine Interne, Hôpital Magenta, Nouméa, France; 10) Laboratoire de biologie moléculaire, Centre Hospitalo-Universitaire, Dijon, France; 11) Service de gastro-entérologie et hépatologie, Centre Hospitalier Régional Universitaire Hôpital Jean Minjoz, Besançon, France; 12) Service de néphrologie, Centre Hospitalo-Universitaire, Dijon, France; 13) Service de radiologie, Centre Hospitalo-Universitaire, Dijon, France; 14) Service d'hépatologie pédiatrique, Centre Hospitalo-Universitaire, Le Kremlin-Bicêtre, France; 15) Service d'endocrinologie moléculaire et maladies rares, Centre de Biologie et Pathologie Est CHU de Lyon-GH Est, Hospices Civils de Lyon, Bron, France.

Ductal plate malformations (DPM) present with a wide phenotypic spectrum comprising Von Meyenburg complexes (VMC), Caroli disease (CD), Caroli syndrome (CS), and autosomal recessive polycystic kidney disease (ARPKD). Familial forms of CD have been rarely described. In a family of 3 siblings presenting with the whole spectrum of severity of DPM (CS with renal cysts, CD and VMC), we performed whole exome sequencing and identified 2 compound heterozygous mutations of PKHD1, including in cis c.10444G>A; p.Arg3482Cys and in trans c.5521C>T; p.Glu1841Lys, segregating with the symptoms. Thereafter, we screened two other familial cases of DPM with at least one patient presenting with CD, and identified 2 compound heterozygous PKHD1 variants in each case, including one hypomorphic mutation. PKHD1 gene encodes fibrocystin, which is expressed in the primary cilium of renal epithelial cells and bile ducts. Mutations in PKHD1 are responsible for ARPKD, with an inconstant liver phenotype that could remain asymptomatic. Only few patients are diagnosed at the adult age with liver-related complications that could be related to CS in association with mild kidney disease. Exceptional cases of CD are associated with PKHD1 mutations. This report, together with the recent literature, confirmed that PKHD1 is a major gene responsible for isolated Caroli disease, and also expands the clinical manifestations associated with mutations of PKHD1 to VMC and thus to the whole spectrum of DPM. In such mild phenotypes, systematic hepatic and renal screening appears mandatory in asymptomatic siblings of affected patients, in order to determine at risk patients that should beneficiate from regular follow-up. These results give another example of the major inter- and intra-familial phenotypic variability associated with PKHD1 mutations. Whereas truncating mutations of PKHD1 are associated with the more severe phenotype, the pathogenicity of missense mutation remains unclear and the association of missense / hypomorphic mutations could explain mild phenotypes. Finally, this work adds an example of the contribution of exome sequencing, not only in the discovery of new genes but also in expanding the phenotypic spectrum of well-known disease-associated genes, using reverse phenotyping.

3133S

Stem Cells from Offspring of Mothers Demonstrate Evidence for Developmental Programming in Obesity. *P.R. Baker, K.E. Boyle, A.L. Butti, D. Dabelea, L.A. Barbour, J.E. Friedman.* University of Colorado, Aroura, CO.

Background: Developmental programming is a major contributor to obesity. Although there is strong evidence for the long-term effects of in utero exposure to maternal diabetes and obesity, and potential fetal overnutrition, the exact mechanisms by which these effects are transmitted from human mother to offspring remain poorly understood. Skeletal muscle of obese adults shows alterations in key mitochondrial energy pathways (including branched chain amino acid (BCAA) dependent intermediary metabolism and compromised fatty acid oxidation (FAO)). Myocytes derived from multipotent MSC populations have the potential to reveal early changes that may predispose infants to obesity. Here we test the hypothesis that umbilical cord-derived MSCs from infants of obese (Ob) vs normal weight (NW) mothers differentiated into myocytes (skMSC) manifest impaired mitochondrial metabolism when challenged with lipid exposure. Design/Methods: We used data from the Healthy Start Study. This pre-birth cohort study of maternal-infant pairs provides detailed maternal data, including pre-pregnant body mass index (BMI), and cord tissue samples. MSCs derived from full term infants born to NW (BMI<25, n=5) and Ob mothers (BMI>30, n=5) were grown and differentiated into myocytes under control and lipid exposed conditions. On day 21 of differentiation media from all cells was tested for amino acid concentrations using the Biochrom 30 Amino Acid Analyzer. Two-way Analysis of Variance compared the effects of maternal pre-pregnant BMI, treatment, and their interaction. Results: SkMSC from infants of Ob mothers had elevations in BCAA (leucine and isoleucine), Alanine, and Citrulline versus skMSC from infants of NW mothers (p=0.05, 0.04, 0.01, and 0.05). All of these increased with exposure to lipid. Only 2-aminobutyric acid, an under-recognized methionine and threonine intermediary metabolite, decreased with lipid exposure in Ob but increased in NW samples (p= 0.02 interaction). Conclusions: Differential handling of amino acids suggest differences in mitochondrial energy metabolism in offspring of Ob mothers. More data is needed to demonstrate use in intermediary metabolism rather than differences in cell transport or anabolism, however our data is consistent with amino acid findings in obese adult serum. This MSC model will further be valuable in the targeted study of energy metabolomics, transcriptomics, and epigenomics in fetal programming and obesity.

3134S

Aberrant activation of the sex-determining gene in early embryonic development results in postnatal growth retardation and lethality in transgenic mice. *T. Kido, Z. Sun, Y. Li, Y. Lau.* Dept Med, VA Med Ctr 111C5, Univ California, San Francisco, San Francisco, CA.

Sexual dimorphisms are prevalent between the sexes, particularly in neurodevelopment. Currently, the role of the Y chromosome in such phenomena has not been clearly defined. Among the Y chromosome genes, the sex-determining gene, *SRY*, could be a significant candidate capable of exerting male-specific effects on sexual dimorphisms. *SRY* is the founder of the SRY-box (*SOX*) genes, which are key regulators for various developmental processes. SRY and SOX proteins harbor a conserved HMG box DNA-binding domain. An ectopically expressed SRY could compete with the SOX factors in regulating the respective SOX targets, thereby disrupting the corresponding developmental processes. To test this hypothesis, we have established a transgene activation mouse system, in which a *SRY* transgene could be activated by a tissue-specific Cre. The Ddx4-Cre transgenic line expresses the Cre recombinase exclusively in the germ lines of both sexes. When a female Ddx4-Cre is crossed with a male SRY responder, the Cre recombinase is transferred to the single-cell embryo from the fertilized oocyte, thereby activating the SRY transgene in all tissues. However, when a male Ddx4-Cre is crossed with a female SRY responder, the *SRY* transgene could not be activated till later in germ cells. Hence, *SRY* transgene could be activated differentially with the respective sex of the Ddx4-Cre mice. Our results show that early activation of *SRY* transgene during embryonic development results in significant postnatal growth retardation of the offspring, which do not survive beyond two weeks of age while late and male germ cell activation of *SRY* transgene result in normal postnatal development. Characterizations of pups with early *SRY* transgene activation show significant abnormalities in heart, liver and brain development. Comparative transcriptome analysis of the brains between mutant and control pups show increases in functions of organismal death, hypoplasia, abnormalities in head and cerebrum, and lack of dentate gyrus development but significant decreases in functions of transcription, cell quantity, dendritic cell migration, long-term potentiation, neuritogenesis, behavior, learning, cognition, and dendritic growth/branching. Our results suggest that ectopic activation of *SRY* in non-gonadal tissues could have significant disruptive effects in embryonic development, including neurodevelopment while a mild *SRY* activation could be responsible for dimorphisms observed between the sexes.

3135S

Autophagy retards inflammatory mRNA decay and elicits a white phenotype during adipocyte maturation. *J. Shan, A. Worschech, R. Thomas, L. Chouchane.* Laboratory of Genetic Medicine and Immunology, Weill Cornell Medical College in Qatar - Qatar Foundation, Doha, Qatar.

Background: Recently, the role of autophagy in glucose and lipid metabolism has been emerging. Mice experiments showed that autophagy deficiency could prevent diet-induced obesity, characterized by less fat and a browning phenotype of white adipocyte (WAT). However, the underlying molecular mechanism is not well explored and the data from human are limited. Method: The mRNA sequencing data of undifferentiated and differentiated human adipocyte cell lines, including two white adipocyte (WAT) and one brown adipocyte (BAT) were included in our analysis. Gene expression was reduced by RNA interference in human adipocyte and was enhanced by glucocorticoid, respectively. Qualifying the lipid droplet content and quantifying the adipolysis and differentiation marker expression were applied to evaluate WAT differentiation. LC3 was used as a marker to examine autophagy function of adipocyte. Result: We found a remarkable feature of adipocyte differentiation that inflammation signaling was significantly strengthened during WAT maturation, but not during BAT maturation. The alteration of Zinc Finger protein 36 (ZFP36), which mediates the decay of mRNA transcripts of inflammation molecules, obviously affected the phenotype of mature WAT: silencing of ZFP36 gene resulted in a more whitening phenotype and induction of ZFP36 resulted in a browning phenotype. ZFP36 activity was associated with p38 MAPK signaling that was regulated by autophagy. Conclusion: ZFP36 links autophagy to the determination of mature adipocyte phenotype. Therefore, ZFP36 is a potential target to prevent obesity and improve glucose and lipid metabolism.

3136S

RNA-Seq to identify novel markers for neural tissue differentiation. *Y. Sun[1], K. Giorda[1], E. Frey[2], M. Taylor[1], T. Barron[2], C. Davidson[1], D. Piper[2], G. Meredith[1].* 1) Thermo Fisher Scientific, South San Francisco, CA; 2) Thermo Fisher Scientific, Madison, WI.

Neural tissue differentiated and cultured from patient-derived stem cells is expected to revolutionize treatment of patients with brain and spinal injuries and diseases. Critical for these cellular therapies is accurate control and monitoring of differentiation but current methods for such cell typing are limited to qPCR and immunocytochemisty (ICC) which is not sufficient to discriminate between the numerous (likely >100,000) possible neural cell-types. RNA-Seq profiling using next-generation sequencing systems permits characterization and discovery of much-needed novel markers. RNA was isolated over a time course from human embryonic stem cells (H9) and induced neural stem cells (NSCs) in triplicate. ICC was performed on the putative NSC pools at d7 and d14 for markers of pluripotency (Oct4) and neural differentiation (nestin, Sox1, and Pax6) and H9 cells were stained on d14 for markers of pluripotency (Oct4 and SSEA4). Ion Torrent semiconductor sequencing libraries were created to profile expression of miRNAs and whole transcriptomes for each of the 15 cell populations. We generated ≥1.5 million small RNA reads and ≥29 million whole transcriptome reads per sample. Cluster analysis of the RNA-Seq profiles indicates that the cell populations have characteristic molecular signatures. Among genes that are decreased in induced cells are OCT4 (POU5F1), JARID2, NANOG , consistent with the differentiation of iPSCs into neurons. Among genes that showed increased expressions are NTRK2, POU3F2, and a number of HOX family genes. We also find lincRNA are involved in cell differentiation. For Research Use Only. Not for use in diagnostic procedures.

3137S

A conserved role for IRF6 in neurulation. *Y.A. Kousa[1], H. Zhu[2], W. Fakhouri[3], Y. Lei[2], A. Kinoshita[4], R.R. Roushangar[1], E.J. Leslie[5], T.D. Busch[6], T.J. Williams[7], Y. Chai[8], B.A. Amendt[9], J.C. Murray[6], G.M. Shaw[10], A.G. Bassuk[6], A. Ashley-Koch[11], S. Gregory[11], R.H. Finnell[2], B.C. Schutte[12].* 1) Biochemistry and Molecular Biology Department, Michigan State University, 48824 East Lansing, Michigan, USA; 2) Dell Pediatric Research Institute, Department of Nutritional Sciences, University of Texas at Austin, 78723 Austin, Texas, USA; 3) Department of Diagnostic & Biomedical Sciences, School of Dentistry, University of Texas at Houston, 77054 Houston, Texas, USA; 4) Department of Human Genetics, Nagasaki University, Nagasaki, Japan; 5) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh; 6) Department of Pediatrics, University of Iowa, 52242 Iowa City, Iowa, USA; 7) Department of Craniofacial Biology, University of Colorado Denver at Anschutz Medical Campus, 80045 Aurora, Colorado, USA; 8) Center for Craniofacial Molecular Biology, Ostrow School of Dentistry, University of Southern California, 90033 Los Angeles, California, USA; 9) Department of Anatomy and Cell Biology, University of Iowa, 52242 Iowa City, Iowa, USA; 10) Department of Pediatrics, Stanford University School of Medicine, 94305 Stanford, California, USA; 11) Duke Molecular Physiology Institute, Department of Medicine and Molecular Genetics and Microbiology, Duke University, 27701 Durham, NC, USA; 12) Department of Microbiology and Molecular Genetics, Michigan State University, 48824 East Lansing, Michigan, USA.

IRF6, TFAP2A and GRHL3 are part of a gene regulatory network that is required for development of the lip and palate in humans and mice. However, in mice, Tfap2a and Grhl3 are also required for neurulation. To test the hypothesis that Irf6 is also required for neurulation, we evaluated an allelic series of Irf6 mutant mice for neural tube defects. We found that overexpressing Irf6 led to anencephaly by directly repressing Tfap2a transcription. Reducing Irf6 expression led to the mouse equivalent of spina bifida by regulating Tfap2a and Grhl3. In addition, we observed Irf6 expression in neural and non-neural ectoderm and early migrating neural crest cells. Moreover, we observed that MCS9.7, an enhancer for IRF6, was active in pre-migratory neural crest during neurulation. Considering these data in mice, we asked if rare or common IRF6 variants are associated with human spina bifida. To find rare variants, we sequenced all IRF6 exons in 92 individuals with spina bifida. In one individual, we identified a rare nonsynonymous substitution previously reported in dominantly inherited orofacial clefting. In the 3'UTR, we identified a conserved region that regulated Irf6 stability in a cell culture assay. Within this element, we found a common SNP (rs17317411 T/C) whose derived allele (C) is predicted to create a new miRNA-binding site. Genotyping in 735 trios with spina bifida showed a significant association between the derived allele of rs17317411 and lipomyelomeningocele (N = 68 trios; p value = 0.04), but was not statistically significant for other types of spina bifida. To date, we have also sequenced MCS9.7 in 158 cases and 171 controls. We observed a significant association between the derived allele of rs76145088 and spina bifida (p=0.016). Interestingly, the derived allele at rs76145088 is predicted to abrogate a highly conserved cis-regulatory motif for the TEAD/TEF family of transcription factors. The TEAD family members mediate Hippo signaling in the neural plate border and pre-migratory neural crest cells and regulate neural progeniter cell number. In sum, we identified a conserved role for IRF6 in neurulation between mouse and humans that is in addition to its role in orofacial clefting. This is a rare example of two common variants, located in the same enhancer element, that contribute risk for more than one common, complex human disease.

3138S

Analysis of *CAPZB* function in cleft pathogenesis and lower jaw extension. *K. Mukherjee[1,4], M. Grimaldi[1], M. Talkowski[2,4], J. Gusella[2,4], R. Maas[3,4], C. Morton[3,4], E. Liao[1,4].* 1) Center for Regenerative Medicine, Massachusetts General Hospital; 2) Center for Human Genetic Research, Massachusetts General Hospital; 3) Brigham and Women's Hospital; 4) Harvard Medical School, Boston, MA.

Purpose: Orofacial clefts are among the most common congenital birth anomalies worldwide. Clefts and other craniofacial anomalies arise due to defects in craniofacial morphogenesis. In an ongoing gene discovery effort, the Developmental Genome Anatomy Project (DGAP) has developed whole genome sequencing strategies to characterize genes contributing to such human congenital anomalies. Through DGAP, we discovered a new candidate gene for cleft palate (CP), *CAPZB* that encodes an actin-capping protein. The isolated disruption of *CAPZB* was identified in a 6-month old female presenting cleft palate, micrognathia and hypotonia. We have exploited the zebrafish model to determine the function of *capzb* in craniofacial morphogenesis. **Methods:** The zebrafish embryonic Meckel's cartilage and ethmoid plate are analogous to the mammalian mandible and primary palate respectively, making it an ideal model system to study palate and lower jaw morphogenesis. The spatiotemporal gene expression of *capzb* was determined by whole mount *in situ* hybridization (WISH) during embryogenesis. Craniofacial cartilaginous structures and muscles were examined in the *capzb* mutant, identified from an insertional mutagenesis screen. **Results:** WISH analysis shows that *capzb* is ubiquitously expressed, demonstrating its potential requirement in the function of many tissue types. Moreover, preliminary analysis of the *capzb* mutants show that the lower jaw elements are smaller and retrusive and the palate is only partially fused with a cleft in the anterior palate. The actin cytoskeleton is in disarray without capzb, with loss of cell morphology in the palate chondrocytes and highly disorganized myofibrils, leading to atrophied muscles. **Conclusion:** We have identified *CAPZB* to be important in craniofacial and muscle morphogenesis, disruption of which is pathologic for palate development. Preliminary results from characterization of the *capzb* mutants suggest that *capzb* plays a role in palate fusion and lower jaw extension, and affects tissue types where actin organization is critical to cell morphology and higher order organ morphogenesis. Experiments are under way to delineate the mechanisms of *capzb* function in craniofacial morphogenesis and in regulation of fundamental cellular mechanisms where actin dynamics and cell signaling pathways intersect. This study illustrates how clinically based studies can uncover fundamental mechanisms that govern cell biology and tissue morphogenesis.

3139S

A novel transcriptional regulatory pathway in cardiac and skeletal muscle. *J. Bharj[1], D. Osborn[1], D. Zheng[1], M. Uysaloglu[1], J. Ross[2], F. Conti[2], Y. Jamshidi[1].* 1) Cardiogenetics Lab, St George's University of London, London, London, United Kingdom; 2) Dubowitz Neuromuscular Centre, Institute of Child Health, London, United Kingdom.

Gene expression can be regulated at the level of transcript elongation, when RNA polymerase pauses along the partially transcribed mRNA. Following the recruitment of transcription elongation factors (TEFs), RNA polymerase resumes transciption allowing for gene expression. Our aim was to determine the role of trancriptional elongation in the development of cardiac and skeletal muscle. Using zebrafish as a model organism, we confirm the expression of a cardiac and skeletal muscle specific elongation factor in these tissues. Morpholino-mediated knockdown of of this TEF leads to severe defects in development of the heart and somites, including aberrant looping, reduced circulation and detachment of somites from myosepta. Interestingly, these defects resemble those seen in zebrafish models of cardiomyopathy and muscular dystrophy, suggesting that defects in transciptional elongation may underlie these conditions. Analysis of gene expression shows deregulation of a large number of genes, including *titin* and *MEF2,* which are involved in heart and somite development. Strikingly, no known proteins involved in somite attachment appeared deregulated. This suggests either that TEFs regulate the expression of yet unkown effectors, or that the detachment of fibres is caused by small changes in the expression of a number of proteins. Our zebrafish model will be useful to further study transcriptional elongation, and may also lead to the identification of viable drug targets for cardiomyopathies and muscular dystrophies.

3140S

The Role of *RERE* in Cardiovascular Development. *H.P. Zaveri[1], B.J. Kim[1], A. Hernandez-Garcia[1], D.A. Scott[1,2].* 1) Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX.

Deletions of chromosome 1p36 affect 1:5000 newborns. Over 70% of these infants have cardiovascular malformations. The arginine-glutamic acid dipeptide (RE) repeats gene (*RERE*) is located in one of five critical regions for cardiovascular malformations we have delineated on chromosome 1p36. *RERE* encodes a nuclear receptor coregulator that positively regulates retinoic acid signaling. In some tissues, this regulation involves the formation of a transcriptional complex that includes RERE and NR2F2, a nuclear receptor implicated in the development of cardiovascular malformations. Since perturbations in retinoic acid signaling can cause cardiovascular malformations, we hypothesized that RERE deficiency contributes to the development of cardiovascular malformations in children with 1p36 deletions. To test this hypothesis, we generated an allelic series of RERE-deficient mice using an *Rere* null-allele and an *Rere* hypomorphic allele (*eyes3*). *Rere*$^{-/-}$ embryos die of cardiac failure around E9.5 with enlarged midline hearts that have failed to undergo rightward looping. *Rere*$^{-/eyes3}$ embryos escape early embryonic lethality but have cardiovascular malformations—outflow tract, septal and valve abnormalities—that are similar to those seen in individuals with 1p36 deletions. Using these mouse models, we identified *in vivo* genetic interactions between *Rere* and *Raldh2*—which encodes the main enzyme that catalyzes retinoic acid production in the heart—and *Gata4*—which encodes a retinoic acid-responsive transcription factor that causes outflow tract and septal defects in humans. Further experiments revealed that the cardiac expression patterns of RERE and NR2F2 overlap in the atrioventricular canal and that *Nr2f2* interacts genetically with *Gata4* during septal development. We also found that manipulation of *in utero* retinoic acid levels was insufficient to rescue RERE-related cardiovascular malformations. These results suggest that RERE and NR2F2 may form a complex during heart development that acts downstream of RALDH2 to modulate the transcription of retinoic acid responsive genes—like *GATA4*—that play a critical role in the development of atrioventricular septum and its associated valves. We conclude that *RERE* contributes not only to the development of cardiovascular malformation caused by 1p36 deletions, but may also modulate the penetrance of cardiovascular malformations caused by mutations in *GATA4* and *NR2F2* and deletions of 8p23.1 and 15q26.2 where these genes reside.

3141S

Characterization of a knock-in mouse model expressing the Stormorken syndrome mutation. *T.H. Gamage[1], S.S. Amundsen[1], G. Gunnes[2], W.E. Louch[3], T.M. Pedersen[4], P.A. Holme[5,6], A. Holmgren[1], G.E. Tjønnfjord[5,6], A. Klungland[7], D. Misceo[1], E. Frengen[1].* 1) Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway; 2) Faculty of veterinary medicine and biosciences, Norwegian University of Life Sciences, Norway; 3) Institute for Experimental Medical Research, Oslo University Hospital and University of Oslo, Norway; 4) Research Institute of Internal Medicine, Oslo University Hospital and University of Oslo, Norway; 5) Department of Hematology, Oslo University Hospital, Norway; 6) Institute of Clinical Medicine, University of Oslo, Norway; 7) Department of Molecular Medicine, Oslo University Hospital, Norway.

The *STIM1* [MIM *605921] R304W mutation was recently identified as the cause of Stormorken syndrome [MIM 185070] in patients from 8 independent families. STIM1 encodes the stromal interacting molecule 1, which is required for activation of the calcium release-activated calcium (CRAC) channel in the plasma membrane of most cell types. Previously, patients with complete STIM1 deficiency were described with combined immunodeficiency (CID), and patients with heterozygous missense mutations in the *STIM1* EF-hand present with tubular aggregate myopathy (TAM [MIM #160565]). Stormorken syndrome patients, however, present with a muscle phenotype (spasms, muscle weakness, lack of endurance capacity and TAM) as well as hematological defects (mild bleeding tendency, mild anemia, thrombocytopathy, thrombocytopenia), miosis, ichthyosis and asplenia. This wide range of clinical features indicates the importance of the *STIM1* function in several organs.

Our aim is to study the functional effects of the *Stim1* R304W mutation in a knock-in mouse model. We have established the knock-in mouse model on a CBA c57bl/6 background using Zinc finger technology. We compare the hematological aspects of knock-in vs wild-type mice to assess the functional consequences of the mutation. General behavior of mice is being observed and endurance testing is being performed to study the muscle phenotype. We are also presently performing a complete pathological and histological examination of the mouse. We expect that studying the biological consequences of the *STIM1* R304W mutation will provide new insight into STIM1 function, and reveal its role in the etiology and pathogenesis of Stormorken syndrome.

3142S

Haploinsufficiency of *RERE* contributes to the development of cleft palate in 1p36 deletion syndrome. *B. Kim[1], H.P. Zaveri[1], V. Jordan[2], D.A. Scott[1,2].* 1) Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX.

Orofacial clefts—cleft palate and cleft lip with or without cleft palate—affect 1.7:1000 newborns and can lead to long-lasting adverse outcomes for health and social integration. Deletions of chromosome 1p36 are the most common telomeric deletions in humans and carry a high risk of orofacial clefts. The arginine-glutamic acid dipeptide (RE) repeats gene (*RERE*) is located in one of three non-overlapping critical regions for orofacial clefts we have defined on chromosome 1p36. *RERE* encodes a nuclear receptor coregulator that positively regulates vitamin A/retinoic acid signaling in the developing embryo. Since cleft palate can be caused by abnormalities in retinoic acid signaling, we hypothesized that haploinsufficiency of *RERE* may contribute to the development of cleft palate in children with 1p36 deletions. To test this hypothesis, we generated an allelic series of RERE-deficient mice using an *Rere* null-allele and an *Rere* hypomorphic allele (*eyes3*) identified in our laboratory. On a C57BL/6 background, cleft palate was seen in 80% (4/5) of $Rere^{-/eyes3}$ embryos at E15.5. When we ablated *Rere* in neural crest cells using a transgenic Wnt1-Cre, we found that $Rere^{flox/flox}$;Wnt1-Cre embryos and mice also developed cleft palates in which the opposing palatal shelves elevated into the normal horizontal position but failed to make contact in the midline. Further studies revealed that RERE is expressed in the bend region of the developing palate; a region where retinoic acid signaling controls cell proliferation and the expression of key palatal genes including *TBX1* that causes cleft palate associated with 22q11.2 microdeletions. We conclude that *RERE* plays a critical, cell-autonomous role in the neural crest cells during palatal development and that deletion of *RERE* contributes to the development of cleft palate in children with 1p36 deletions.

3143S

An animal model to investigate genetic variants in patients with 46,XY Disorders of Sex Development. *H. Barseghyan, V. Arboleda, R. Baxter, A. Eskin, S. Nelson, E. Délot, E. Vilain.* Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California.

Disorders of Sex Development (DSD) are defined as congenital conditions in which development of chromosomal, gonadal, or anatomic sex is atypical. These conditions have a frequency of 0.5-1% of live births and encompass a wide variety of urogenital abnormalities ranging from mild hypospadias to sex reversal. At present, a specific molecular diagnosis is identified in only a minority of DSD cases. We used whole exome sequencing (WES) with gene list approach specific to DSD to identify the underlying genetic etiology for 46,XY DSD cases with a range of phenotypes. In 35% of these cases, exome sequencing identified mutations in genes, classified as pathogenic or likely pathogenic, that could explain the given phenotype. In the unexplained 65% of the cases, we identified a large number of variants of uncertain clinical significance (VUS). In order to investigate these variants, we utilized a powerful mouse model for studying undervirilization in 46,XY individuals. In this model, the presence of a Y chromosome originating from an M. domesticus strain (YPOS) in C57BL/6J (B6) background results in XY sex reversal and severe hypovirilization. We hypothesized that abnormal gonadal expression of specific genes in the B6-YPOS mouse model will correlate with VUS in genes of 46,XY DSD patients identified through WES. Thus, we isolated gonadal tissue at embryonic day 11.5 and performed RNA sequencing in order to assess the differential gene expression levels in the bipotential gonads of wild type (WT) B6 and undervirilized B6-YPOS males. The obtained differential expression data from the mouse model was analyzed in conjunction with WES data. We identified 12 genes in which the differential expression between WT and B6-YPOS males was significantly lower in B6-YPOS gonads, and in which a missense variant with an alternate allele frequency of less than 1% was identified in 46,XY DSD cases. Among these genes, four (FBLN2, SMTNL2, ADAMTS16, COL9A2) were predicted to be damaging by in silico tools. Additional studies and validation techniques are necessary to investigate the role of these new candidate genes for 46,XY DSD.

3144S

An Allelic Series Reveals Novel Roles of Fgf Ligands in Skeletogenesis. *I.H. Hung[1,2], G.C. Schoenwolf[1], M. Lewandoski[2], D.M. Ornitz[3].* 1) University of Utah, Salt Lake City, UT; 2) National Cancer Institute, Frederick, MD; 3) Washington University School of Medicine, St. Louis, MO.

Chondrodysplasia and craniosynostosis syndromes are caused by activating mutations in fibroblast growth factor (FGF) receptors. Although the roles of the FGFRs in bone development have been relatively well-characterized, only two FGF ligands, FGF9 and FGF18, have been identified to regulate embryonic skeletogenesis. We have generated an Fgf allelic series to further elucidate the functions of these growth factors during skeletal development. Mice lacking both Fgf9 and Fgf18 exhibit severe defects in endochondral and intramembranous ossification, demonstrating partial functional redundancy between these ligands. We provide evidence that FGF signaling is required at early developmental stages to prevent an arrest in chondrocyte differentiation of the proximal limb and to promote calvarial osteogenesis .

3145S

Phenotypic and functional characterization of $Bst^{+/-}$ mouse retina. *G. Sun[1,5], H. Riazifar[2,5], X. Wang[1], F.N. Ross-Cisneros[3], V. Carelli[4], A.A. Sadun[3], T. Huang[1,2].* 1) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Department of Pediatrics, Division of Human Genetics; University of California, Irvine, CA; 3) Doheny Eye Institute, Department of Ophthalmology, University of Southern California, Los Angeles, CA; 4) Dipartimento di Science Neurologiche, Universita di Bologna, Bologna, Italy; 5) Equal Contributor.

Current animal models for subretinal neovascularization (SRN) depend on using chemistry or physics methods to stimulate subretinal vessels grow. The animal model for retinal degenerative disease is also very limited. Genetic animal model for retinal disease is needed. The Belly spot and tail (Bst) mouse phenotype is caused by mutations of the ribosomal protein L24. $Bst^{+/-}$ mouse has striking ocular phenotypes, with the feature of delayed closure of choroid fissures, decreased ganglion cells, and subretinal vascularization. In seeking mouse model for stem cell therapy for retinal degenerative disease due to retinal ganglion cell (RGC) loss, we further characterized the $Bst^{+/-}$ mice and investigated the underlying molecular mechanisms. We found that although RGC was significantly reduced in retinal ganglion cell layer in $Bst^{+/-}$ mouse, melanopsin-positive RGC, also called ipRGCs seemed preserved. Pupillary light reflex is complete absent in $Bst^{+/-}$ mouse, while circadian rhythm is normal. In order to examine the pathological abnormalities in $Bst^{+/-}$ mice, we performed electronic microscope test and found mitochondria morphology was deformed with irregular borders and lacking cristae. The complex activities of mitochondrial electron transport chain were decreased significantly. Finally, for subretinal vascularization, the delay of angiogenesis was observed in $Bst^{+/-}$ mice associated with delayed haloid regression. Our characterization of $Bst^{+/-}$ mouse retina suggests that $Bst^{+/-}$ mouse could be a useful model for stem cell therapy.

3146S

Osteoblast development is driven by trans-acting regulations. *K. Choi[1], K. Shultz[1], D. Godfrey[1], M. Hibbs[2], C. Ackert-Bicknell[1], G. Churchill[1].* 1) The Jackson Laboratory, Bar Harbor, ME; 2) Trinity University, San Antonio, TX.

Understanding the genetic and molecular bases of normal osteoblast development is a key to new anabolic treatments for osteoporosis. To investigate development process of osteoblast, we generated time course RNA-seq data (Illumina HiSeq 2000) covering 9 time points from pre-osteoblast stage to mature osteoblast. The similar sets of data were derived from three different genetic backgrounds of C57BL/6J (B6) inbred, B6xCAST/EiJ, and B6x129S1/SvImJ hybrid mice.

We found that over 5,000 genes have either monotonically increasing or decreasing temporal expression patterns using Weighted Gene Co-expression Network Analysis tool version 1.34. For each strain, we ordered these monotone-patterned genes *temporally* with respect to their maximal change time by fitting to a four-parameter logistic growth model. We observed the temporal order of maximal expression change is largely preserved across the three mouse strains. We applied these time order to generate causality hypotheses in the MouseMap, a publicly available tissue-specific functional networks. For example, MouseMap provides the neighboring genes of Col3a1 but we do not know whether Col3a1 is triggering its neighboring genes or any of its neighbors is causing it to transcribe. But when we introduce temporal orders to the subnetwork, we can see Col3a1 is a requirement for its neighboring genes to function. Volk and colleagues reported that Col3$^{-/-}$ mice showed a significantly reduced mineralization capacity. We have found hundreds of interesting causal relationships, many of which are novel.

Time course transcriptome analysis of F1 hybrid coupled with its founder inbred strains is a great resource for identifying modes of gene regulation: whether a certain gene is regulated by cis- and/or trans-acting elements. Suppose a gene's total expression in F1 hybrid has different temporal pattern from those of its founder strains. If the two alleles are expressing similarly in the F1, then the gene has to be under trans-regulation since we know (1) local structural variations in those alleles do not influence their expression levels but (2) what is different between F1 and its founder strains is the trans-acting environment regulating this gene's transcription. We identified over 1,000 genes that are in trans-regulation in the process of osteoblast development using a statistically conservative filtering process.

3147S

The Role of SOX7 and SOX17 in Cardiovascular Development. *A. Hernandez-Garcia[1], M. Wat[1], R. Udan[2], A. Renwick[1], M. Garcia[2], Z. Yu[1], C.A. Shaw[1], M. Dickinson[2], D.A. Scott[1,2].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX.

SOX7 is located on chromosome 8p23.1—a region that is commonly deleted in individuals with complex cardiovascular malformations. *SOX7* encodes a transcription factor that is highly expressed in the endocardium and in endothelial cells but its role in these cells remains unclear. To determine if SOX7 plays a critical role in cardiovascular development, we targeted the second exon of *Sox7* to create standard and conditional *Sox7* knockout mice. Although *Sox7*$^{+/-}$ mice were viable and fertile, *Sox7*$^{-/-}$ embryos died around E10.5 with signs of cardiac failure including pericardial edema and failure of yolk sac vasculature remodeling. The same phenotype was observed when *Sox7* was ablated in endothelial cells using a Tie2-Cre. Rare *Sox7*$^{-/flox}$;Tie2-Cre embryos that were recovered at E15.5 had pericardial effusions, vascular hemorrhage, ventricular septal defects, aberrant liver histology and evidence of developmental arrest. Using these mouse models, we also found that *Sox7* interacts genetically with *Sox17*—a known regulator of vascular sub-type specification—during vascular development with *Sox7*$^{+/-}$;*Sox17*$^{+/-}$ embryos dying around E10 with signs of cardiac failure. At E8.5—a time point prior to the onset of cardiac failure—*Sox7*$^{-/-}$ and *Sox7*$^{+/-}$;*Sox17*$^{+/-}$ embryos had abnormally high levels of endothelial cell proliferation when compared to wild-type littermate controls. Transcriptome and gene expression analyses carried out at E8.5 and E9.5, revealed several genes involved in cardiac, liver and vascular development—including *Tbx2* and *Vegfa*—that were dysregulated in the absence of SOX7. These results suggest that *Sox7* interacts genetically with *Sox17* during vascular development and that *Sox7* acts in a cell-autonomous fashion to modulate the transcription of genes that regulate endothelial cell proliferation. SOX7-deficiency leads to abnormal vascular development and may contribute to the development of cardiovascular malformations associated with recurrent 8p23.1 microdeletions.

3148S

Unraveling the genetic architecture of anencephaly: Identification and analysis of coding variants in *Cecr2* and candidate modifier genes of *Cecr2* in mice and humans. *R.Y.M. Leduc[1], D.R. Krupp[2], A.E. Ashley-Koch[3], S.G. Gregory[2], N. Mola[3], E.E. Davis[3], N. Katsanis[3], H.E. McDermid[1].* 1) Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada; 2) Duke Molecular Physiology Institute, Duke University, Durham, North Carolina, USA; 3) Center for Human Disease Modeling, Duke University, Durham, North Carolina, USA.

Anencephaly, the failure of the cranial portion of the neural tube to close during early embryonic development, is an invariably lethal birth defect in humans. To investigate the mechanistic underpinnings of neural tube closure, we have examined genetic variants in mice and humans. In mice, a homozygous loss-of-function mutation of the gene *Cecr2* causes exencephaly (equivalent to the human anencephaly phenotype) in a strain-dependent manner. Approximately 54% of BALB/c *Cecr2* mutant mice develop exencephaly compared to 0% of FVB/N *Cecr2* mutant mice. This suggests the presence of DNA sequence variants that modify the penetrance of the exencephaly phenotype. Linkage analysis previously identified a modifier region on mouse chromosome 19. Microarray analysis was performed to identify genes differentially expressed between BALB/c and FVB/N within this region, and expression differences for prioritized candidate genes were confirmed with qRT-PCR. To detect possible functional variants, whole exome sequencing (WES) was also performed on Balb/c and FVB/N, and prioritized candidate genes were confirmed with Sanger sequencing. A combination of microarray and WES analyses yielded a list of 24 candidate modifier genes. It is not known whether *Cecr2* or any of the 24 candidate genes identified in mice are involved in the etiology of anencephaly in humans. Therefore, we performed custom capture sequencing of the coding regions of the human homologues of *Cecr2* and the 24 candidate modifier genes in a human anencephaly cohort of 157 probands. Fifty coding variants of interest (minor allele frequency (MAF) ≤ 3) were identified in 18 genes, including *CECR2*. All fifty variants were Sanger sequence verified in the probands and in available parental DNA samples. As the mutation in *Cecr2* is homozygous recessive in mice, we were also interested in looking for compound heterozygosity in the probands. To do this, nonsynonymous variants with a European MAF > 0.03 were also Sanger sequenced in proband and parental DNA samples if the proband contained more than one coding variant in the same gene. Six probands were identified as compound heterozygotes, three of which are in *CECR2*. These data demonstrate the utility of comparative genomics in dissecting the mechanisms of neural tube closure. Future functional assays will be necessary to determine causality of the identified variants.

3149S

Rapid identification of kidney cyst mutations by whole exome sequencing in zebrafish. *J.R. Willer[1,2], S. Ryan[2], L. Marjoram[2], J. Bagwell[2], J. Mankiewicz[2], I. Leshchiner[3], W. Goessling[3,4,5], M. Bagnat[2], N. Katsanis[1,2].* 1) Center for Human Disease Modeling, Duke University, Durham, NC; 2) Department of Cell Biology, Duke University; 3) Genetics Division, Brigham and Women's Hospital, Harvard Medical School, Boston; 4) Genetics and Gastroenterology Divisions, Brigham and Women's Hospital, Harvard Medical School, Boston; 5) Harvard Stem Cell Institute, Cambridge.

Forward genetic approaches in zebrafish have provided invaluable information about developmental processes. However, the relative difficulty of mapping and isolating mutations has limited the number of new genetic screens. Recent improvements in the annotation of the zebrafish genome coupled to a reduction in sequencing costs prompted the development of whole genome and RNA sequencing approaches for gene discovery. Here we describe a whole exome sequencing (WES) approach that allows rapid and costeffective identification of mutations. We used our WES methodology to isolate four mutations that cause kidney cysts; we identified novel alleles in two ciliary genes as well as two novel mutants. The WES approach described here does not require specialized infrastructure or training and is therefore widely accessible. This methodology should thus help facilitate genetic screens and expedite the identification of mutants that can inform basic biological processes and the causality of genetic disorders in humans.

3150S

A *Tgds* mutation in a novel mouse model of Pierre-Robin sequence-type clefting. *B.C. Bjork[1], I. Saadi[2], L. Pitstick[1], S. Docksey[1], D.R. Beier[3].* 1) Department of Biochemistry, Midwestern University, Downers Grove, IL; 2) Department of Anatomy and Cell Biology, University of Kansas Medical Center, Kansas City, KS; 3) Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA.

Non-syndromic orofacial clefting is one of the most common human birth defects with both multigenic and environmental influences that contribute to its etiology. Pierre Robin sequence (PRS)-type clefting is a "palate-extrinsic" mechanism of cleft secondary palate (CP) that results as an indirect consequence of an abnormally small mandible and ankyloglossia. We identified the *little chin (lc)* mutant in a three-generation ENU-mutagenesis screen designed to identify autosomal recessive mouse mutants that model human congenital birth defects. *lc* mutants die shortly after birth due to respiratory distress and PRS-type CP, as evidenced by mandibular hypoplasia. *lc* mutant palate shelves fail to elevate above the tongue. Mutants also exhibit a shortened snout and a bifid xyphoid process. We initially mapped the *lc* mutation to a 1.1 Mb critical region on chromosome 14 by positional cloning. The region contains six genes, at least two of which, *Sox21* and *Gpc6*, represented attractive candidates for involvement in craniofacial development. However, initial sequencing of coding exons and highly conserved genomic sequences within the critical region failed to reveal any causative mutation.

Recently, we performed Next Generation whole-exome sequencing of genomic DNA from *lc* and *wild type* embryonic tissues and identified a homozygous point mutation in the *Tgds* gene that encodes the dTDP-glucose 4, 6-dehydratase enzyme. The C to T missense mutation creates a non-conservative Alanine to Valine substitution at amino acid 26 (A26V) of TGDS that is predicted to be damaging via PolyPhen. This is the only homozygous damaging sequence variant identified from this study. The exact function of TGDS and its relationship to craniofacial development is unknown. It has primarily been studied for its role in in cell wall integrity in plant species and Candida albicans. TGDS also plays an important role in NAD+/NADH shuttling, and the A26V mutation in *lc* mutants alters its highly conserved NAD binding motif (GGAGFIG). We are validating the mechanism of *Tgds* mutation in the etiology of clefting in *lc* mutants. Impaired TGDS function in *lc* mutants represents a novel mechanism to be explored as a contributing factor in the incidence of PRS-type clefting in humans.

3151S

The transcriptional co-regulator Jab1 is required for skeletogenesis. *G. Zhou, L. Bashur, Z. Chen, S. Murakami.* Orthopaedics, Case Western Reserve University, Cleveland, OH.

Skeletogenesis is a finely-tuned process governed by a spatiotemporal-specific transcriptional circuit. RUNX2 is a master transcription factor required for BMP signaling-mediated chondrocyte maturation and osteoblast differentiation. RUNX2 haploinsufficiency results in cleidocranial dysplasia, while mutations in various BMP signaling components lead to a spectrum of human skeletal developmental defects. Through analyzing a novel mouse model of lethal chondrodysplasia (Jab1flox/flox; Col2a1-Cre), we have recently shown that Jab1, a highly conserved transcriptional co-regulator, is a novel factor crucial for chondrocyte maturation, likely by inhibiting Runx2 and BMP signaling. To further dissect the role of Jab1 at successive steps of osteogenesis, we first generated Jab1flox/flox; Prx1-Cre conditional knockout (Prx1-Cre/cKO) mutant mice in which Jab1 was deleted in the osteochondral progenitor cells of the limb buds. At E18.5, Prx1-Cre/cKO mutants exhibited significantly shorter limbs with: very few hypertrophic chondrocytes, much smaller primary ossification centers with reduced von Kossa staining, decreased expression of Runx2 and Col1a1, and reduced BMP signaling response. Thus, Jab1 is required for early osteoblast differentiation of osteochondral progenitor cells. Next, we generated Jab1flox/flox; Osx-Cre conditional knockout (Osx-Cre/cKO) mutant mice in which Jab1 was deleted in the osteoblast precursor cells. Osx-Cre/cKO mutants appeared grossly normal in size at birth. However, histological study revealed reduced trabecular thickness and fewer trabecular areas in newborn Osx-Cre/cKO mutant long bones. Osx-Cre/cKO mutant long bones grew less after birth and all the mutants died before weaning age. Thus, Jab1 is likely to be required for the proliferation and function of osteoblast precursor cells. Finally, to specifically target differentiating osteoblasts, we generated Jab1-flox/flox; Col1a1-Cre conditional knockout (Col1a1-Cre/cKO) mutant mice, using a 2.3 Col1a1-Cre driver. Col1a1-Cre/cKO mice were viable, fertile, and showed no gross phenotype until 12-month age, indicating that Jab1 plays a less prominent role at later stage of osteoblast differentiation. We are further characterizing these novel Jab1 mutant models to study the stage-specific effect of Jab1 on Runx2 and BMP signaling during skeletogenesis. Our study will provide novel insights into the pathogenesis of skeletal disorders caused by dysregulated RUNX2 and BMP signaling.

3152S

SPECC1L modulation of adherens junctions and PI3K-AKT signaling is required for collective cell migration in facial morphogenesis. *N. Wilson[1], A. Olm-Shipman[1], E. Kosa[1], D. Acevedo[1], K. Stumpff[1], G. Smith[1], L. Pitstick[2], B. Bjork[2], A. Czirok[1], I. Saadi[1].* 1) Department of Anatomy and Cell Biology, University of Kansas Medical Center, Kansas City, KS; 2) Department of Biochemistry, Midwestern University, Downers Grove, IL.

Orofacial clefts are among the most frequent birth defects, with cleft lip and palate (CL/P) alone affecting 1/800 live-births. While a number of contributory genes have been identified, there is continued need to understand underlying pathogenetic mechanisms. Previously, we identified *SPECC1L* as the first gene mutated in a severe cleft that extends from the oral cavity to the eye, termed Oblique Facial Cleft (ObFC). We proposed that mechanisms underlying ObFC and SPECC1L function are directly related to common CL/P malformations. Indeed, new heterozygous missense *SPECC1L* mutations have been identified in two patients with nonsyndromic CL/P (Dr. Murray, U. Iowa) and in two families with hypertelorism and bilateral CL/P (Dr. Zackai, CHOP), underscoring the significance of this novel cytoskeletal protein. Thus, we have created severe and hypomorphic mouse models of *Specc1l* deficiency. Hypomorphic homozygous mutants exhibit perinatal lethality with incompletely penetrant micrognathia, cleft palate, and subepidermal blebbing phenotypes. Severe homozygous *Specc1l* gene-trap mutants die during early embryogenesis with defects in neural tube closure and delamination/migration of neural crest cells (NCCs). Severe *Specc1l* mutant embryos show increased apoptosis and reduced AKT signaling. SPECC1L protein is stabilized at cell-cell boundaries upon confluence and interacts with both β-catenin and E-cadherin, two canonical components of adherens junctions (AJs). Further, SPECC1L-deficient cells (kd) show altered AJ staining predicted to reflect *more dense* AJs, which leads to *poor* cell migration. Electron micrographs of kd cells indicate jagged cell-cell boundaries and loss of apical-basal distinct electron-dense regions of AJs, consistent with increased AJ density. Interestingly, live-imaging analysis indicates that the migration defect is not due to slower individual speed, but rather due to *poor collective movement*. A significant proportion of kd cells show misoriented cell divisions along the long axis, suggesting a defect in cell polarity and directional cell division - two important considerations in collective migration of epithelial sheets. Modulation of cell-cell contacts is important not only for NCC delamination from the neural folds, but also for collective migration of NCCs to their defined destinations. Thus, SPECC1L is an entirely novel modulator of AJ density and of AKT signaling, affecting collective cell movement in facial morphogenesis.

3153S

Retinoic acid induced-1 (Rai1) regulates craniofacial and brain development in *Xenopus*. *R. Tahir[1], A. Kennedy[2], A.J. Dickinson[2], S.H. Elsea[3,4].* 1) Center of the Study of Biological Complexity, Virginia Commonwealth University, Richmond, VA; 2) Department of Biology, Virginia Commonwealth University, Richmond, VA; 3) Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Retinoic acid induced-1 (RAI1 [MIM 607642]) is an important yet understudied histone code reader that when mutated in humans results in Smith-Magenis syndrome (SMS [MIM 182290]), a complex neurobehavioral disorder accompanied by signature craniofacial abnormalities, including a broad square-shaped face, a flat nasal bridge, a tented upper lip, and midface hypoplasia. Despite previous studies in mouse and human cell models, very little is known about the function of RAI1 during embryonic development. Here, we have utilized *Xenopus laevis* and *Xenopus tropicalis* to better understand the developmental roles of Rai1. Protein sequence analysis shows that frog Rai1 is conserved, especially in known functional domains. By *in situ* hybridization, we revealed expression of *rai1* in the developing craniofacial tissues and the nervous system, including the developing neural crest, eye, otic vesicles, branchial arches, and brain, and that excess retinoic acid enriched *rai1* in the developing embryo. Knockdown of Rai1 using antisense morpholinos resulted in defects in the developing brain and face, and in particular, Rai1 morphants displayed midface hypoplasia and malformed mouth shape analogous to defects in humans with SMS. These craniofacial defects were accompanied by aberrant neural crest migration and reduction in the size of facial cartilage elements. Rai1 morphants also had defects in axon patterns and decreased forebrain ventricle size. Such brain defects correlated with a decrease in the neurotrophic factor, *bdnf*, and an approximately 3-fold increase in forebrain apoptosis. These results emphasize a critical role of Rai1 for normal neural and craniofacial development, and further the current understanding of potential developmental mechanisms that cause Smith-Magenis syndrome.

3154S

The chromosome 3p22.3 region is a potential novel locus for complex heart disease, anorectal malformation and intellectual disability. *G. Akler[1], RA. Denchy[2], A. Anguiano[3], S. Sharma[2], MK. Geiger[1], L. Mehta[1].* 1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) Elmhurst Hospital Center, Elmhurst, NY; 3) Quest Diagnostics Nichols Institute, San Juan Capistrano, CA.

Deletions in the 3p22.3 chromosomal region are rarely reported. Here we describe a female infant, currently 11 months old, with a de novo 1.5 Mb deletion encompassing 14 genes in the 3p22.3 region (no deletion of the 22q11 region detected). Significant phenotypic findings present included tetralogy of Fallot with absent pulmonary valve and a variant imperforate anus (rectoperineal fistula). A small extrarenal pelvis was reported in early renal sonograms. The association of these anomalies in the otherwise non dysmorphic child is suggestive of incomplete VACTERL association [MIM 192350]. Currently the patient is status-post repair of the heart defect, but has had motor delays, recurrent respiratory infections and failure to thrive prior to the repair. During embryologic development, aortopulmonary outflow tract septation, and down-growth of the urorectal septum with disintegration of the anal canal occurs between 5-8 weeks suggesting that these birth defects may result from factors disrupting development at this time. There are two previously reported interstitial deletions of this region: deletion 3p22.2-p24.2 in a patient with a small perimembranous ventricular septal defect, global developmental delay and short stature, and deletion 3p22.3p22.2 encompassing the ARPP21 [MIM 605488] and CLASP2 [MIM 605853] genes in a family with intellectual disability and dysmorphic facial features. The CLASP2 gene found in the overlapping region of all three cases may be a potential candidate gene for intellectual disability, given recent evidence of CLASP2 functioning as a key regulator of neuronal polarity and synaptic activity function. Of the other genes deleted in our patient, TRIM71 has been shown to be evolutionarily conserved and its regulation has been shown to play an important role in development. Recent literature has identified several CNV's and SNV's in patients with VACTERL/ VACTERL-like association. However, currently no consistent candidate regions or genes have emerged. This report adds to the literature on potential genetic etiologies for congenital heart defects and associated birth defects as in the VACTERL spectrum.

3155S

The sex-determining factor SRY is involved in numerous early events of testis differentiation including testis cord formation. *Y. Lau[1], Y. Li[1], M. Zheng[2].* 1) Dept Medicine/ VAMC-111C5, Univ California, San Francisco, San Francisco, CA; 2) Dept Anesthesia, Stanford Univ, Palo Alto, CA.

The Y-located sex-determining region Y (*Sry*) gene encodes the testis-determining factor, which plays a critical role in male sex determination during embryogenesis. A current hypothesis suggests that SRY collaborates with steroidogenic factor 1 (SF1) and activates a related transcription factor gene, SRY-box 9 (*Sox9*), the essential mediator for testis differentiation. SOX9, in turn, interacts with SF1 and propagates its own expression and testis differentiation. Using the ChIP-Chip strategy, we have identified large numbers of targets for both SRY and SOX9 in the fetal gonads of mouse embryos at the time of sex determination. Subsequent studies indicate that SRY or SOX9 bindings could be confirmed independently to about 90% of respective targets, which could be regulated by SRY or SOX9 in various reporter and transient gene transfection assays. SRY and SOX9 share about half of the respective targets, suggesting that these two transcription factors regulate a significant number of common target genes. Importantly, SRY, but not SOX9, binds to numerous ovarian differentiating genes and represses their activation by the WNT/β-catenin signaling. Among the SRY and SOX9 targets, numerous ones have been demonstrated to play critical roles in early events of testis differentiation. Notable ones include *Ptgds* (prostaglandin D2 synthase), critical in Sertoli cell recruitment; *Cyp26b1* (cytochrome P450, family 26, subfamily B, polypeptide 1), important for determining male lineage of the germ cells; and *Fgf9* (fibroblast growth factor 9) and *Gdnf* (glial cell derived neurotrophic factor), vital for Sertoli cell proliferation and male germ cell niche development. Gene ontology analysis identifies the Sertoli cell-Sertoli cell junction signaling, important for testis cord formation, as the top canonical pathway among the SRY and SOX9 targets. Our findings challenge the current paradigm and suggest that SRY is involved in regulating genes critical in early events of testis differentiation. Hence SRY determines the Sertoli cell fate by repressing ovarian and activating testicular differentiating genes, promotes Sertoli cell proliferation and recruitment, establishes the male germ cell lineage, and induces the early Sertoli cells to form testis cord. It then passes on its functions to SOX9, which regulates a set of common targets and activates its own gene regulatory program, beyond SRY action, in mammalian sex determination.

3156S

New insights in holoprosencephaly inheritance : Exome sequencing in families reveals new double-hit and recessive cases. *C. MOUDEN[1], C. DUBOURG[1,2], S. ROSE[1], G. VIOT[4], B. HERON[5], W. CARRE[2], M. DE TAYRAC[2,6], V. DUPE[1], S. ODENT[1,3], V. DAVID[1,2].* 1) Genetics of pathologies related to development, IGDR UMR 6290 CNRS, Rennes, France; 2) Laboratory of molecular genetics and genomics, CHU Pontchaillou, Rennes, France; 3) Service de Génétique médicale, CHU Hôspital Sud, Rennes, France; 4) Service de gynécologie-obstétrique 1 GHU Cochin - Saint Vincent de Paul, Paris, France; 5) Service de neuropédiatrie, Hôpital Jean Verdier, Bondy, France; 6) Génomique Fonctionnelle Intégrée et Biomarqueurs, IGDR, UMR6290 CNRS, Rennes, France.

Holoprosencephaly (HPE) is an anomaly of early development resulting from cleavage default of prosencephalon. It represents the most frequent cerebral malformation, with an occurrence of approximately 1 in 250 embryos and 1 in 10,000 births. About 60% of HPE cases are due to chromosomal defaults, to environmental causes like maternal alcoholism or diabetes, or are part of a polymalformative syndrome. A part of the remaining 40% of HPE cases, called genetic cases, are due to genetic alterations involving four main genes: SHH, ZIC2, SIX3 and TGIF. About 10 other minor genes are also implicated. But all these genic alterations explain only 30% of HPE genetic cases. Although these genes are formally involved, their penetrance is usually incomplete and mutations are often inherited from a healthy parent. Thus the outbreak of HPE should be due to a second event, like the implication of another gene. In this context, two high-throughput sequencing strategies (NGS) were used to identify new HPE genes. On one hand, we sequenced the whole exome in 15 families, in which known mutations were inherited from one of the two parents asymptomatic or presenting a microform of the disease, in order to point a potential inherited or de novo mutation in a new candidate gene. Analysis of one family with a known SHH mutation, revealed a second mutation in DISP1 in the HPE fetus. The screening for these two mutations in the entire family showed a co-segregation with HPE. This suggests for the first time a double hit involving SHH and DISP1. On the other hand, recessive inheritance of HPE can also be suspected in consanguineous families with intrafamilial recurrence. Homozygosity mapping was undertaken in eight families with history of consanguineous marriage to search for regions harboring mutations that are identical by descent. Exome sequencing in these consanguineous families revealed a homozygous mutation in two HPE children in a new candidate gene, involved in centrosome duplication. The KO of this gene in mouse leads to HPE phenotype. Functional tests are in progress to validate the deleterious role of this mutation in this good candidate gene. The discovery of new genes by these different strategies should permit to improve the prenatal diagnostic, and more generally the knowledge on early brain development.

3157S

Whole Exome Sequencing (WES) to Analyze the Genetic Basis of Non Syndromic Cleft Lip and Palate. *M. BASHA[1], M. QUENTRIC[1], R. HELAERS[1], N. REVENCU[2], B. BAYET[3].* 1) Laboratory of Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Brussels, Europe, Belgium; 2) Center for Human Genetics, Cliniques universitaires St Luc, Université catholique de Louvain, Brussels, Belgium; 3) Centre Labiopalatin, Division of Plastic Surgery, Cliniques universitaires St Luc, Université catholique de Louvain, Brussels, Belgium.

Cleft lip with or without cleft palate (CL/P) is the most common craniofacial birth defect with an incidence of ~ 1/700 live births, varying with ethnicity and cleft type. It is a debilitating condition requiring an expensive and lifelong treatment. We performed whole exome sequencing (WES) on 20 CL/P subjects, by using True Seq enrichment kit capture on Hiseq2000 (illumina) (38x mean coverage, 100bp 2x paired-end reads). Our pilot WES cohort predominantly consisted of familial cases (n=15) and a few isolated syndromic CL/P subjects (n=5). We WESed 1 affected subject per family. Their sub-phenotypes ranged from a full- blown bilateral CL/P to a subtler velopharyngeal insufficiency (VPI). After bioinformatic processing of raw data with an in-house developed pipeline, we analyzed our samples on Highlander (a software developed in our group by R. Helaers). Each sample had on an average ≈50,000 detected variants. We retained variants for further analysis if they met the following criteria: (i) passed the GATK filter (≈49,000), (ii) reported allele count of ≤10 in the population from the 1000 genomes project (≈40,000), (iii) not reported in the population from the GO-NL (genome Netherlands) project (≈6,000), (iv) occurrence in our in-house WES variant list in CL/P samples that passed GATK and at most 2 subjects afflicted with a different condition than CL/P that did not pass GATK (≈957). As “likely pathogenic“ variants, we considered those with a high impact (stops) or a moderate impact (non synonymous-NS), as predicted by SnpEff software. On average, a sample harbored 12 stops and 78 NS variants. All NS variants were predicted as damaging by at least 3 softwares. To identify causative gene(s) shared by multiple subjects, we used the following combinatorial filtering: {a} ≥2 samples with high impact variants in the same gene: 2 variants in 1 gene retained; {b} ≥2 samples with high or moderate-impact variants: 37 variants in 16 genes retained ; and {c} ≥2 samples with moderate-impact variants: 267 variants in 122 genes retained. Our results demonstrate important locus heterogeneity for familial CL/P. To be able to distill out the causal gene from the aforementioned data, we will WES additional affected subjects from multiplex families, for which continued collection is also ongoing.

3158T

CHROMOSOME ABERRATIONS IN A MEXICAN PEDIATRIC HOSPITAL. RING CHROMOSOMES 4, 13 AND 18. *I. Hurtado-Hernadez[1], J.M. Aparicio Rodríguez[2, 5], M.P. Barrientos[3], S. Chatelain Mercado[4].* 1) Cytogenetics,; 2) Genetics; 3) Endocrinology, Hosp para el Nino Poblano, Puebla; 4) Biotechnology, Universidad Autónoma Metropolitana; 5) Estomatology Benemérita Universidad Autónoma de Puebla, México.

he chromosome alteration due to a ring formation, is rare associated to chromosomes 4, 13 and 18 which is in relation to phenotypic malformations, neurologic problems and genital abnormalities. To clinical poly malformed case with treboliform skull, dimorphism with early seizures and malformed genitals with micropenis is presented. Among chromosome alterations, the ring of autosomic chromosome 4, 13 and 18 are not frequent, the main phonotypical alterations in this study are in relation to neurological, genital and craniofacial malformations. Taking in consideration that mutations or chromosome aberrations are considered alterations in the chromosome number or structure. They are mainly considered due to gametogenesis inborn error (meiosis) or during the zygote first cellular divisions. All these alterations might be observed during metaphase from the cellular cycle, where DNA loses are seen due to DNA repair processes deficiency o total absence, among others. Almost 5000 chromosomal studies were performed at Hospital Para El Niño Poblano (Pediatric Hospital) in Mexico (1n 19 years) were 34.6% (1596 patients) showed different chromosomal alterations and only two patients showed ring chromosome aberrations. These chromosome changes are classified as structural alterations. All pediatric patients with these genetic diseases are described in this study analyzing their clinical characteristics, medical or surgical treatments according to the phenotypic alterations.

3159T

Use of MLPA of buccal smear for Pallister-Killian diagnosis. *L.S.A. Costa, M. Moreira, A. Zandona-Teixeira, A. Dias, L. Kulikowski, R. Honjo, D. Bertola, C. Kim.* Dept Pediatrics, Universidade de São Paulo , São Paulo , Brazil.

Pallister-Killian syndrome (PKS) is a sporadic genetic disorder caused by the presence of a tissue-specific mosaicism for isochromosome 12p - i(12p) and is characterized by facial dysmorphism including coarse facies, upslanting palpebral fissures; bitemporal alopecia, pigmentary skin anomalies, developmental delay, hypotonia and seizures. Karyotype of skin, amniotic fluid, buccal smear or others tissues are essential for PKS diagnosis because cytogenetic analysis of peripheral lymphocytes usually fails to detect the mosaicism of isochromosome 12p - i(12p). We reviewed the medical records of patients with confirmed PKS followed in our service (since 1981 to 2014) and performed MLPA (Multiplex Ligation-dependent Probe Amplification) for diagnosis and confirmation in two of them. In this study we propose a new methodology for PKS diagnosis: MLPA of buccal smear. We describe eight patients with PKS. Clinical features of all patients were consistent with those described in the literature. Age at diagnosis varied from prenatal to 3 years. In all patients, peripheral blood karyotypes were normal. Seven patients were diagnosed by a skin fibroblast karyotype with i(12p) in 50% to 100% of cells. Recently, it was performed in two patients MLPA tecnique of buccal smear, showing 4 copies of 12p. Two patients have been lost during follow up and one patient died from pneumonia at 14 years old. The current age of five patients ranges from 9 months to 18 years, all of them with severe mental retardation. A precocious PKS diagnosis is important to optimally manage the disease and to provide genetic counseling. MLPA on buccal smears is a good and non-invasive method to detect extra copies of 12p and should be considered at the first exam, before a skin biopsy for a fibroblast karyotype is performed.

3160T

Cytogenetic characterization of 2 patients with supernumerary marker chromosomes (sSMCs) derived from complex genetic rearrangements. *A.L. Penton, R.D. Burnside, I. Gadi, V. Jaswaney, S. Schwartz, K. Phillips, J.B. Schleede, J.H. Tepperberg, J. Whiley, P. Papenhausen.* Cytogenetics, Laboratory Corporation of America, Durham, NC.

Small supernumerary marker chromosomes (sSMCs) that cannot be identified with standard cytogenetic G-banding techniques are estimated to occur in approximately 0.04% of live births potentially causing congenital abnormalities if associated with copy number gains in regions of clinically significant genes. We present cytogenetic findings from two newborn patients referred for cytogenetic analysis . Patient 1: Clinical indications for patient 1 included respiratory distress and suspected sepsis. G-banded chromosome analysis of lymphocytes collected from this patient showed the variable presence of 4 to 6 marker chromosomes in 20 cells counted . Whole genome SNP microarray analysis revealed a total copy number gain of 51.14 MB consisting of the pericentric regions of chromosomes 1, 6 and 9, as well an interstitial region of chromosome 17, suggesting the possibility of at least one composite marker. Subsequent FISH studies confirmed that these genomic regions were present in three distinct sSMCs . Patient 2: This patient was born at 29 weeks with congenital anomalies including a ventricular septal defect (VSD), patent ductus arteriosus (PDA), low set ears and widely spaced nipples. Whole genome SNP microarray analysis detected a total copy number gain of 15.80 MB originating from the pericentromeric regions of chromosomes 1, 8 and 13, and interstitial regions of chromosomes 6, 16 and X. A consistent percentage of mosaicism was detected for most of these regions, again suggesting the presence of several mitotically unstable composite marker chromosomes. Chromosome studies confirmed the variable presence of 3 to 6 supernumerary marker chromosomes in 30 cells counted . These results indicate that sSMCs can originate from complex genetic rearrangements and illustrate the necessity and efficacy of whole genome microarray analysis to precisely identify the overall genomic imbalance. Designation of the specific integration of the additional interstitial segments in each marker will entail numerous FISH studies which can be facilitated by multiprobe painting cocktails.

3161T

Double translocations in individuals and multiple family members. *D.H. Zaleski, K.J. Anderson, A.Z. Van Dyke, S.D. Nielsen, J.N. Sanmann, R.T. Hagelstrom, B.J. Dave, W.G. Sanger.* Human Genetics Laboratory, UNMC/ Munroe Meyer Institute, Omaha, NE.

The incidence of multiple, apparently balanced translocations in the general population is rare. Our lab has identified five families with multiple translocations, two of which have been previously reported. This study details the cytogenetic findings in the remaining three families. Family 1: Chromosome studies on a proband with cerebral palsy and intellectual disabilities revealed a balanced translocation: t(5;6)(p15.2;q23). Parental studies showed the father to be a carrier of a double balanced translocation: 46,XY,t(1;18)(p34.2;q11.2), t(5;6)(p15.2;q23). Two other siblings were also carriers of either one or both balanced translocations. Family 2: Karyotypic analysis of a newborn with osteogenesis imperfecta revealed two balanced translocations : 46,XX,t(2;6)(p11.2;p21.1),t(3;7)(q25;q11.2). The mother had a normal karyotype and the father was unavailable for study. Family 3: Cytogenetic analysis of a product of conception (POC) specimen on an individual with a history of multiple miscarriages showed the presence of an abnormal karyotype: 46,XY,der(17)t(6;17)(p21.3;p11.2),t(14;18)(q24.1;q23). Chromosome studies on peripheral blood of the patient revealed that she was a carrier of two balanced translocations: 46,XX,t(6;17)(p21.3;p11.2), t(14;18)(q24.1;q23). Typically, balanced translocations do not affect the health of the carrier. However, they are at an increased risk for miscarriages and/or children with chromosomal imbalances associated with clinical involvement. This risk is greater for individuals with two balanced translocations. For example in family 3, a translocation risk specialist calculated that at the given breakpoints, the t(6;17) poses a 38% risk of miscarriage and a 2.5% chance of having a live born with an unbalanced translocation. The t(14;18) results in a 31% risk of miscarriage and 2-4.5% chance of a live born with an unbalanced translocation. Considering the additional background risk of 20%, the calculated theoretical risk for miscarriage in this patient is ~89% and the theoretical risk for live birth with an unbalanced translocation is 4.5-7%. This leaves only 4-6.5% chance of having a child with normal or balanced chromosomes. For individuals with double translocations the risk of miscarriage/abnormal live born is much higher due to the cumulative effect of both translocations. Familial studies are important to derive accurate recurrence risk and appropriate genetic counseling.

3162T

Reversing differences in chromatin accessibility that distinguish homologous mitotic metaphase chromosomes. *W.A. Khan[1], P.K. Rogan[2,3,4], J.H.M. Knoll[1,4].* 1) Department of Pathology, London, ON; 2) Department of Biochemistry, London, ON; 3) Department of Computer Science, London, ON; 4) Cytognomix, London, ON Canada.

Treatment with agents that alter histone modifications, DNA conformation or its sequence are well established strategies for cancer chemotherapy. Apart from their effects on interphase chromatin (G1, S), little is known about how these compounds act during and preceding metaphase. We have previously demonstrated that most individual single copy (SC) FISH probes (1.5-5 kb) exhibit equivalent fluorescence hybridization intensities to allelic metaphase targets. A subset of these probes show reproducible, non-random hybridization intensity differences between homologous targets. Atomic and super-resolution microscopy analyses indicate that the differences correspond to 'differential accessibility' (DA) at these genomic loci. We investigated epigenetic characteristics of DA with agents that inhibit histone deacetylation, phosphorylation, cytosine methylation, and alter superhelicity in lymphoblastoid cell lines. Compared to controls, DA was impervious to chemical treatments targeting histone modifying enzymes or cytosine methylation from five distinct genomic regions (*RGS7*, 1q43; *CACNA1B*, 9q34.3; *ADORA2B*:IVS1, 17p12; *PMP22*:IVS3, 17p12 and *ACR*, 22q13.33). We also hybridized etoposide-treated metaphase cells with probes to the same genomic regions that exhibited DA in untreated cells. DA was reversed at all of these loci at various physiological etoposide concentrations ($p < 0.05$, two-proportion z-test, n = 30-40 cells/region). Differences in fluorescent probe intensity ratios measured between homologs from etoposide-treated cells were also 4 to 5-fold lower relative to untreated controls. With 3-D structured illumination microscopy, we demonstrated that the mean differences in hybridized probe volume between allelic targets significantly decreased ($p < 0.05$, two-tailed t-test) in treated cells ($\Delta\mu = 0.237\mu m^3$) compared to untreated controls in which one allele was less accessible ($\Delta\mu = 0.730\ \mu m^3$). Inhibition of the topoisomerase IIα-DNA cleavage complex decreased DNA supercoiling and axial metaphase chromosome condensation. Our findings suggest that allelic differences in metaphase chromosome accessibility represent a stable chromatin mark that is largely maintained on mitotic metaphase chromosomes; and that the local differences in condensation are due to topoisomerase activity, which result in DA.

3163T

Longitudinal shortening in telomere length as a biomarker for dementia status of adults with Down syndrome. *E.C. Jenkins[1], L. Ye[1], S.J. Krinsky-McHale[1], W.B. Zigman[1], N. Schupf[1,2], W.P. Silverman[3].* 1) Department of Human Genetics, NYS Institute for Basic Research in Human Genetics, Staten Island, NY; 2) Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, New York, NY; 3) The Kennedy Krieger Institute and The Johns Hopkins University School of Medicine, Baltimore, MD.

We have previously reported shorter telomeres in T-lymphocytes from participants in a cross-sectional study of 11 people with Down syndrome (DS) and mild cognitive impairment (MCI) or dementia versus those with DS only, using PNA (peptide nucleic acid) probes for telomeres and the chromosome 2 centromere (cen (2) PNA probe was a gift from DAKO for our research studies) (Jenkins et al., 2012). When measuring individual metaphase chromosomes, the length and/or light intensity of the cen (2) probe can be used to generate an "internal standard" to normalize telomere data. To calculate such a ratio (see Jenkins et al. 2012), the fluorescent light intensity of the chromosome 1 telomeres, for example, becomes the numerator while the denominator becomes the light intensity for the cen (2) PNA probe. The intensity/length for the cen (2) PNA probe should remain stable, while the light intensity/length of the telomeres is expected to shorten with each successive cell division. Multiple methods comparing telomere lengths of individuals with and without MCI revealed clear and consistent effects of clinical status in preparations using an FITC-labeled PNA telomere probe (DAKO) but longitudinal data were needed to verify that the effects reflect true accelerated telomere loss. Preliminary results from our first longitudinal study (N = 6; 4F, 2M) have now shown that interphase and metaphase chromosomes from T-lymphocytes of repeat samples from people who did not have MCI/dementia at Time 1 but who had converted to MCI (N = 5) or dementia (N = 1) by the time a second sample was obtained, all had substantially shorter telomeres after conversion to MCI/dementia. Compared to our previous cross-sectional results, group means prior to MCI/dementia onset were (0.12 - 0.17)[Note: 0.12 is the ratio obtained when chromosome 1 telomere length in microns is divided by chromosome 1 length minus chromosome 1 telomere length.] versus the "non-MCI/demented" cross-sectional group (0.19 - 0.23), while the MCI group telomere lengths following transition to MCI were 0.08 - 0.1 comparable to the cross-sectional MCI/demented group (0.08 - 0.13). We are anxious to extend this study with additional prospective follow-up of a larger sample of individuals with Down syndrome who have not yet converted to MCI/dementia. Supported in part by NYS OPWDD, Alz. Assoc. IIRG-96-077, IIRG-07-60558; NIH PO1-HD035897, RO1-HD037425, RO1-AG014673, and P30-HD024061.

3164T

The role of copy number variation in non-syndromic cleft lip and palate. *L.A. Harney[1,3], B.W. Darbro[1,3], A. Long[2], J. Standley[1], J.C. Murray[1,3], J.R. Manak[1,2,3].* 1) Department of Pediatrics, University of Iowa, Iowa City, IA; 2) Department of Biology, University of Iowa, Iowa City, IA; 3) Interdisciplinary Genetics Program, University of Iowa, Iowa City, IA.

Clefts of the lip and/or palate (CLP) are common congenital abnormalities affecting approximately 1 in 700 live births. It is estimated that 70% of CLP cases are non-syndromic (NS) which excludes individuals with CLP who exhibit cognitive or multiple congenital abnormalities. A variety of candidate gene, genome-wide association studies, and animal models have been used to study CLP, but no large-scale studies have explored the role of copy number variation (CNV) in clefting. We performed array-based comparative genomic hybridization on a NSCLP cohort from the Philippines to identify CNVs associated with clefting. After quality controls minimized false-positives, we processed 84 NSCLP cases and a replication cohort of 854 NSCLP cases for further analysis. Specifically focusing on deletions for this study, we used an analysis pipeline which identified CNV losses that overlapped with exons of genes in regions sharing 50% or less overlap with segmental duplications and common CNVs. Analysis of CNVs in the 84 NSCLP cases identified 33 deletions containing 53 genes, 1 of which overlapped with a previously identified clefting locus *CDH1* (MIM 192090). We also investigated the GWAS region significantly associated with NSCLP, *8q24* (MIM 612858). Analysis of *8q24* revealed three small deletions within a 1.4Mb region containing the most associated SNP *rs987525* (MIM 612858) and a recently reported murine medionasal enhancer of *MYC* (MIM 190080). Characterization of these deletions is underway. Seventeen regions from the small cohort, containing 23 genes, replicated in the larger cohort. In total, 178 regions appeared in two or more individuals in the large cohort, including six previously identified clefting loci *XRCC3* (MIM 600675), *JAG2* (MIM 602570), *8q24*, *NTN1* (MIM 601614), *DLG1* (MIM 601014), and *LPL* (MIM 609708). To elucidate the role of these CNVs in CLP, we are performing expression analysis of genes within the altered copy number regions, as well as altering their dose in zebrafish. In addition, we are conducting a trio study using the loci identified in the small cohort to determine if the CNVs are *de novo* or familial. Since this large dataset provides immense potential for the identification of novel CLP loci, we plan to extend the analysis to intronic and intergenic CNVs. It is our hope to define how CNVs contribute to NSCLP and identify novel causative variants for the disease.

3165T

Characterization of a fusion gene involving the leptin gene generated by a *de novo* 7q32.1 duplication associated with severe anorexia. *J. Lévy[1,2,3], E. Pipiras[1,4,5], L. de Pontual[5,6,7], V. El Ghouzzi[1,3], N. de Roux[1], AC. Tabet[2], P. Gressens[1,3], B. Benzacken[1,2,4,5], S. Lebon[1,3], A. Delahaye[1,4,5].* 1) Inserm U1141, Hôpital Robert Debré, Paris, France; 2) AP-HP, Service de Génétique et de Cytogénétique, Hôpital Robert Debré, Paris, France; 3) Université Paris Diderot, Sorbonne Paris Cité, Paris, France; 4) AP-HP, Service d'Histologie, Embryologie, et Cytogénétique, Hôpital Jean Verdier, Bondy, France; 5) Université Paris 13, Sorbonne Paris Cité, UFR SMBH, Bobigny, France; 6) AP-HP, Service de Pédiatrie, Hôpital Jean Verdier, Bondy, France; 7) Inserm, U781, Université Paris Descartes, Paris-Sorbonne Cité, Institut IMAGINE, Paris, France.

Leptin is a polypeptide hormone secreted by white adipocytes which plays a major role in the regulation of appetite and weight. Leptin action is mediated by receptors that are located in the arcuate nucleus of the hypothalamus and in many peripheral tissues. Homozygous mutations in either the leptin or leptin receptor genes cause extreme obesity in mice and humans. Here we describe a *de novo* 7q32.1 duplication spanning the *LEP* gene discovered in a 2-year-old boy referred for unexplained precocious anorexia associated with postnatal ponderal growth retardation.

Whole-genome sequencing revealed that the duplication was a direct tandem duplication. This duplication generated a fusion between a potential pseudogene paralogous to the *TLK2* gene (thereafter referred as *TLK2P3*) and the *LEP* gene. Several fusion transcripts generated by alternative splicing of *TLK2P3* were detected by RT-PCR on patient's fibroblasts and lymphocytes. They contained the complete coding sequence of the *LEP* gene. Constructions of expression vectors encoding HA-tagged leptin, TLK2P3 and TLK2P3-LEP were performed to characterize the potential fusion proteins. The expression of TLK2P3-LEP, LEP and TLK2P3 were analyzed by immunoblotting and western blot after transfection of human HeLa cells.

The first results of the fusion gene expressed in HeLa cells showed the production of a protein with the same size and the same cellular localization as leptin. Additional alternative fusion transcripts expressions in HeLa cells are ongoing. ENCODE analysis showed a highly active promoter regulating *TLK2P3*, predicted by a chromatin state segmentation for each of nine human cell types. This highly active promoter may induce deregulated and ectopic expression of leptin.

Several hypotheses of the mechanisms by which this 7q32.1 duplication affect the phenotype of the patient will be discussed. Although our results are preliminary, this study suggests that fusion genes can be a mechanism by which CNVs contribute to Human diseases.

3166T

Automated Dicentric Chromosome Identification by Machine Learning-Based Image Processing. *P.K. Rogan[1,2], Y. Li[1], A. Subasinghe[1], J. Samarabandu[1], R. Wilkins[3], J.H. Knoll[1,2].* 1) University of Western Ontario, London ON; 2) Cytognomix, London ON; 3) Health Canada, Ottawa CA.

Accurate detection of chromosome centromeres is a critical element of cytogenetic diagnostic techniques, including chromosome enumeration, karyotyping and radiation biodosimetry. We present an approach for identification of normal and dicentric chromosomes (DC) in either Giemsa- or DAPI-stained metaphase cells. Existing centromere segmentation methods perform poorly in the presence of irregular chromosome boundaries, shape variations and sister chromatid separation. Different sample preparation protocols can produce inconsistent centromere localizations. We apply a profile thickness measurement technique to chromosome structures of intensity-integrated Laplacian segmented images. A set of centromere candidates is generated, evaluated by a support vector machine (SVM) based on a set of image measurements, ie. features. We partition the chromosome contours to isolate telomere regions, then detect and correct for sister chromatid separation. This algorithm accurately localized 1220 of 1400 centromeres (87%). Distinction of true from false centromeres was enhanced by adding new features to the SVM. The new features incorporated pixel intensity, angular deviation from the projected centerline of each chromosome, and fitted a curve to the chromosome width. The original and new versions of the SVM were evaluated and assessed for correct assignments of both centromeres in known DCs. Candidate centromeres in 247 DCs were assessed based on their respective distances to the SVM hyperplane. The original and new SVMs were 89% and 92% accurate, respectively, for the identification of both centromeres among the 4 closest candidates to the hyperplane. For the true centromeres, ie. the 2 closest candidates to the hyperplane, accuracy was reduced (58% and 57%). SVMs based on predominantly monocentric chromosomes detected individual centromeres in these chromosomes and prioritized candidates in DCs, but did not reliably identify DCs de novo. This is caused by weak feature combinations in the SVMs for some DCs. A set of 158 DCs was used exclusively to train and test the SVMs by cross-validation. Accuracy was unchanged for the original SVM (59%), but the new SVM showed considerable improvement (75%). The SVMs are also more likely to miss candidate centromeres in normal acrocentric chromosomes or acrocentric-derived DCs. Expansion of the DC training set and custom segmentation of acrocentric-derived DCs are promising approaches to further improve performance.

3167T

Detectable mosaic 13q14 deletions in non-hematologic cancers and healthy controls. *M. Yeager[1,2], M. Machiela[2], W. Zhou[1,2], M. Dean[3], S.J. Chanock[1].* 1) Cancer Genomics Research Laboratory, Leidos Biomed, FNLCR,, Frederick, MD; 2) Division of Cancer Epidemiology and Genetics, NCI, Bethesda, MD; 3) Center for Cancer Research, NCI, Frederick, MD.

Mosaic loss of 13q14.3 is a common chromosomal event found in approximately 50 percent of B-cell chronic lymphocytic leukemia (CLL) cases. Studies suggest 13q14.3 loss may be evident in somatic DNA of other non-hematologic tumor types, but to date no comprehensive study has been performed on the frequency of 13q14.3 mosaic loss in lymphocyte-derived DNA of solid tumor cases and healthy controls. Based on preliminary findings indicating an excess of 13q14.3 mosaic loss in our previous studies of detectable clonal mosaicism, we performed a comprehensive scan of chromosome 13 from 46,254 non-hematologic cancer cases and 36,229 controls using modified B-allele frequency and log R ratio based detection algorithms. We detected a total of 60 individuals with 13q14.3 mosaic loss, 1 occurrence of mosaic copy neutral uniparental disomy, and 13 individuals with stretches of homozygosity that spanned the 13q14.3 region. While detected 13q14.3 mosaic losses were variable in size, the minimally deleted region (MDR) in our analysis (chr13:49,590,000-49,983,100, NCBI36/hg18) closely resembled that seen in CLL and included transcripts of DLEU1 and DLEU2. A breakpoint analysis on 13q14.3 loss boundaries indicates a clustering of breakpoints with statistically significant enrichment for breakpoints around gene rich regions and areas of open chromatin when compared to random permutations of same-sized events spanning the chromosome 13 MDR. Our analysis indicated the frequency of 13q14.3 loss significantly increases with increasing age (p-value=0.028). The frequency of 13q14.3 mosaic loss did not significantly differ between non-hematological cancer cases (0.084%, 95% CI: 0.058-0.111%) and controls (0.058%, 95% CI: 0.033-0.083) nor did frequency estimates differ from age and sex standardized SEER estimates. While 13q14.3 deletions may serve as early biomarkers for CLL, our study indicates such events are present and potentially tolerated in peripheral blood lymphocytes of disease free controls and non-hematologic cancer cases.

3168T

Microdeletions and Microduplications in Brazilian Children with intellectual disability from a public health service. *A.D. da Cruz[1,2,3,4,5], I.P. Pinto[1,2], L.B. Minasi[1,2], A.V Melo[1,3], D.M.C. Cunha[1,2], A.S. da Cruz[1,5], C.L. Ribeiro[1], D.M. e Silva[1,2,6], C.C. da Silva[1,2,3,4].* 1) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group; 2) Pontifical Catholic University of Goias, Genetics Master's Program; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program; 4) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health; 5) Federal University of Goias, Biology Master's and PhD Programs; 6) Federal University of Goias, Genetics and Molecular Biology Master's and PhD Programs.

Intellectual disability is a complex, variable, and heterogeneous disorder, representing a disabling condition diagnosed worldwide and the etiologies are multiple and highly heterogeneous. Microscopic chromosomal abnormalities and well-characterized genetic conditions are the most common causes of intellectual disability. Here, we report 15 probands with undiagnosable intellectual disabilities referred by doctors from the public health system in Central Brazil. For all probands, G-banding karyotypes showed no visible alterations (46 XX or 46XY). Chromosomal Microarray Analysis (CMA) using Affymetrix CytoScanTM HD Array showed a total of 18 CNVs that were identified in 9/15 (60%) patients and no chromosome rearrangement was observed in 6/15 (40%) cases. Moreover, molecular karyotyping of all 30 progenitors included in our study showed no evidence of chromosome rearrangements. Pathogenic and likely pathogenic CNVs were classified based on their size, gene content and diversity, and previously reported cases of potential involvement with pathogenic mechanism in human and animal models. In our cohort, pathogenic CNVs represented about 22% of all observed rearrangements and were associated with chromosomes 17, 18, and X, involving genes that were related to the formation and/or maintenance of the central nervous system. Also, pathogenic CNVs included morbid genes from OMIM. Moreover, all pathogenic CNVs in this study had a de novo origin. 22% CNVs were classified as likely pathogenic and 56% CNVs were classified as of unknown clinical significance because they overlapped by more than 90% of the CNVs observed in the databases of normal control groups. CMA technology is a relatively new strategy useful as an additional tool for genetic diagnosis. The method has been recommended as the first-tier diagnostic test for patients with global developmental delay, intellectual disability, autism spectrum disorders, and multiple congenital anomalies. Furthermore, the correct diagnosis of a neurological disorder is crucial for predicting the probands' clinical follow up, to establish accurate prognostic, and to provide adequate genetic counseling.

3169T

First case of homozygous deletion in the ABAT gene leading to GABA-T deficiency and severe neonatal neurologic disease. *A. Mosca-Boidron[1,2], G.S. Salomons[3], E.A. Struys[3], C. Vianey-Saban[4], L. Faivre[2,5], M. Lefebvre[5], N. Marle[1,2], F. Feillet[6], M. Payet[1], C. Ragon[1], F. Mugneret[1], A. Masurel-Paulet[5], J. Thevenon[2,5], M.G. Mourot de Rougemont[7], C. Thauvin-Robinet[2,5], P. Callier[1,2], S. El Chehadeh[2,5].* 1) Laboratoire de cytogénétique, Plateau technique de biologie, CHU de Dijon, France; 2) GAD, EA4271, Génétique et Anomalies du Développement, Université de Bourgogne, Dijon, France; 3) Metabolic Unit, clinical chemistry, VUmc Medical center, de Boelelaan, Amsterdam, the Netherlands; 4) Service des maladies héréditaires du métabolisme et dépistage néonatal, Groupement hospitalier EST, Bron, France; 5) FHU TRANSLAD, Centre de référence maladies rares « anomalies du développement et syndromes malformatifs » de l'Est, Centre de Génétique, CHU de Dijon, France; 6) Service de médecine infantile, Hôpital d'Enfants de Brabois, CHU de Nancy, Vandoeuvre-les-Nancy, France; 7) Service de radiologie, Hôpital d'Enfants, CHU de Dijon, France.

Background 4-gamma-aminobutyrate transaminase (GABA-T) catalyzes the conversion of gamma-aminobutyric acid (GABA), one of the most important inhibitory neurotransmitters in the central nervous system of mammalians, into succinic semialdehyde and is encoded by the ABAT gene. GABA-T deficiency is a very rare autosomal recessive disorder characterized by severe psychomotor retardation, hypotonia, hyperreflexia, seizures, associated with early infantile death. It has been reported in only 3 patients to date, 2 of whom were siblings. Electroencephalographic and brain magnetic resonance imaging (MRI) abnormalities including respectively burst suppression pattern, cerebellar hypoplasia, posterior fossa cyst, corpus callosum agenesis and abnormal gyration, have been previously described. We report here on a fourth male patient born to consanguineous healthy Moroccan parents who presented with neonatal epileptic encephalopathy including major axial hypotonia, lethargy with hypomotility and intractable seizures, without any dysmorphic features. He died at 15 days of life. Methods Investigations that were done in the first week after birth included array-CGH, brain MRI, and wide metabolic screening. Results Brain MRI showed atrophy of the cerebellar vermis and hemispheres with a cerebellar cyst aspect, reduced white matter, very thin corpus callosum and diffuse thickening of the cortex leading to the suspicion of pachygyria. Electroencephalograms were abnormal showing burst suppression patterns. Array-CGH (Agilent 180 K) showed a homozygous 28 kb intragenic deletion of the 16p13.2 region encompassing exons 3 to 10 of the ABAT gene. This deletion was inherited from both parents who carried the deletion at a heterozygous state. Chromatography of urinary aminoacids showed an increased level of GABA (197 μMol/mmol creat; normal < 10 μMol/mmol creat) while GABA concentration was normal in blood and cerebrospinal fluid. Conclusion This observation is the first reported case of a homozygous deletion in the ABAT gene with clinical, imaging and biochemical features consistent with a diagnosis of GABA-T deficiency. It confirms the power of array-CGH in identifying intragenic rearrangements in genes implicated in rare diseases, in particular when there is no clinical clue leading to diagnosis, as is the case in many inborn errors of metabolism.

3170T

Manifestations of Xp22.2-22.13 and Xp21.3 microduplications. *M. Goto[1,2], A. Matsumoto[1], K. Kojima[1], E.F. Jimbo[1], M. Mori[1,3], H. Osaka[1], T. Yamagata[1].* 1) Pediatrics Dept, Jichi Medical University, Shimotsuke, Tochigi, Japan; 2) Pediatrics Dept, Kitaibaraki Municipal General Hospital, Kitaibaraki, Japan; 3) Pediatrics Dept, Matsudo City Hospital, Matsudo, Chiba, Japan.

Three Japanese patients with intellectual disability (ID) were reported to have two duplications on chromosome Xp. One was at Xp22.2-22.13 including exons 2-18 of REPS2 (MIM 300317) and exons 1-3 of NHS (MIM 300457). The other was at Xp21.3 including exon 2 of IL1RAPL1 (MIM 300206). Although the mechanisms underlying these duplications and their contributions to ID were not fully understood, they were strongly considered to be responsible for ID because they were not detected in unaffected males. Here we report on a boy with the same duplication showing ID, involuntary movements, and epilepsy to delineate the phenotype associated with this duplication. Case Report: The patient was a 7-year-old boy born to non-consanguineous healthy parents. He was bedridden because of severe ID and hypotonia. His clinical phenotype included epilepsy, athetoid movements, strabismus, hypodontia, delayed myelination, and cerebral atrophy. Array CGH analysis using an Agilent Human Genome CGH 180K detected a 728- and 95-kb duplication at Xp22.2-22.13 and Xp21.3, respectively. FISH analysis using two probes that were designed for Xp21.3 and Xp22.2-22.13 showed that the duplicated sequence at Xp21.3 was adjacent to the duplicated sequence at Xp22.2. His mother had the same duplication. Discussion: The duplications detected in our patient were identical to those detected in three previously reported Japanese patients (Honda, et al. J Hum Genet. 2010). One of these patients had West syndrome, severe ID, epilepsy, aphasia, and hippocampal atrophy; the other two were dizygotic twins with moderate ID, speech delays, and autism. Moderate to severe ID, epilepsy, athetoid movements, cerebral atrophy, delayed myelination and hypodontia were candidates for the phenotypes of this syndrome. In addition to these patients, females from a Korean family were found to be carriers for identical double duplications (Kondo, et al. J Hum Genet. 2012). This suggests that these chromosomal rearrangements could have been transmitted from a common ancestor in eastern Asia. The presence of excess genes localized in these duplicated regions was not pathogenic because these genes were partially duplicated. Inverse PCR analyses reported by Kondo et al. suggested that these duplications were related to each other and that they had occurred simultaneously. It was considered that complex genomic rearrangements including NHS and IL1RAPL1 had occurred and that they had consequences on neuronal development.

3171T

Comparison of de novo and inherited copy number variants of unknown clinical significance. *L. Matyakhina, S. Aradhya, D. Pineda, A. Janze, J. Nieto, G. Richard, J. Meck.* GeneDx, Gaithersburg, MD.

Variants of uncertain/unknown clinical significance (VUS), defined as a variation in copy number whose association with disease risk is unknown, comprise a significant percentage of the reports generated by a clinical microarray laboratory. Inheritance of the aberration from a parent is often used as a means of determining whether or not an aberration classified as VUS has clinical significance. If the aberration is de novo or is inherited from a similarly affected parent, then the aberration is most likely pathogenic in nature. By contrast, those VUS aberrations, which are found in a phenotypically normal parent, are more likely to be benign. We reviewed our VUS data from a custom-designed 180K whole genome microarray to determine patterns characteristic of inherited versus de novo VUS calls. Two hundred fifty six cases were tested for inheritance: 203 aberrations (79%) were inherited, 34 (13%) cases were de novo, and the remainder was inconclusive since only a single parent was tested. Of the 203 inherited cases, 134 were maternally inherited and 69 were paternally inherited (maternal: paternal ratio was 1.9). Duplications were inherited three times more frequently than deletions. In contrast to the inherited VUS aberrations, de novo deletions and duplications were represented in roughly equal proportion, with a slightly higher percentage of deletions (58%). Comparison of de novo versus inherited aberrations showed that the majority of the de novo variants contained significantly more genes, with a maximum of 35 genes, versus a maximum of 12 genes for the inherited aberrations. In addition, although there is overlap in aberration size, de novo and inherited VUS copy number variants differed with respect to overall size with 38% of de novo deletions and duplications being over 1 Mb versus 11% for inherited. For many de novo cases, parental testing was helpful to reclassify a VUS call to likely pathogenic. These data demonstrate the value of parental testing for better classification of copy number variants identified by chromosomal microarray and underscore the need for collaborative efforts to better resolve and understand the significance of inherited copy number variants.

3172T

Whole Genome Characterization of Copy Number Variation Regions in 2000 Phenotypically Normal Individuals. *A. Roter[1], J. Sampathkumar[1], C. Du[1], S. Schwartz[2], S. Scherer[3,4], M. Uddin[3], B. Thiruvahindrapuram[3], C. Marshall[3], R. Fpundt[5], T. Jonson[6], A.C. Thuresson[7], I. Quintela[8], F. Barros[8], J. Cuevas[1].* 1) Informatics, Affymetrix, Inc., Santa Clara, CA; 2) Cytogenetics Laboratory, Laboratory Corporation of America, Research Triangle Park, NC; 3) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 5) Radboud University Medical Centre, Nijmegen, Netherlands; 6) Department of Clinical Genetics Skåne University Hospital - 221 85 Lund, Sweden; 7) Department of Immunology, Genetics and Pathology, Rudbeck and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 8) Grupo de Medicina Xenómica, UdC, CeGen- CIBERER, Santiago de Compostela, Spain.

A large cohort of 2,321 genomic DNA samples from phenotypically normal individuals was characterized for genomic copy number variations to understand the frequency and genomic distribution of non-pathogenic CNVs in the population. DNA samples were extracted from blood samples obtained at 6 sites in 6 countries, including Sweden, the Netherlands, Norway, Spain, Canada and the United States. Samples were processed with the Affymetrix CytoScan® HD Microarray Kit and the data analyzed for copy number calls using Affymetrix Power Tools. Copy number calls were extracted from the output files and compiled into a large dataset, including only CNVs that exceeded a minimal size threshold, 25 markers and 25 kbp for copy number losses or 50 markers and 50 kbp for copy number gains, using a composite profile from 100 blood and 280 cell line samples as a reference. The CNV data set was evaluated for the frequency of finding a loss or gain at every genomic location. Frequencies of adjacent genomic locations were smoothed using an exponential decay window-smoothing algorithm and the data were segmented to reconstruct common regions of CNVs. We found less than 5.5Mb of the human genome to be associated with CNV regions having population frequencies of 10% or more. These common CNV regions were analyzed for genomic structure, such as segmental duplications and chromosomal position to help understand the genomic attributes associated with regions with frequent CNV polymorphisms.

3173T

Familial transmission of 5p13.2 duplication due to maternal der(X)ins(X;5). *L.C. Walters-Sen[1], K. Windemuth[1], J. Nandhlal[1,2], J.M. Milunsky[1].* 1) Center for Human Genetics, Inc., Cambridge, MA; 2) Boston Medical Center, Boston, MA.

Submicroscopic duplications of 5p13 have been recently reported in a number of cases, warranting the description of a new clinical entity (Chromosome 5p13 Duplication Syndrome, MIM 613174). These microduplications, while variable in size, all contain at least part of the *NIPBL* gene (MIM:608667). Patients with duplications in this region present with intellectual disability/developmental delay (ID/DD) and dysmorphic facies. In addition, skeletal and brain abnormalities have been variably reported, as well as propensity for obesity and hypotonia. We report a family with two affected sons, each carrying a duplication at 5p13.2 encompassing the 5' portion of *SLC1A3* (MIM:600111) and the 5' portion of *NIPBL*. The proband was originally referred for a SNP microarray analysis, which identified the 341 kb 5p13.2 duplication, also found in his brother by a different laboratory. Both brothers had delays in speech development and diagnoses of possible autism and ADHD. The proband's weight was in the 95-97th percentile, while that of his brother was in the 75-90th percentile. In addition, the brothers shared facial dysmorphisms, including long eyelashes, long palpebral fissures, a broad nasal tip with anteverted nares, and a small chin. Other anomalies included a high palate, pes planus, and prominent breast tissue. Upon confirming the SNP microarray finding by FISH in the proband, it was discovered that the 5p13.2 duplication was located on the short arm of the X chromosome. Further FISH studies on the mother demonstrated that both she and the proband carried a derivative X chromosome with insertion of material from 5p13.2 into the intermediate region of Xp [der(X)ins(X;5)(p2?2.1;p13.2p13.2)]. As the mother did not share facial dysmorphism or ID/DD with her sons, X-inactivation studies were performed. The mother showed skewed inactivation (82:18), providing a mechanism for suppression of expression of the duplicated 5p13.2 material. To our knowledge, this is the first report of an inherited duplication of 5p13.2 with multiple affected family members. This family underscores the need to confirm array findings by FISH, both in the proband and family members, to discern implications for pathogenicity and more accurately define the recurrence risk.

3174T

Submicroscopic chromosomal imbalances in patients with intrauterine growth retardation and features of Silver-Russell syndrome. *A. Bonaldi[1], A.C.S. Fonseca[1], D.R. Bertola[2], C.A. Kim[2], P.A. Otto[1], A.M. Vianna-Morgante[1].* 1) Department of Genetics and Evolutionary Biology, University of São Paulo, São Paulo, São Paulo, Brazi; 2) Genetics Unit, Children's Institute, Clinics Hospital, Faculty of Medicine, University of São Paulo, São Paulo, Brazil.

Silver-Russell syndrome (SRS, MIM 180860) is characterized by severe intrauterine and postnatal growth retardation associated with typical small triangular face and other variable features. Most frequently, SRS is caused by altered gene expression on chromosome 11p15 due to hypomethylation of the telomeric imprinting center (ICR1) (40%). Maternal uniparental disomy of chromosome 7 (matUPD7) is less frequent (5-10%). Thus, about half of SRS patients remain without a genetic diagnosis. Analysis of the known SRS genetic causes in a cohort of 70 patients, whose clinical picture included growth retardation and SRS features, revealed hypomethylation of ICR1 in 28 (39.4%), and matUPD7 in four patients (5.6%). Two patients (2.8%) had maternally inherited microduplications at 11p15. One of them carried a microduplication of the centromeric imprinting center, ICR2 (Bonaldi et al., AJMG 155A:2479,2011); in addition to the ICR1 and ICR2 microduplication, the other patient had a 4p microdeletion due to a cryptic unbalanced translocation, der(4)t(4;11)(p16.3;p15.4)mat, as documented by aCGH. Among the 36 patients (51.4%) without a causative diagnosis, 21 were submitted to aCGH analysis (CytoSure 180K/UPD180K, OGT). We found 32 rare variations in 14 patients: one probably pathogenic, 11 of unknown significance (VOUS) and 20 probably benign. One patient carried a paternally inherited Yp11.32 microdeletion, encompassing ~40 Kb of SHOX downstream regulatory region. Four VOUS identified in different patients were particularly interesting: (1) a paternally inherited ~365 Kb 2p23.3 microdeletion encompassing eight protein-coding genes; (2) a de novo ~1.4 Kb Xp11.23 microduplication in a male patient with intellectual disability (ID), partially comprising the 5'UTR and first exon of the FTSJ1 gene, which has been associated with ID in men; (3) a de novo ~28 Kb 11p15.2 microduplication at SOX6 intron 4, which contains highly conserved segments. SOX6 is involved with skeletal growth and chondrogenesis, and its dysregulation might explain the patient's phenotype. A boy has been described with a similar phenotype and a balanced de novo t(9;11) translocation, the chromosome 11 breakpoint mapping to SOX6 intron 4 (Tagariello et al., JMG 43:534,2006); (4) a ~544 Kb 1p36.32 microduplication, encompassing three potein-coding genes; inheritance could not be tested. Potentially pathogenic microimbalances were found among patients with atypical SRS phenotypes. Financial support: FAPESP.

3175T

Multiple homozygosity regions in a girl with unexplained intellectual disability. *C.C. da Silva[1,2,3,5,6], F.G. Reis[1,4], L.B. Minasi[1,2], I.P. Pinto[1,2], A.V. Melo[1,3], D.M.C. Cunha[1,2], C.L. Ribeiro[1], T.C. Vieira[5,6], D.M. e Silva[1,2,4], A.D. da Cruz[1,2,3,5].* 1) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group, Brazil; 2) Pontifical Catholic University of Goias, Genetics Master's Program, Brazil; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program, Brazil; 4) Federal University of Goias, Genetics and Molecular Biology Master's and PhD Programs, Brazil; 5) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health, Brazil; 6) State University of Goias, UnU Eseffego, Brazil.

Intellectual Disability (ID) is a complex disorder characterized by significant limitations in both intellectual functioning and adaptive behavior that begin before the age of 18 years. The detection of excessive long continuous stretches of homozygosity (LCSH) on multiple chromosomes may suggest consanguinity, and could be useful in determining candidate genes for further testing for autosomal recessive Mendelian disorders. Herein, we report a first case of a Brazilian girl who presented moderate intellectual disability and LCSH detected by Chromosomal Microarray Analysis (CMA). A 9 years old girl born of first-cousin parents, at 40 weeks gestation to a 27-year-old mother and 25-year-old father, her birth weight was 1210g and crown-heel length 42 cm, and she presented mild facial dysmorphism, highly arched eyebrows, congnitive impairment, and divergent strabismus. Affymetrix's CytoScan™ HD SNP array revealed several LCSH in 9 chromosomes (2p25.3; 2p13.2; 7q32.3; 8q24.23; 10q26.12; 11q24.1; 14q24.3; 14q32.2; 15q13.1; 21q11.2 e Xp21.1) and none pathogenic CNVs. The morbid genes included *TMEM18, SNTG2, COLEC11, MYT1L, TSSC1, SOX11, RNF144; DCTN1, PCGF1, LRRTM4; CASP2, CNTNAP2; KCNK9, TRAPPC9, BAI1; FGFR2; BSX, SCN3B, NRGN, KIRREL3; NGB, NRXN3, KCNK10, TDP1, CCDC88C, ATXN3; DYNC1H1; CHRFAM7A, CHRNA7, SLC12A6, TTBK2, MAP1A, GATM, UBR1; BTG3; TSPAN7,* respectively. NCBI database analyses showed that mutations in *COLEC11* gene are related to Malpuech-Michels-Mingarelli-Carnevale Syndrome (3MC Syndrome [MIM265050]) which has autosomal recessive inheritance. The results suggest that LCSH involving the *COLEC11* gene may be responsible for the intellectual disability of a child. Although the several LCSH not diagnostic a specific condition, it indicate a causative gene located within of these regions. Thus, further studies require Exome Next Generation Sequencing to identify the mutations for candidate genes and confirm the pathogenicity of such changes. Homozygous segments increase the likelihood the proband has inherited two copies of a deleterious allele for an autosomal recessive disorder. Therefore, the genetic counseling is recommended for families to help to understand that autosomal recessive disorders are strongly associated with consanguinity and that sporadic undiagnosed case of intellectual disability, congenital anomalies and dysmorphism may have an autosomal recessive etiology with risks of recurrence in future pregnancies.

3176T

Mate-pair sequencing analysis of karyotypically balanced chromosomal rearrangements associated with cryptic imbalances reveals additional structural variants and complex genomic reorganization typical of chromothripsis or replication mechanisms. *A.C.S. Fonseca[1,2], A. Bonaldi[2], M. Bak[1], K.T. Abe[3], A.M. Vianna-Morgante[2], N. Tommerup[1].* 1) Wilhelm Johannsen Centre for Functional Genome Research, Department of Cellular and Molecular Medicine, University of Copenhagen, Denmark; 2) Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo, São Paulo, Brazil; 3) Sarah Network of Rehabilitation Hospitals, Brasilia, DF, Brazil.

About 7% of karyotypically balanced chromosomal rearrangements (BCR) are associated with intellectual disability (ID), developmental delay (DD) and/or congenital malformations (CM). Cryptic imbalances on the rearranged chromosomes appear as an important cause of the clinical phenotypes, being detected in about 27% of the BCRs analyzed by aCGH (Feenstra et al. Eur J Hum Genet 19:1152, 2011). Aiming at investigating if structural variants (SVs) undetected by aCGH co-occur with such imbalances, we applied mate-pair sequencing (MPS) to characterize eight sporadic BCRs (four translocations and four complex rearrangements) carried by patients with ID, DD and CM. Previously performed aCGH (180K, OGT) had revealed one or more imbalances (2.2kb to 5.8Mb) at or in cis to the BCR breakpoints. Mate-pair libraries were prepared using the Nextera kit (Illumina), and paired-end sequenced (2×100 bp) on HiSeq2000. Filtering strategies and cluster analysis of discordant mate pairs were performed using an in-house script. SVDetect and Delly were used for SV prediction. In total 109 breakpoints (BPs; resolution <1kb) were identified, which resulted in 60 balanced and 29 unbalanced SVs. The number of BPs and SVs varied among the 8 BCRs (n=BP/SV; n1=4/2, n2=20/18, n3=18/17, n4=10/7, n5=17/17, n6=6/5, n7=12/11; n8=22/12). Among the 89 SVs identified by MPS, 25 unbalanced and 14 balanced SVs had been detected, respectively, by aCGH or karyotyping. The rate of SVs detected exclusively by MPS varied among the different SVs: deletions (4/23), duplications (0/6), inversions (2/2), intrachromosomal direct (15/16) and inverted insertions (10/11), translocations (2/12), interchromosomal direct (7/8) and inverted insertions (10/11). All but one BCR had SVs identified only by MPS (n1=0; n2=10; n3=12; n4=4; n5=10; n6=1; n7=9; n8=4). Our findings indicate that karyotypically balanced rearrangements carrying imbalances are likely associated with inter/intra chromosomal SVs, particularly insertions, which remain cryptic to aCGH, thus deserving reinvestigation by MPS. The chromosome shattering with clustered double-strand breaks we observed in three BCRs carrying deletions suggest they arose via chromothripsis, an evidence that this mechanism can result in gross DNA losses (>1kb) in the germline. A fourth BCR with complex reorganization contained several duplications, a hallmark of rearrangements involving replication mechanisms (Liu et al. Cell 146:889, 2011). (Financial support: FAPESP).

3177T

Novel H3K4me3 marks are enriched at human- and chimpanzee-specific cytogenetic structures. *G. Giannuzzi, E. Migliavacca, A. Reymond.* Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland.

Human and chimpanzee genomes are 98.8% identical within comparable sequence. They however differ structurally in nine pericentric inversions, one fusion that originated human chromosome 2 and content and localization of heterochromatin and lineage-specific segmental duplications. The possible functional consequences of these cytogenetic and structural differences are not fully understood and their possible involvement in speciation remains unclear. We show that subtelomeric regions—that have a species-specific organization, are more divergent in sequence, and are enriched in genes and recombination hotspots—are significantly enriched for species-specific histone modifications that decorate transcription start sites in different tissues in both human and chimpanzee. Human lineage-specific chromosome 2 fusion point and ancestral centromere locus as well as chromosome 1 and 18 pericentric inversion breakpoints showed enrichments of human-specific H3K4me3 peaks in prefrontal cortex. Our results reveal an association between plastic regions and potential novel regulatory elements.

3178T

Detection of copy number variation in single cells by next generation sequencing. *F. Kaper[1], C. April[1], X. Cai[1], H.Y. Chuang[1], T. Royce[1], H. Joshi[2], C.E. Michel[2], P. Burns[2], J.B. Fan[1].* 1) Illumina Inc, San Diego, CA; 2) Illumina Inc, Cambridge, UK.

The detection of copy number variations (CNVs) in single cells has many applications, including cell lineage mapping, cancer heterogeneity research, neuronal mosaicism research and preimplantation genetic screening (PGS). Current protocols employ an upfront whole genome amplification (WGA) step such as multiple displacement amplification (MDA), primer extension preamplification PCR (PEP-PCR) or degenerate oligonucleotide-primed PCR (DOP-PCR) followed by a next generation sequencing (NGS) library preparation. Here we describe results obtained with a novel protocol that combines NGS library preparation and WGA into one robust and rapid protocol. Single, three and five cell samples were isolated through FACS sorting or narrow-bore cell transfer pipetting from commercially available human lymphoblastoid cell lines with well-described subchromosomal abnormalities. NGS libraries were sequenced on a MiSeq® or a HiSeq® 2000 sequencing system. After alignment using Bowtie, only perfectly and uniquely aligning reads were kept for tag counting. The reference genome was divided into non-overlapping bins such that each bin is expected to contain 100 uniquely mapping reads. Deviations from this expectation are correlated with copy number. Unlike MDA-based methods, the protocol does not result in overamplification pile ups, resulting in increased signal to noise. Furthermore, the library fragments are generated in a random, uniform fashion across the entire genome unlike the reproducible bias of PCR-based WGA methods. Therefore genome coverage is correlated with increasing cell number. Using 1 million reads per NGS library, confirmed CNVs of 4Mb were reproducibly detectable at single, three and five cell input levels. The entire work flow from the start of sample preparation to available sequencing data takes 9 hours when using a MiSeq sequencing system. For the detection of smaller events, NGS libraries were sequenced to greater depth on a HiSeq 2000 system. In conclusion, this technique provides a simple and rapid protocol to detect CNVs in single cells by NGS; experiments to further validate the method are currently ongoing.

3179T

Mosaicism in a Mosaic: Reduced capacity of female X chromosomes to resist age-related structural erosion. *M.J. Machiela[1], W. Zhou[1,2], M. Dean[3], J.N. Sampson[1], N. Rothman[1], N.D. Freedman[1], S. Wacholder[1], M.A. Tucker[1], F.X. Real[4,5], L.A. Perez-Jurado[5,6], M. Yeager[1,2], S.J. Chanock[1], NCI GWAS Mosaicism Consortium.* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research Inc., Frederick, Maryland, USA; 3) Laboratory of Experimental Immunology, Center for Cancer Research, NCI-Frederick, Frederick, Maryland, USA; 4) Spanish National Cancer Research Centre (CNIO), Madrid, Spain; 5) Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain; 6) Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona, Spain.

Previous studies characterizing large clonal mosaic events have not focused on the X chromosome. Therefore, we analyzed deviations in $\log_2$ intensity ratio and B allele frequency on the X chromosome using existing genome-wide association data of 38,581 female participants from cancer association studies. We detected 131 mosaic events ≥2 Mb in size on the X chromosome of 104 females (0.27%). The rate of mosaicism per base-pair was approximately 5 fold higher for the X chromosome than the autosomes (1.10 vs 0.22 events per 100,000 Mb). Similar to autosomal events, most detected X events were mosaic copy neutral losses of heterozygosity (53%) followed by mosaic losses (34%) and mosaic gains (14%). Contrary to autosomal observations, mosaic X events tend to be larger in size (mean 86.1 vs 44.6 Mb) and far more likely to involve the whole chromosome (49% vs 5%). Interestingly, when compared to individuals with no X mosaicism, individuals with detected X events were observed to have a 15 times increased odds of additionally harboring an autosomal event (p-value= 2.20×10^{-8}). Frequency of X mosaicism ranged from 0.1% in 50 year old women to 0.4% in women 75 years or older (trend p-value=0.0066). While no overall cancer association was observed (p-value=0.30), subtype analyses suggest an association of X mosaicism with increased lung cancer risk (p-value=0.04), an association also observed for autosomal mosaicism, but further studies are required to confirm this finding. Methylation arrays were hybridized for a subset of 10 samples in an attempt to localize events to either the active (X_a) or inactive (X_i) X chromosome. Initial results indicate 8 out of 10 events have a shift in methylation profile favoring more copies of X_i and fewer copies of X_a; however, the observed trend was not significant in our limited sample (p-value=0.11). Our results indicate a higher per base pair rate of mosaicism on female X chromosomes than on autosomes and show that age-related erosion of the genome, previously observed in studies as mosaic events on the autosomes, also occurs on female X chromosomes.

3180T

Array-based analysis reveals partial 11q14 duplication in a familial case with intellectual disability, short stature and mild dysmorphic features. *R. Satomi[1], M. Ohta[1], K. Matsumoto[1], T. Miyashita[1], M. Hayata[1], H. Shiraku[1], N. Matsuda[1], K. Kurosawa[2], K. Enomoto[1].* 1) Department of Pediatrics, JA Toride medical center, Toride, Ibaraki, Japan; 2) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan.

Partial chromosomal duplications of 11q without additional imbalances are rare. A total of seven cases associated with 11q14 duplication without any other chromosomal involvement have been reported. The phenotypes in these previous cases were characterized by a wide variety of clinical findings and variable intellectual disability or psychomotor delay with facial dysmorphic features. This may be due in part to varying size of the duplicated segments. Herein, we report a new case with a 9.6Mb duplication involving 11q14.1 to q14.3 defined by array-based comparative genomic hybridization analysis: arr 11q14.1q14.3(79,131,430-88,758,610)x3 (hg19). The patient is a 9-year-old boy who exhibits developmental delay, intellectual disability, minor dysmorphic features, and short stature. He was born at 32-week gestation. His birth weight and length are 1854g (25-50th percentile) and 45cm (75-90th percentile), respectively. He was healthy and uncomplicated in infancy. Initial visit to our hospital was at age 9, complaining short stature. His height was 119.5cm (below -2.5SD). Physical examinations found that he also had mildly upslanting palpebral fissures and cryptorchidism. Subsequent examinations reveal that his intelligence quotient is 50 and ADHD Rating Scale based on DSM IV criteria is 26. His brother also has intellectual disability, learning disorder, and short stature. Furthermore, we have discovered that he has some relatives presenting similar features on his mother's side. Though, this family history strongly shows as if they have the X-linked recessive inheritance, the copy-number change is even in the autosome. Our observations delineate the phenotypic spectrum associated with a clearly defined duplication of chromosome 11q14. Phenotype-genotype correlation will be discussed in view of all the reported cases and our case's family members.

3181T

15q11.2 duplication encompassing only the *UBE3A* gene is associated with developmental delay and neuro-psychiatric phenotypes. *A. Noor[1], L. Dupuis[2], K. Mittal[3], T. Stockley[1,4], J.B. Vincent[3], R. Mendoza-Londono[2], D.J. Stavropoulos[1,4].* 1) The Hospital for Sick Children, Department of Paediatric Laboratory Medicine, Toronto, Ontario, Canada; 2) The Hospital for Sick Children, Department of Pediatrics, Division of Clinical and Metabolic Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Molecular Neuropsychiatry & Development Lab, Neurogenetics Section, The Campbell Family Brain Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 4) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada.

Due to the presence of low copy repeats (LCRs) in chromosome 15q11-q13 region is prone to recurrent structural rearrangements including deletions, duplications and inversions. This region is also subject to genomic imprinting, thus associated phenotypes show parent-of-origin specific effect. Interstitial duplications of the 15q11-q13 region with maternal imprints are associated with a wide spectrum of neuro-psychiatric disorders including, autism spectrum disorders (ASD), developmental delay, learning difficulties, schizophrenia and seizures. On the other hand, individuals with duplications of the same region with paternal imprints are often phenotypically normal or in rare instances mildly affected. These observations suggest that a dosage sensitive imprinted gene or genes within this region that underline the risk for neuro-psychiatric phenotypes. Recent studies have shown that *UBE3A* is the only gene within this region that is solely expressed from the maternal allele in mature neurons. To date, no case with duplication of only *UBE3A* has been reported, therefore the contributions of this gene to the 15q11-q13 duplication phenotypes remains unclear. We present a female patient with developmental delay in whom we identified a 129 Kb duplication in chromosome region 15q11.2 encompassing only the *UBE3A* gene. The duplication was found to be maternally inherited in the proband. We further tested the segregation of this duplication and it was found to be segregating with a neuro-psychiatric phenotype in four generations. Expression analysis in cultured fibroblast confirmed the over-expression of the *UBE3A* in the proband compared to age-matched controls. Our study shows for the first time the clinical features associated with over-expression of *UBE3A* in humans, and underscores the significance of this gene in the phenotype of individuals with 15q11-q13 duplication.

3182T

A duplication of the *CDKL5* gene identified in a boy with developmental delay with autistic behavior, short stature and microcephaly. *K. Takano[1], T. Nishimura[2], K. Wakui[1], S. Takahashi[3], Y. Inaba[2], T. Kosho[1], Y. Fukushima[1].* 1) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 2) Department of Pediatrics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Department of Pediatrics, Asahikawa Medical University, Asahikawa, Japan.

Mutations in the X-linked cyclin-dependent kinase-like 5 (*CDKL5*) gene [MIM#300203] cause neurological developmental disorders involving early infantile epileptic encephalopathy, atypical Rett syndrome, severe intellectual disability (ID), infantile spasms and epilepsy. Chromosomal microdeletions including *CDKL5* also cause similar neurodevelopmental disorders. Although duplications of the short arm of X chromosome including *CDKL5* have been reported in patients with ID, dysmorphic features and/or epilepsy, the effect of *CDKL5* copy number gains on neurological development remains unclear because of those duplications encompassing other X-linked ID genes. Here, we present a boy with developmental delay and autistic behavior having a microduplication including the only gene *CDKL5* that could affect developmental brain. The patient is a 3-year-old boy, born at 38 weeks of gestation with the weight as 2,226 g (-2.1 SD), length as 46 cm (-1.0 SD), and OFC as 29 cm (-2.8 SD). His mother has normal intelligence and short stature with the height as 145 cm. He had surgical repair of inguinal hernia at the age of 9 month. His motor development was delayed with sitting alone at 7 months and walking independently at 2 years and 1 month. His DQ was 51 at 1 year and 8 months and he presented autistic behavior. He began to refuse food at 2 years and 8 months, and nasogastric tube feeding was started. At 3 years, he showed growth impairment with the height as 84.0 cm (-3.1 SD), weight as 11.5 kg (-1.6 SD), and OFC as 46.5 cm (-2.1 SD). G-banded chromosomes were 46,XY. Chromosomal microarray analysis revealed a 200-kb duplication in Xp22.13 spanning along the most of the genomic region of *CDKL5* which was derived from his mother. This is the first report on a microduplication of *CDKL5*, implying neurodevelopmental effects of increased gene dosage of *CDKL5*.

3183T

Apparent fetal developmental correction of partial monosomy 4 secondary to possible inheritance of a single translocation derivative. *V. Potluri[1], R. Burnside[2], R. Pasion[2], P. Papenhausen[2].* 1) Dynagene, Div Lab Corp America, 3701, Kirby Drive, Houston, TX 77098; 2) Labcorp, 1904 Alexander Drive, Research Triangle Park, NC 27709.

A 13-year-old female was referred for chromosome microarray analysis due to multiple abnormalities which includes scoliosis, hepatomegaly, chronic lymphocytic thyroiditis, oliguria and anuria among others. The whole genome SNP microarray analysis detected an 88 Mb terminal chromosome 4q gain [arr [hg19](4q24q35.2)(103,160,816-190,195,071)x3]. In addition, there was a complete allele homozygosity of the 103 Mb remainder of the chromosome 4 (4pter to q24)(46,691-103,160,815). The large duplicated region of 4q includes numerous OMIM annotated genes (proximal OMIM gene SLC39A8). Chromosome analysis identified the duplicated portion of the terminal 4q fused to Xpter with no DNA lost from Xp by array analysis. The original genomic imbalance could be due to unbalanced meiotic segregation of a parental translocation involving chromosomes 4 and X with subsequent single derivative (3:1) segregation leaving the zygote monosomic for chromosome 4. The observed homozygosity contiguous with the translocated gain of 4q is a strong indication that the single chromosome 4 normal homolog duplicated creating UPD for all of chromosome 4, although the partial monosomy apparently was corrected by this probable pathway, it results in trisomy for the distal half of chromosome 4 that was fused to the X chromosome. Since there is no apparent loss of DNA at Xpter, selective inactivation of the derivative X chromosome (and subsequently the 4q) may not be 100% or may be slow to select against, allowing early gestational effects of the partial trisomy 4. In addition, the 4q segment may not be completely inactivated . Recessive mutations in the homozygotic region of 4 cannot be excluded as also contributing to the abnormal phenotype. Although corrections through trisomy and monosomy rescue are well documented in literature, this apparent correction which results in imbalance is highly unusual, demonstrating the apparent selective disadvantage of monosomy over trisomy.

3184T

Complex mosaic chromosome rearrangement in a patient with Phelan-McDermid Syndrome. *C. Purmann[1,2], W. Froehlich[1,3], J. Bernstein[3], R. Dolmetsch[4,5], J. Hallmayer[1], A.E. Urban[1,2].* 1) Department of Psychiatry & Behavioral Sciences, Stanford University, Palo Alto, CA; 2) Department of Genetics, Stanford University, Palo Alto, CA; 3) Department of Pediatrics, Stanford University, Palo Alto, CA; 4) Department of Neurobiology, Stanford University, Palo Alto, CA; 5) current address: Novartis Institutes for Biomedical Research, Cambridge, MA.

Phelan-McDermid Syndrome is caused by deletions or unbalanced translocations of chromosome band 22q13, for which few mosaic cases have been described. Here, we report a detailed genetic analysis of a patient with a 22q13 terminal deletion showing a complex rearrangement containing three different breakpoints. The affected patient exhibits global developmental delay, moderate intellectual disability, autism, and severely impaired speech. She was initially diagnosed with Phelan-McDermid Syndrome using multiplex ligation-dependent probe amplification (MLPA). Molecular karyotyping of a fibroblast sample using a high-density Illumina HumanOmni5Exome array revealed a complex rearrangement of a 1 Mb region missed by the MLPA assay pointing to two or possibly three different cell populations in the sample. This finding was followed up by digital droplet PCR and microsatellite analysis. Subsequent analyses include Fluorescence in situ hybridization (FISH) as well as targeted chr22 capture followed by deep sequencing for fine-mapping of the breakpoints. The combined approach revealed a complex rearrangement consisting of two different breakpoints of the 22q13 terminal deletion (BP1 and BP2). Furthermore, the patient showed an additional rearrangement between BP2 and a third breakpoint (BP3) pointing to a possible third cell population. To the best of our knowledge, this case represents the first mosaic chromosome rearrangement of this kind in Phelan-McDermid Syndrome. This case also highlights the effective use of established and new laboratory technologies for studying complex chromosomal rearrangements.

3185T

CMIP, a Strong Candidate Gene Involved in the Autism Spectrum Disorder. *M. Luo[1], S. Mulchandani[1], M. Harr[2], T. Wenger[3], E. Zackai[2], R. Schultz[3], N. Spinner[1], L. Conlin[1].* 1) Dept of Pathology & Lab Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Dept of Clinical Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Center for Autism Research, The Children's Hospital of Philadelphia, Philadelphia, PA.

Autism spectrum disorder (ASD) is a group of conditions that involve severe deficiency in language and social communication, along with inappropriate repetitive behaviors. Despite the evidence of high heritability, a clear genetic diagnosis can only be established in a minority of patients. Genetic heterogeneity has been appreciated, and both chromosomal aberrations and monogenic causes have been found. We report two patients with deletions involving the CMIP gene, identified by genome-wide SNP array analysis. Both patients presented with ASD, specifically with pervasive developmental delay and attention deficit hyperactivity disorder (ADHD). Patient 1 has a 517kb deletion within 16q23.2q23.3, including the entire CMIP gene and part of the PLCG2 gene. Patient 2 has a 1.59Mb deletion within 16q23.3 encompassing 12 genes, including part of the CMIP gene and three genes known to be associated with disease (GCSH, BCMO1, GAN). The minimal region of overlap includes only the CMIP gene. Parental studies confirmed that the deletions in both patients were de novo and were not associated with a familial chromosomal rearrangement. These two deletions were observed in a cohort of 11,000 pediatric samples tested for various clinical reasons, with at least 1,000 patients with ASD. In the literature, only one patient with a comparable de novo deletion has been reported at this time. This patient shared very similar phenotypes with our patients. In addition, one 280kb de novo deletion was found in Decipher (258184) and one 720Kb deletion without inheritance information was found in Clinvar (nsv532569). Although no detailed clinical information was available, both of these individuals were reported to have ASD. Association studies showed that the CMIP gene is a susceptibility locus for specific language impairment (SLI), which is a comorbid disorder with ASD. Our findings provide further support for the role of CMIP haploinsufficiency in ASD patients. Additionally, our study also provides evidence for the genetic overlap between SLI and ASD.

3186T

Cytogenetic studies of the drug Methotrexate (MTX) on the blood lymphocytes of colon cancer patients. *Z.MT Jaafar.* Biotechnology Center, Ministry of Science and Technology - Iraq, Baghdad, Baghdad, Iraq.

The aim of the study is to investigate the cytogenetic parameters in order to define the effect of methotrexate (MTX) on blood lymphocytes , manifested by blast genic index (BI) , mitotic index (MI) , replicative index (RI), and sister chromatid exchanges (SCEs). The collected blood samples were cultivated in RPMI 1640 supported with 10% fetal bovine serum , 0.3 ml of phytohaemagglutinin (PHA) , antibiotics (penicillin and streptomycin) and different concentrations of MTX (0,0.2,0.5, 0.5, 1, 2, 4, 8 μg/ml). The results showed significant reduction in BI and MI and RI indices with concentrations (1 , 2, 4, 8 μg/ml) when comparing with the control . While in 0.2, 0.5 μ/ml of MTX gave no significant decrease when compared with the controls . The results of BI shows significant reduction according to increasing of conc. Of drug , it gave 26% in (0)conc. and increase to 19.1% in 0.2 μg/ml and 11% at 2ug/ml and 4.5% in 8ug/ml . The replicative index also show reduction which is proportional with increasing conc. Of MTX , it gave 0.72% at zero conc. and 0.44% at conc . of 0.2 μg/ml of MTX and 0.22% at conc. of 0.5 μg/ml and 0.02% at conc. of 1 μg/ml of MTX . On the other hand the SCE which represent the breakage in the two double helices of DNA shows increases in the SCEs frequency in colon cancer samples which gave 8.69 in zero conc. and 8.81 in 0.2 μg/ml , 10.66 at conc. of 0.5ug/ml . The SCEs in the other conc. of drug can not be detected due to the toxicity of drug in high concentrations . In control samples , the SCEs frequency gave 5.45 at zero conc. , 5.66 at 0.2 μg/ml and 6.56 at conc. of 0.5 μg /ml .The mutation fraction increased in colon cancer patient which gave 8.93 in 0.2 μg/ml , and 15.72 in 0.5 μg/ml and 4.1 in 1ug/ml while in control gave 3.17 in 0.2 μg/ml and 2.17 in 0.5μg/ml and 0.74 in μg/ml . The conclusion showed decrease in the (MI) , (BI) and (RI) and increase in SCEs frequencies and mutation fraction in colon cancer patients when compared with control . The presence of mutant cells in lymphocytes of colon cancer that resistant to MTX, give an idea about amplification in DHFR gene that regulate the activity of DHFR enzyme and also indicate the resistance of cancer cells to drugs .

3187T

Molecular cytogenetic characterization of a patient diagnosed with dimorphic anemia carrying de novo rare ring chromosome 7 along with t(7;9). *S.K. Bhattacharya, V. Lal.* Cytogenetics, Dr. Lal Path Labs. Pvt. Ltd., Newdelhi, Delhi, India.

Cytogenetic abnormalities are described in a few patients with typical dimorphic anemia (DA). The possible clonal nature of this disease is still controversial. We present a case of a derivative chromosome 7 formed by a ring chromosome 7 and t(7;9). The patient is phenotypically normal. This is a unique case report from India, on occurrence of r(7) along with t(7;9). Patient is born to cytogenetically normal parents. We have performed conventional cytogenetic technique using GTG banding and molecular FISH technique using LSI D7S486 probe and WCP for chromosome 7 and 9. However, both the clinical features and the extent of cytogenetic imbalance of chromosome 7 and 9 are variable, and few reported cases of r(7) have been studied by molecular techniques. To our knowledge, this is the first report of r(7) and t(7;9) found in a child with dimorphic anemia, characterized by molecular cytogenetic analysis.

3188T

Reciprocal microdeletions and microduplications of *CNTN6* gene (3p26.3) in patients with intellectual disability. *I. Lebedev[1], A. Kashevarova[1], N. Skryabin[1], N. Chechetkina[2], O. Salyukova[2], L. Nazarenko[2].* 1) Laboratory of Cytogenetics, Institute of Medical Genetics, Tomsk, Tomsk region, Russian Federation; 2) Laboratory of Hereditary Pathology, Institute of Medical Genetics, Tomsk, Tomsk region, Russian Federation.

To date several reciprocal microdeletion / microduplication syndromes or «genomic sister disorders» have been described. Due to application of high resolution microarray technologies new submicroscopic chromosomal rearrangements are being published. The reciprocal rearrangements are particularly valuable, since they allow to determine new dosage-sensitive pathogenic genes in human genome. We performed array-CGH analysis for 80 patients with idiopathic intellectual disability using CGH Microarray Kits 4×44K and 8×60K (Agilent Technologies, USA). Pathogenically significant cases were confirmed by qPCR. We present two patients with microdeletion (369 kb) and patient with microduplication (766 kb) at 3p26.3 containing the only gene - *CNTN6*. The microduplication was inherited from apparently healthy father and grandmother. The brother and sister with microdeletion were orphans. The microdeletion was the same size and localization in both sibs. Recently, the microduplication in 3p26.3 (containing *CNTN6*) has been shown to be associated with autism spectrum disorders. Contactin 6 is also suggested to play a neuroprotective role in ischemic injury and contribute to granule cell maturation and/or synaptic formation in the developing cerebellum. Considering the experiments with mice we found myotonic syndrome, late development of sit and walk ability in the anamnesis, current fine motor skills impairment and dysarthria in patient with dup3p26.3. Dysarthria was also observed in the boy with del3p26.3. Obviously, *CNTN6* can be a novel pathogenic gene associated with autism spectrum disorders, intellectual disability, and motor functions impairment. This study was supported by Russian Scientific Foundation (grant N 14-15-00772).

3189T

Co-occurrence of non-mosaic trisomy 22 and inherited balanced t(4;6)(q33;q23.3) in a live-born female: Case report and review of the literature. *J. Liu[1], F. Kehinde[2], C. Anderson[3], J. McGowan[2], R. Jethva[3], M. Wahab[1], A. Glick[1], M. Sterner[1], J. Pascasio[1].* 1) Dept Path & Lab Med., St Chris Hosp for Children, Drexel University College of Medicine, Philadelphia, PA; 2) Department of Neonatology, St. Christopher's Hospital for Children, Drexel University College of Medicine, Philadelphia, PA 19134; 3) Department of Medical Genetics, St. Christopher's Hospital for Children,- Drexel University College of Medicine, Philadelphia, PA 19134.

Trisomy 22 is the third most common autosomal trisomy occurring in about 0.4% of all clinically recognized pregnancies. Complete non-mosaic trisomy 22 is extremely rare in live births, and most of the affected children died before one year of age. To date, only 29 cases have been reported in the literature, and none of them carried an additional genetic lesion. In this report, we describe clinical presentations, cytogenetics and cytogenomics findings in a live-born female with a complete non-mosaic trisomy 22 as well as a paternally inherited, balanced reciprocal chromosomal rearrangement t(4;6)(q33;q23.3). The proband manifested features commonly seen in individuals with non-mosaic trisomy 22 such as intrauterine growth retardation (IUGR), single umbilical artery, cranial abnormalities, short neck, cleft lip and palate, dysmorphic ears, congenital heart defects, dysplastic kidneys, genital anormalies, hypoplastic nipples, and digital malformations. In addition, she had lobar holoprosencephaly, aqueductal stenosis, limb and eye problems that have not been associated with complete trisomy 22. The proband died 35 days after birth due to complex heart disease and renal failure. We are hereby expanding cytogenetic and clinical spectrum of this rare chromosome disorder. Clinical features of live-born children with non-mosaic trisomy 22 are reviewed and compared to those in our proband. The impact of genomic content in relation with the survival of trisomies in humans is also discussed.

3190T

Identification And Characterization Of Marker Chromosome In a Patient with Turner's Syndrome. *s. sharma, p. srivastava.* indraprastha apollo hospital, New Delhi, new delhi, India.

Introduction: We are describing a 14 year old female who was referred to us with severe growth retardation, primary amenorrhea, poor development of secondary sexual characteristics, web neck, increased FSH and LH. A provisional diagnosis of Turners syndrome was made clinically. On routine cytogenetic analysis she was found to be mosaic for 45,X and 46,X,+marker chromosome. The presence of a marker chromosome in Turner syndrome generally implicates a sex chromosome origin but it may also originate from a non-sex chromosome. If the marker chromosome originates from Y, the patient is at risk of developing Gonadoblastoma. Therefore, FISH test was performed for marker chromosome to rule out the presence of Y chromosome so as to delineate the risk of gonadoblastoma in the patient. Methods: Peripheral blood sample was collected from the proband and cytogenetic examination was carried out using standard techniques of 72hr phytohemagglutinin stimulated peripheral blood lymphocyte culture. FISH was done using centromere enumerating probe for X and Y chromosome. Results: The marker chromosome identified in the chromosomal analysis was found to be centromeric part of X chromosome on FISH analysis, thus ruling out the Y chromosome and the risk of gonadoblastoma. Conclusion: We strongly recommend that FISH should be done in all cases of Turner syndrome where a marker chromosome is identified on chromosomal analysis so as to rule out the risk of gonadoblastoma and to provide appropriate genetic counseling. Keywords: Turner's Syndrome, Marker chromosome, Cytogenetics.

3191T

Cytogenetic Abnormalities in Products of Conception; Analysis and Review of NSUH series of cases. *J. Paul, I. Piechocki, C. Sreekantaiah.* Cytogenetics, North Shore LIJ Health System, Manhasset, NY.

Chromosomal abnormalities account for approximately 50 to 70% of sporadic first trimester miscarriages with the remainder usually classified as idiopathic. A majority of the abnormalities are numerical with structural chromosomal abnormalities or other genetic mechanisms accounting for less than 10% of cases. Cytogenetic analysis of products of conception (POC) is essential to determine the cause of pregnancy loss, aid the prenatal diagnosis of subsequent pregnancies, calculations of recurrence risks, and delineation of the different kinds of abnormalities causing spontaneous abortions. We report our case series of over 5000 POC samples referred to the North Shore University Hospital and compare our data with the published data.

3192T

Haploid-insufficiency and triploid-insensitivity of the same 6p25.1p24.3 region in a family. *Z. Qi[1], L. Jeng[2], A. Slavotinek[3], J. Yu[1].* 1) Dept. of Lab Medicine, University of California San Francisco, San Francisco, CA; 2) Dept. of Medicine, Pediatrics and Pathology, Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD; 3) Dept. of Pediatrics, University of California San Francisco, San Francisco, CA.

Chromosome 6pter-p24 deletion syndrome is a clinically recognized chromosomal disorder (OMIM #612582). Of more than 40 reported cases of this disorder, only four showed interstitial deletions within the 6p25.3p23 region. To date, only two interstitial duplications involving the 6p25.3p23 region have been reported. We present a family carrying both deletion and duplication within the 6p25.1p24.3 region (4,745,144-10,384,769 bp, Build hg19). A 5.6 Mb interstitial deletion within the region was detected by array CGH in a 26-month-old female proband who presented learning disabilities affecting motor and speech skills, bilateral conductive hearing loss, dysmorphic features including hypertelorism, a left preuaricular pit, downturned corners of the mouth and small teeth, a hemangioma of the right neck and a small pectus excavatum. Instead of deletion, her typically developing 5-year-old brother carried an interstitial duplication of the same region. High-resolution G-banding and FISH studies demonstrated that the deletion and duplication were apparently inherited from their 35-year-old mother who carries both cell lines of the deletion (~70%) and duplication (~10%), as well as a normal cell line (~20%) in her peripheral blood. The proband showed phenotypes reportedly associated with the chromosome 6pter-p24 deletion syndrome. Their mother had an embolic stroke at age of 26 years, a patent foremen ovale, and an atrial septal aneurysm. Our findings are consistent with the reported haploid-insufficiency of the 6p25.1p24.3 region, and suggest possible triploid-insensitivity of the same region. In addition, the cells with the duplication may compensate the phenotypic effect of the cells with the deletion as implied by the maternal phenotype and karyotype that was most likely derived from a rare somatic inter-chromatid non-allelic recombination that occurred early in embryogenesis.

3193T

A (2;12)(p12;p13) translocation in a Down Syndrome patient with B-ALL. *Y. Kim[1], P. Traum[2], B. Lasky[2], L. Yang[2], D. Ngo[2], J. Pelkey[2], JN. Rao[2], C.A. Tirado[2].* 1) UCLA Department of Pathology & Lab Medicine 248 Backs Lane, apart B Placentia, CA92870; 2) UCLA Department of Pathology & Lab Medicine 1010 Veteran Ave, 2212F Los Angeles, CA 90024.

Children Children with Down syndrome have a greatly increased risk of developing acute leukemia, including B-Acute Lymphoblastic Leukemia (B-ALL). Herein we present a 9-year-old male Down syndrome patient who presented with pancytopenia, circulating lymphoblasts and increased megathrombocytes. His white blood cell count was 3,010 per μL, the hemoglobin was 8.9 g/dL and the platelet count was 43,000 per μL with a differential showing 7% neutrophils, 1% metamyelocytes, 1% myelocytes, 52% lymphocytes, 32% lymphoblasts and 4% monocytes. A bone marrow biopsy showed 90% marrow cellularity, with 95% of replacement by lymphoblasts. Flow cytometry of the bone marrow aspirate showed the blasts were positive for CD10, CD19(dim), CD22(dim), CD34(bright), CD38(dim, partial), HLA-DR, CD16(subset), and CD79a. Based on these clinical and pathologic findings, a diagnosis of B-ALL was made. Cytogenetic analysis revealed an abnormal karyotype described as 47,XY,+21c[25]/47,idem,t(2;12)(p12;p13),t(8;14)(q11.2;q32)[5] indicating two reciprocal translocations , one at 2p12; 12p13 and the second one at 8q11.2;14q32. The *IGH@* rearrangement seen in the second translocation was confirmed with FISH using the IGH@ break-apart probe. A FISH panel for *ETV6,RUX1* revealed an abnormal signal pattern in 51% of the cells examined with an extra signal for *ETV6* on 12p which confirmed the (2;12) translocation and all cells examined showed an extra signal for *RUNX1* which correlates with the constitutional +21. The t(8;14)(q11.2;q32) involving the *IGH@* gene has been observed as a rare abnormality in CD10+ B-ALL and more than 1/4 of these cases are Down syndrome patients . However to the best of our knowledge, the 2;12 translocation has been reported in cyclin-D1 negative/cyclin-D2 positive mantle cell lymphoma but not in ALL. Therefore the exact significance of this unique finding in this patient is unclear. Further case reports of similar cytogenetic abnormalities in patients with Down syndrome presenting with B-ALL is warranted.

3194T

An uncommon coincidation of multiple myeloma and myelodysplastic syndrome. *S. Berker Karauzum[1], C. Aydin[1], O. Salim[2], O. Kemal Yucel[2].* 1) Department of Medical Biology and Genetics, School of Medicine, Akdeniz University, Antalya, Turkey; 2) Department of Hematology, School of Medicine, Akdeniz University, Antalya, Turkey.

Multiple myeloma (MM) is a subgroup of disease called plasma cell dyscrasias and characterized by clonal proliferation of neoplastic plasma cells. The most common genetic abnormalities that cause multiple myeloma, are translocations which affect IgH chromosome locus on chromosome 14. This translocations are determined by using FISH methods in general. Advanced stage MM requires treatment of chemotherapy. Conventional chemotherapy is known to be an important factor to transformation of MM to MDS or AML. In this report, we present 48 years old case who was first diagnosed with MM in 2012 and our patient's karyotype was found to be normal (46, XY), but in FISH analysis t(11; 14)(q13; 32) translocation was observed in 70%. In addition to this abnormality 13q14, 14q32, 16q22 deletions were also identified. These results supported the MM diagnosis as a sole anomaly. The occurrence of the t(11; 14) and 20q together has marked the coincidence of MM and MDS in this patient. In december 2014 as t(11; 14) translocation was identified by FISH analysis in 19% , in conventional cytogenetic analysis revealed deletion of the 20q as a sole anomaly.

3195T

An easy cIg-FISH Protocol For Multiple Myeloma which can be incorporated as Routine Cytogenetic Laboratory Practice. *L. Gole[1], A. Lin[2], C. Chua[1], W.J. Chng[2].* 1) Department of Laboratory Medicine, National University Health Systems, Singapore, Singapore; 2) Department of Haematology Oncology, National Cancer Institute of Singapore.

The International Myeloma Working Group and the European Myeloma Network recommends that FISH findings in multiple myeloma should not be reported without either first concentrating the plasma cells (PC) or employing some means of PC identification so that only these cells are scored, making the patient report more meaningful. The more commonly practiced techniques seem to be staining plasma cells with lambda/kappa antibodies followed by FISH (cIg-FISH). We have made modifications to the current techniques making it simpler for cytogeneticists to incorporate this as a routine procedure. Bone marrow aspirates from 20 patients for conventional cytogenetics and FISH for myeloma panel were included in this study. Bone marrow cell pellets in Carnoy's fixative were washed with 100% methanol and dropped onto slides. An additional drop of 96% ethanol was applied before drying. Antigen retrieval was performed for 10 minutes in 10mM citrate buffer (0.1M citric acid / 0.1M sodium citrate) at 95°C. Slides were rinsed twice in 1x phosphate buffered saline (PBS) and incubated with 1:10 AMCA goat anti-human kappa and/or anti-human lambda antibodies at 37°C for 1 hour followed by two washes with 1xPBS. They were then incubated with 1:10 AMCA anti-goat antibody at 37°C for 1 hour. Standard FISH protocols were applied following antibody staining. The three FISH probes used were for t(4;14)(*FGFR3;IGH*), t(14;16)(*IGH;MAF*) and *TP53*. Slides were mounted with VECTASHIELD mounting medium without DAPI and viewed using a fluorescence microscope. AMCA-positive cells which stained blue were identified using a triple-bandpass filter and only these were scored. Our modified method showed clear antibody staining along with analyzable bright FISH signals. Plasma cells were well separated and resembled the larger nuclei obtained by routine FISH. All 20 samples had a reportable result and abnormalities were detected in 9 cases by cIg-FISH vs 7 by FISH and 5 by karyotyping. A total of 16/60 abnormalities were detected by cIg-FISH with the 3 probes. This modified method did not miss any abnormalities, produced clear results and increased confidence in setting tighter cutoff limits.

3196T

Homozygous Deletion of TEL (ETV6) in Childhood Acute Lymphoblastic Leukemia (ALL): Prognostic Implications. *G. Velagaleti, C. Mendiola, J. Rodriguez, M. Zhu, V. Ortega.* Dept Pathology, Univ Texas Hlth Sci Ctr, San Antonio, TX.

The fusion of TEL/AML1 (ETV6/RUNX1) gene regions resulting from the t(12;21), is the most common genetic abnormality observed in childhood ALL. The neoplasm in these patients persists not by the initial fusion event, but in most cases by a secondary or subsequent leukemogenic event involving deletion of the TEL gene region on the non-rearranged TEL allele. Although variant patterns are seen by fluorescence in situ hybridization (FISH) including extra copies of AML1, extra fusions of TEL/AML1 and rearrangements of TEL, deletion of the second TEL allele is the most common and can have significant prognostic implications. In these cases, persistence of the original pre-leukemic clone may indicate proliferative advantage and lead to prolonged remission followed by periods of relapse as the residual secondary clones emerge. We report a case on a 26 year old male with ALL harboring homozygous deletion of the TEL gene region. To our knowledge, this is the first reported case of a homozygous TEL deletion without a concomitant TEL/AML1 fusion. His initial presentation to our laboratory was in 2003 status post bone marrow transplant and showing normal result by chromosome analysis (CA). Nine years later (07/2012) the patient returned with relapse. CA showed a complex karyotype with multiple abnormalities. FISH analysis with ALL panel showed 3 copies of cMYC gene region, deletion of p16 gene region and homozygous deletion of TEL gene region with 6 copies of AML1. Follow-up chromosome and FISH studies up until 01/2014 were all normal. Homozygous deletion of the TEL gene region in our patient may suggest that absence of TEL gene region on both alleles behaves similarly to cases with the classic TEL deletion of only one allele. Furthermore, the manifestation of relapse following remission clearly supports TEL deletion being the principle causative factor for subsequent and late-onset relapses by its ability to evade complete elimination during therapy. Since our patient presented to us following BMT, we cannot ascertain if the homozygous TEL deletion, which was never before reported, is the second hit following the initial TEL/AML1 fusion, or is by itself the primary hit. We hypothesize that our patient had the initial TEL/AML1 fusion with deletion of one TEL allele at diagnosis and this in turn may have resulted in the loss of second TEL allele similar to loss of heterozygosity in other cancers.

3197T

A maternally inherited 697.4 kb *SOX3* duplication in a female fetus with neural tube defects. *N. Cohen[1], H. Mei[1], E. Chan[2], M. Babcock[1], J. Reiner[1], A. Li[2], A. Babu[1], S.A. Scott[1], L. Edelmann[1].* 1) Mount Sinai Genetic Testing Laboratory, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 2) New York-Presbyterian Lower Manhattan Hospital, New York, NY 10038.

The SRY-BOX 3 (*SOX3*) gene [MIM 313430] at Xq27.1 belongs to the highly conserved SOXB1 subfamily of transcription factors, which are expressed throughout the developing central nervous system and are involved in embryonic development and neuronal differentiation. Disruption of *SOX3* by polyalanine expansion has been reported among patients with X-linked mental retardation and growth hormone deficiency, and full gene duplication of *SOX3* is a cause of X-linked hypopituitarism with incomplete penetrance. In addition, structural rearrangements involving *SOX3* have been described in patients with XX male sex reversal, and increased expression of *SOX3* by gene duplication has very recently been proposed as a risk factor for neural tube defects (NTDs). Here we report on a female fetus with a maternally inherited submicroscopic duplication of the X chromosome containing the *SOX3* gene. Amniocentesis was performed at 17 weeks due to an abnormal maternal serum alpha fetoprotein (MS-AFP) consistent with an increased risk of NTD. Amniotic fluid AFP (AF-AFP) was also elevated at 60.46 ug/ml and acetylcholinesterase was detected, also consistent with open NTDs. Although chromosome analysis showed a normal 46,XX karyotype, array comparative genomic hybridization (aCGH) using the 180k CGH+SNP (Agilent Technologies) microarray revealed a 697.4 kb duplication on Xq27.1 between 139.1-139.8 (hg19) that includes *SOX3* and exons 1-3 of the long non-coding RNA, *LINC00632*. This duplication was inherited from an apparently phenotypically normal mother (G2P0), with a history of miscarriage and a family history of open NTDs. Ultrasound performed at 19 weeks of gestation showed bilateral ventriculomegaly and the pregnancy was terminated at 21 weeks. This case report provides additional support that overexpression of *SOX3* due to gene duplication may be a risk factor for NTDs and possibly additional brain malformations.

3198T

Phenotype/genotype effect of 14q32.3 terminal deletion and 9p duplication. *S.Ghareeb. Abd Allah, M.O. El Ruby, S.A. Hammad, A.K. Kamel, A.M. Mohamed.* Human Cytogenetics, National Research Centre, Cairo, Egypt.

Molecular cytogenetic studies with thorough clinical evaluation are needed for proper genotype phenotype correlation in specific chromosome deletion and duplication. We reported on a female patient aged 7 years , she had severe intellectual disability (ID), developmental delay, epilepsy. She had characteristic facies with high arched eye brows, epicanthic folds, hypertelorism, broad nasal bridge, broad nose, low set protruded ears, short hands, clinodactyly of little fingers, right semian cease , broad big toe, dysplastic nails, wide spaced hypo pigmented nipples, umbilical hernia, mild hypotonia, normal Echo, normal pelvi-abdominal ultrasound, EEG revealed right central epileptogenic activity. Cytogenetic study on periphersl blood lymphocytes revealed add(14)(q). Whole chromosome paint for chromosome 14 indicated that the add material to 14q was not originated from chromosome 14.Using total subtelomere probes we found deletion of chromosome 14q subtelomere and duplication of chromosome 9p , using locus specific probes 9p21 revealed duplication 9p. Both parents had normal karyotype. This de novo deletion/duplication occurred as a result of unbalanced translocation. All the clinical findings are typical of 14q32 deletion syndromes. Also there is an overlapping and similarity between 14q deletion syndrome and 9p trisomy syndrome like ID, developmental delay, the depressed nasal bridge, epicanthal folds, semian crease and wide spaced nipples. All manifestations of 14q32 deletion syndrome related to critical small region in 14q32 and also the same of the clinical manifestations of 9p trisomy dependent on the duplication of 9p22. The combined effect of these two chromosomal abnormalities may be the cause of the severe ID.

3199T

Two new cases of chromosome 7p22.1 microduplication detected by array CGH. *R.G. Hutchinson[1], J. Nicholl[1], L. Montgomery[1], C. Barnett[2], K. Boundy[3], S. Yu[1].* 1) Cytogenetics, SA Pathology, Adelaide, South Australia, Australia; 2) Clinical Genetics, SA Pathology at WCH, King William Rd, North Adelaide; 3) Neurology Department, TQEH, Woodville.

Array CGH has enhanced our ability to detect rare microduplications previously too small to be observed cytogenetically. Cytogenetically detectable aberrations are necessarily large and the clinical features associated with them have been variable. The use of microarray technology has enabled the phenotypic features more generally associated with duplication of a chromosomal region to be assigned to smaller chromosomal areas and even specific genes. Microarray also detects abnormalities where the pathogenicity of the change may be unclear. Interpretation of these rare abnormalities relies on the cooperation of institutions involved in genetic diagnosis in sharing and comparing cases where the anomaly is similar. Two cases of 7p22.1 microduplication (1.7Mb and 1Mb) detected by array have been reported in the literature. [1,2] We present two patients with 7p22.1 microduplications for comparison with the reported cases. Our cases include a 4Mb (case 1) and a 700kb (case 2) microduplication . The 700kb duplication is almost entirely contained within the 1mb duplication reported by Prelksaitene et.al. (2012) [2], while both published cases are encompassed within the larger duplication. Brief clinical descriptions of our cases are as follows. Case 1. A de novo 4Mb duplication was detected in a 2 1/2 year old girl with global developmental delay particularly language, autistic features, dysmorphic facies and renal pelvis dilatation. Case 2. A 700kb duplication was detected in a 22 year old male reported to have intellectual disability with language delay and dysmorphic features including macrocephaly. Parental studies are in progress to determine if this finding is de novo. . The clinical features common to both reported cases for this duplication include abnormal craniofacial development, developmental delay and skeletal abnormalities. The 7p22.1 region has been suggested as a candidate region for autism [3] with RNF216L highlighted as a candidate gene. This gene has been shown to be strongly expressed in adult and fetal brain. RNF216L was duplicated in both case 1 and 2. Our cases will be compared in more detail with those reported in the literature. 1.Chui J.V. et.al.(2011) Am J Med Genet 155A:2508-2511 2. Prelksaitene et.al (2012) Am J Med Genet 158A:1200-1203 3. Bayou et.al. (2010) J Biomed Biotechnol 2010:423893. doi: 10.1155/2010/423894.

3200T

Partial Microduplication of PTEN in a girl with multiple congenital abnormalities including agenesis corpus callosum. *C.P. Oliveira[1, 2], R.S.H.G Silva[2], R.O.A. Benicio[2], M.T.O. Cardoso[2], S.F Oliveira[3], A. Pic-Taylor[3], J.F. Mazzeu[4].* 1) Programa de Pós Graduação em Ciências Médicas, Faculdade de Medicina, Universidade de Brasilia (UnB), Distrito Federal, Brasil; 2) Núcleo de Genética (NUGEN), Hospital de Apoio (HAB), Secretaria de Saúde, Distrito Federal, Brasil; 3) Departamento de Genética e Morfologia, Instituto de Ciências Biológicas, Universidade de Brasilia (UnB), Distrito Federal, Brasil; 4) Programa de Pós graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasilia (UCB), Distrito Federal, Brasil.

The corpus callosum is the largest commissure connecting the cerebral hemispheres. Agenesis corpus callosum (ACC) is one of the most prevalent brain malformations with an incidence of 0.5-70 in 10,000, and can occur isolated or as part of many syndromes. Although ACC is predominantly genetic, few genes have as yet been identified and chromosome abnormalities have been described in a large number of patients. We report a patient with multiple congenital malformations including agenesis corpus callosum, asymmetrical face, microcephaly, hypertelorism, microphtalmia and short neck. She has two accessory nipples, asymmetric limbs, sacral dimple and a sacral hemangioma. ECG showed interatrial communication. To screen for chromosomal imbalances high-resolution array Genomic Hybridization Assay was performed using the Affymetrix® CytoScan™ HD platform. Analysis revealed a 49 Kb de novo microduplication 10q23.31 (89,629,908-89,679,426 - hg19) including exon 2 of PTEN gene. The phosphatase and tensin homolog (PTEN) gene on chromosome 10 was initially identified as a tumor suppressor gene that is frequently mutated in human cancers. Further studies showed that PTEN plays an important role in brain development involved in cell migration, cell number and cell size regulation. Clinical manifestations of inherited PTEN mutation occur with highly variable penetrance and also include benign hamartomas in multiple tissues, macrocephaly, seizures, ataxia, mental retardation and autism in addition to cancer predisposition. These disorders add to the already well-established role of PTEN mutations in causing the PTEN hamartoma tumor syndromes which include Cowden, Bannayan-Riley-Ruvalcaba, Lhermitte Duclos, and Proteus Syndromes. The patient here reported shows a complex phenotype that does not correspond to any of the syndromes listed above and includes ACC. Thickening of corpus callosum have been reported in patients with Bannayan-Riley-Ruvalcaba syndrome. Our report describes the first de novo PTEN microduplication in a patient with ACC. This finding suggests that PTEN may warrant further consideration as an ACC candidate gene. Financial support: UnB, CNPq, FAPDF.

3201T

Is it time to retire the standard 15colonies/20 cell chromosome analysis if chromosome microarray analysis (CMA) is concurrently ordered in prenatal testing? *A. Patel, J. Smith, A. Breman, P.A. Ward, S. Peacock, T. Vaughn, P. Hixson, W. Bi, C. Bacino, S.W. Cheung.* Dept Molecular & Human Gen, Baylor Col Medicine, Houston, TX.

The present standard of care for prenatal cytogenetic testing is AneuVysion FISH and a complete chromosome analysis. We present our experience with concurrently ordered CMA and a limited chromosome analysis of 5 cells. A total of 353 prenatal samples (amniotic fluid, CVS, cultured cells) submitted between 3/2012 and 3/2014 were analyzed on a custom-designed 180K or a 400K oligonucleotide array manufactured by Agilent Technologies and a 5 cell chromosome analysis performed. Of those 353 CMA/limited chromosome analyses, 90 (25%) showed a copy number variation (CNV) by CMA: 21 (6%) cases showed CNVs >10Mb that were consistent with chromosome analysis, 35 (10%) cases had a smaller familial CNV, 15 (4%) cases showed a CNV in an autosomal recessive locus, three cases (0.8%) showed regions of absence of heterozygosity suggesting consanguinity, and in 9 (2.5%) cases a copy number variant of unknown significance (VOUS) was identified. A VOUS is defined as a CNV < 5Mb in a non syndromic region with multiple genes, and either inherited or one or both parents were not available. Three cases showed a CNV in regions that are associated with variable expressivity (16p11.2, 16p13.11 and 15q11.2 - BP1-BP2). The losses in 16p11.2 and 16p13.11 were de novo, and the loss in BP1-BP-2 was paternally inherited. In addition 4 cases (1%) with normal cytogenetics showed CNVs <10Mb in clinically significant regions and were classified as abnormal: 16p11.2 deletion (2 cases), 14qter deletion and duplication of 22q11.2 (22q11.2 duplication syndrome). Two cases of balanced translocations were detected by chromosome analysis which could not be detected by CMA. Also, cytogenetics detected three cases with an abnormal colony or cell in the first 5 cells; when counts were extended, it was determined that the abnormality was pseudomosaic and of unlikely clinical significance. In this cohort, the CMA/limited chromosome analysis provided a comprehensive cytogenetic diagnosis with no apparent decrease in sensitivity compared to a full chromosome analysis. In the two cases with balanced rearrangements, the 5-cell analysis was sufficient for interpretation. Our data shows that CMA should be the first line test for prenatal cytogenetic testing with limited chromosome analysis (5 cells) as an adjunct to identify balanced rearrangements and distinguish between translocation and trisomy Down/Patau syndromes and to confirm mosaicism.

3202T

Pigmentary mosaicism with 45,X and an extra marker containing the Xp11.22-p11.23 region. *P. Pérez-Vera[1], V. Ulloa-Avilés[1], E. Lieberman-Hernández[1], C. Durán McKinster[2], S. Gómez Carmona[1], R. Cruz-Alcívar[1], M.P. Navarrete-Meneses[1], A. Reyes-León[1], V. Del Castillo-Ruíz[1], C. Salas-Labadía[1].* 1) Genética Humana, Instituto Nacional de Pediatría, Mexico, Distrito Federal, Mexico; 2) Dermatología, Instituto Nacional de Pediatría, Mexico, Distrito Federal, Mexico.

Pigmentary mosaicism is characterized by hypo and/or hyperpigmented macules following Blaschko lines. These lesions are strongly associated with the chromosomal mosaicism. We have studied a cohort of 52 patients under these clinical criteria. We present a patient with 2 abnormal cell lines, one with 45,X and the other with a small supernumerary marker chromosome. The proband is a 2 year and 7 months old female, first child of healthy non-consanguineous parents. She has psychomotor delay and suffers from epileptic seizures. Height is 88 cm (PC<25), weight 12.6 kg (PC>50), and head circumference 58.8 cm (PC 95). She has macrocrania, high hair implantation, prominent forehead, hypertelorism, telecanthus, left dacryostenosis, nasal bridge with broad root, bulbous nose and anteverted nares. She presents hyperpigmented lineal macules in arms and legs following Blaschko lines, and a dorsal inverted V macule. Cytogenetic analysis with GTG bands in lymphocytes revealed mos 47,XX,+mar[28]/45,X[6]/46,XX[16]. In fibroblasts from both types of skin, two cell lines were detected; light skin: mos 45,X[35]/47,XX,+mar[2]/46,XX[6], dark skin: mos 45,X[20]/47,XX,+mar[3]/46,XX[13]. The origin of the marker chromosome was determined by FISH using DXZ1, p11.23-p11.22 (49687273-49865097) and q11.1-q11.2 probes. The marker chromosome comprised Xp11.23-p11.22: ish der(X)(DXZ1+,49687273-49865097+). Based on this result, cytogenetic analysis in lymphocytes from the mother was performed revealing: mos 45,X[2]/47,XXX[1]/46,XX[181]. The partial trisomy observed in our patient resembles to that found in Xp11.22 duplication cases, who have neurological features such as intellectual discapacity and a typical EEG pattern. Our patient shares these characteristics, however, she has other dysmorphic features besides those compatible with her Turner syndrome karyotype. The mosaic condition and the unidentified regions including in the extra chromosome could influence the phenotype. SNPs array analysis must be performed to complete the genotype-phenotype correlation. Acknowledgements: CONACYT Salud-2012-01-182277.

3203T

The identification of ring 13 chromosome and breakpoint region at a Brazilian child requires karyotype, FISH and microarray analysis. *I.P. Pinto[1,2], L.B. Minasi[1,2], A.V. Melo[2,3], J.G. Almeida[1,2], D.M.C. Cunha[1,2], C.L. Ribeiro[2], G.P. Silva[2], M.G Brasil[4], D.M. e Silva[1,2,5], C.C. da Silva[1,2,3,6], A.D. da Cruz[1,2,3,6].* 1) Pontifical Catholic University of Goias, Genetics Master's Program; 2) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program; 4) Clinical Hospital of Federal University of Goias; 5) Federal University of Goias, Genetics and Molecular Biology Master's and PhD Programs; 6) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health.

Deletions of chromosome regions are generally the result of double strand chromosome breaks with loss of produced acentric fragments during the next cell division, and could lead to ring chromosome. Patients with r(13) presented different phenotypic abnormalities according to different breakpoints. Here, we report the first postnatal diagnosis in Central Brazil of a girl with cytogenetic abnormalities involving chromosome 13 using 3 different laboratory methodologies. At the age of 8 months, she presented low tubular renal function in both kidneys, craniofacial dysmorphisms, hypotelorism, bilateral microphthalmia, deep set eyes with reduction and change in shape of the eyeballs, atrophy of the optic nerves and chiasm, and epicanthal folds. Magnetic Nuclear Resonance (MNR) revealed decreased cervical and axial tonus, reduced fontanelle, lack of visualization of the septum pellucidum with persistent of the cavum septum pellucidum and cavum Vergae, atrophic hippocampus, absence of falx cerebri, partial fusion of the thalami with an appearance suggestive of holoprosencephaly. She also had hands with elongated fingers and disproportionate in size, and heart problems. At the age of 6 years, physical examination of the proband reveled severe delayed psychomotor development and intellectual disability. Cytogenetics analysis revealed a constitutive 46,XX,r(13)[77]/45,XX,-13[17]/46,XX,idic r(13)[6]. FISH analyses also showed the absence of 13qter and the presence of 13q14.3 in the cells with ring chromosome 13, and it was also observed the presence of an isodicentric chromosome. Chromosome Microarray Analysis (CMA) with Affymetrix CytoScan™ HD Array detected a *de novo* 15.39Mb deletions at 13q32.3-q34. This deletion involved 44 morbid genes from OMIM. Both monosomy and isodicentric were confirmed with FISH, which was also useful to determine the deletion of the terminal end of the long arm of chromosome 13 in the ring formation. However, banding karyotyping and FISH were not able to precisely define the breakpoint of the terminal deletion. It was evident that the ring chromosome led to a partial deletion of the long arm of chromosome 13. Haploinsufficiency of the genes within the deleted region is the most probable cause of the proband phenotype. Thus, based on probe density, high-resolution CMA made it possible to refine the breakpoint region for the affected chromosome 13 and also it was useful to determine the gene content within the deleted region.

3204T

Proximal 3p deletions: phenotypic characterization and molecular delineation. *I. Song[1], D.H. Lee[1], H.W. Yoo[2], J.O. Lee[3], E.J. Seo[1].* 1) Department of Laboratory Medicine, University of Ulsan College of Medicine and Asan Medical Center, Seoul, Korea; 2) Department of Pediatrics, University of Ulsan College of Medicine and Asan Medical Center, Seoul, Korea; 3) Asan Institute for Life Sciences, Seoul, Korea.

Proximal interstitial deletion of chromosome 3p is a rare intrachromosomal rearrangement known to cause neurodevelopmental delay and intellectual disability, and associated with various clinical manifestation such as autistic features, Waardenburg syndrome type 2, Larsen syndrome and Zimmermann-Laband syndrome. Recently, widespread implementation of chromosomal microarray in diagnostic testing has brought great advances in identifying genomic disorders and studies about microdeletion of relevant genes are being made to delineate genetic basis of spectrum of clinical manifestation. We report two cases of an interstitial deletion of proximal 3p that have been characterized by high-resolution array CGH. Both patients presented abnormal auditory brainstem response or ear canal obstruction, and heart anomalies like secundum atrial septal defect and patent ductus arteriosus. The first patient had a 16 Mb deletion at 3p14.1-p12.1. He had small mouth, epiblepharon, high arched palate and genital malformations: small scrotum and bilateral cryptorchidism. At 21 months of age he couldn't stand or walk alone and could only just babble. His brain magnetic resonance scan showed vermian hypoplasia and enlarged retrocerebellar CSF space. The second patient had a 4 Mb deletion at 3p14.1-p13. She was noted to have heterochromia iridis, strabismus, ulnar deviation of wrist and digit anomaly with overlapping fingers. There was no family history of congenital anomalies or intellectual disability. The genomic characterization and clinical findings in these patients were compared to those in public databases and the medical literature: OMIM, DECIPHER, ISCA, and recently published reports. Within the deleted region, FOXP1 and MITF genes are known to involve in nerve cell regulation and neurodevelopment, and PROK2 known for its mutation causing Kallmann syndrome, is associated with cryptorchidism. This study suggests that proximal 3p deletions show no common breakpoint or deletion size and haploinsufficiencies of those genes play a major role in these clinical phenotype.

3205T

A novel familial gain of 3q25.2-3q25.31 involving OMIM genes SLC33A1, GMPS, and MME. *K. Swisshelm[1], M. Haag[1], D. Hennerich[1], P. Brzeskiewicz[1], B. Lunt[1], K. Ha[1], J. LeRoux[1], G. Bellus[1,2], T. Schreiner[1,2].* 1) University of Colorado Denver, Denver, CO; 2) Children's Hospital of Colorado, Aurora, CO.

Four siblings, exhibiting various phenotypes, were studied for chromosomal deletions and duplications by chromosomal microarray (CMA), using a 180K oligo platform. The first sib, a 7-year-old male, showed phenotypes of developmental delays, dysmorphic facial features consistent with fetal alcohol syndrome (FAS), mild intellectual disability, and low muscle tone. The second sib, a 6-year-old female, showed dysmorphic facial features not consistent with FAS. The third sib, a 4-year-old male, exhibited a phenotype similar to sib 1, and the fourth, a 3-year-old male, exhibited autism, seizure disorder, and an abnormal EEG in addition to dysmorphic facial features consistent with FAS. The female sibling (sib 2) had a normal CMA result. However, the three male sibs' genomic DNA all showed a gain of 1.1 - 1.5 Mb in 3q25.2-3q25.31, which contained three OMIM disease genes: SLC33A1 (associated with autosomal recessive congenital cataracts, hearing loss, and neurodegeneration), GMPS (associated with a case of somatic acute myeloid leukemia), and a partial overlap with MME (associated with antenatal membranous glomerulonephritis). Sibling 4 also showed a partial duplication of KCNAB1 (potassium voltage-gated channel, shaker-related subfamily, beta). This latter gain may contribute to the patient's seizure disorder. This particular duplication has not been reported in the literature, although a smaller de novo duplication has been reported in DECIPHER, detected in a patient with intellectual disability and tics. To date there has been no parental follow-up testing and the clinical significance of this familial copy number gain is uncertain.

3206T

Position effects modify gene expression in a ring chromosome 14? *R.S. Guilherme[1], A.G. Dantas[1], M.M Oliveira[1], V.A. Meloni[1], D. Brunoni[1], A.T. Paes[2], L.D. Kulikowski[3], M.I. Melaragno[1].* 1) Morphology and Genetics, UNIFESP, Rua Botucatu 740, CEP: 04023-900, São Paulo, Brazil; 2) Setor de Estatística Aplicada, UNIFESP, Rua Diogo de Faria, 1087, CEP: 04023-900, São Paulo, Brazil; 3) Department of Pathology, Laboratório de Citogenômica, Universidade de São Paulo, Avenida Dr. Enéas Carvalho de Aguiar 647, CEP: 05403-000, São Paulo, Brazil.

Ring chromosomes (r) usually originate from breakage in their short and long arms followed by fusion of the broken ends resulting in genetic material loss. The most important factor affecting the phenotype of patients with rings is the chromosome involved in the rearrangement and the extension of the deleted region. We present a patient with an apparently complete r(14) characterized by G-banding, genomic array, FISH (Fluorescent in situ Hybridization) and MLPA (Multiplex ligation-dependent probe amplification). Although the r(14) did not have genomic loss, except by the telomere, the patient presents the main characteristics of the syndrome such as seizures, mild intellectual disability, microcephaly, dolichocephaly, downslanting palpebral fissures and bilateral cryptorchidism. Our aim is to verify if the ring configuration affect the expression of genes in two copies in the patient's genome, causing phenotypic alterations. Gene expression of eight genes was assessed using TaqMan assays, in peripheral blood from the patient and seven controls, matched by age and sex. qPCR experiments were run in triplicate and normalized by two endogenous genes (GAPDH and ACTB) using 2-$\Delta\Delta$Ct analysis method. PTGER2, RPS6KA5, RCOR1 and VRK1 genes presented normal level of expression. ARID4A, GPHN, RCOR1 and ACTN1 genes were significantly downregulated in the patient compared to controls. The website GeneMania (www.genemania.org) used to verify gene interaction network showed that, except for GPHN gene, all the other genes studied are co-expressed. We propose that these downregulated genes may affect the expression of other genes explaining the patient's phenotype. One mechanism proposed to explain ring syndrome phenotype is telomere effect position caused by the ring formation that changes the chromatin architecture and consequently the gene expression. The juxtaposition of the DNA from p-arm inactive, centromere and subtelomeric region close to the q-arm active euchromatin could have silenced the RCOR1 gene and others in 14q32 due to inactivation spreading. However, the other downregulated genes (ARID4A, GPHN and ACTN1) may have been silenced by other mechanisms as the repositioning of the ring chromosome in the nucleus that could change cis and/or trans interactions. Thus, the chromatin architecture needs to be taken into account in the evaluation of the genetic consequences of complete rings with clinical phenotype. (Financial support, FAPESP, Brazil).

3207T

Partial trisomy 17q and partial monosomy 20q in a boy with craniosynostosis. *F.A. Marques[1], R.S. Heredia[2], C.P. de Oliveira[2,3], M.T.O. Cardoso[2], J.F. Mazzeu[1,3], R. Pogue[1].* 1) Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, SGAN 916, Asa Norte, Brasília, 70790-160-DF, Brazil; 2) Núcleo de Genética da Secretária de Saúde do Distrito Federal, Hospital de Apoio, SAIN Qd4, 70071-125-DF, Brazil; 3) Programa de Pós-Graduação em Ciências Médicas, Universidade de Brasília, Campus Darcy Ribeiro, Caixa Postal 4569, 70910-900, Brasília-DF, Brazil.

Craniosynostosis is defined as a premature fusion of at least one cranial suture, which can be accompanied by other phenotypes. Of syndromic cases, 14-22% have been associated with chromosomal rearrangements. This report describes a Brazilian boy with syndromic craniosynostosis who also presented with mental retardation, microcephaly, frontal bossing, bitemporal narrowing, short neck, syndactyly and cardiac defects. Chromosome banding showed an apparently normal male karyotype. Subsequent chromosomal microarray analysis (CMA) usingAffymetrix CytoScan 750K Array showed a duplication of 2.1 Mb on chromosome 17q and a deletion of 1.4 Mb on chromosome 20q. The data suggested an unbalanced translocation, which was confirmed by Fluorescence in situ hybridization analysis (FISH). While there are several reports in the literature of chromosome 17q duplication syndrome accompanied by partial monosomies of other chromosomes, this is the first case featuring partial monosomy of 20q. The patient's phenotype is generally consistent with 17q duplication syndrome, however the craniosynostosis has rarely been associated with this chromosomal anomaly. Thus, CMA provides a sensitive, rapid and powerful technology for a ''genotype-first'' approach in conditions such as craniosynostosis that are difficult to associated with a specific genetic defect based on clinical features alone.

3208T

Meiotic I error in a Thai girl with tetrasomy 9p syndrome identified by SNP microarray. *C. Charalsawadi[1], S. Jaruratanasirikul[2], J. Wirojanan[2], V. Praphanphoj[3], S. Puangpech[1], K. Jarauthamsophon[1], P. Limprasert[1].* 1) Department of Pathology, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla, Thailand; 2) Department of Pediatrics, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla, Thailand; 3) Rajanukul Institute, Bangkok, Thailand.

Tetrasomy 9p is a rare chromosomal disorder caused by the presence of an extra isochromosome consisting of the short arm of chromosome 9. The syndrome has a wide variation of phenotypic features, ranging from stillborn babies with multiple malformations to normal healthy adults. We report the case of a 4 year-old Thai girl with mild developmental delay and no major physical malformations. The girl also presented with pilomatricoma on the abdominal wall and skin pigmentary abnormalities. G-banding karyotyping of cultured lymphocytes revealed 47,XX,+mar in all analyzed cells. The Illumina HumanCytoSNP-12 v2.1 array was carried out on uncultured peripheral blood DNA and revealed an elevated logR ratio and an altered pattern of B allele frequency, consistent with tetrasomy of the short arm of chromosome 9. A FISH assay using probes specific to chromosome 9p and 9q subtelomeres confirmed the presence of 4 copies of chromosome 9p and 2 copies of chromosome 9q. Her karyotype was designated 47,XX, +idic(9)(q12).arr[hg19] 9p24.3p12 (46,587-42,374,011)x4 dn. G-banding karyotyping of tissues from hypopigmented skin, normal skin and the pilomatricoma revealed 28.5%, 5%, and none of the aberrant chromosome, respectively. Tissue-limited mosaicism may explain mild developmental delay in the patient and the pilomatricoma presented in this case was likely unrelated to the syndrome. Haplotype information obtained from a SNP microarray revealed meiosis I nondisjunction which determines the meiotic origin of tetrasomy 9p. The rearrangement thereafter which leading to duplication of the short arm and loss of the long arm of chromosome 9 was probably a mechanism of the isochromosome formation. Subsequent postzygotic tetrasomy rescue in certain tissues was a possible cause of tissue-limited mosaicism in our patient. To the best of our knowledge, we are the first to speculate a mechanism insight into a supernumerary isochromosome 9p formation using an extensive collection of SNP markers.

3209T

A case of probable constitutional trisomy 3 mosaicism. *M. Kekis[1], S. Hashimoto[1,4], A. McKinney[1], C. Deeg[1], C. Shuss[2], S. Hickey[2,4], C. Astbury[1,3].* 1) Department of Pathology and Laboratory Medicine, Nationwide Children's Hospital, Columbus, OH; 2) Division of Molecular and Human Genetics, Nationwide Children's Hospital, Columbus, OH; 3) Department of Pathology, The Ohio State University College of Medicine, Columbus, OH; 4) Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH.

Constitutional mosaicism for trisomy 3 is extremely rare, with only a few postnatally diagnosed cases reported in the literature. We report a case of probable constitutional trisomy 3 mosaicism in a 16-year-old female, who presented with chronic joint pain, easy bruising, joint hypermobility and dysmorphic features (long, thin facies, over-folded dysplastic ears, micrognathia, and cleft palate). Microarray analysis with the Agilent GGXChip+SNP revealed a gain of an entire chromosome 3. FISH analysis with the centromere probe for chromosome 3 demonstrated the presence of an extra chromosome 3 in 5.4% (3/55) of the metaphase cells and 22.5% (9/40) of the interphase cells analyzed. Additional FISH analysis with the centromere probe was performed on buccal swab cells, which revealed the presence of the supernumerary chromosome 3 in 22% of the interphase cells counted. Clinical features previously reported in patients with trisomy 3 mosaicism include variable survival ranging from neonatal deaths to survival in adulthood, variable cognitive phenotype, intrauterine growth retardation, short stature, dysmorphic facies, cleft lip/palate, cardiac defects, hip dislocations, eye and/or ear abnormalities, macro/microcephaly, abnormal digits, spina bifida, scoliosis, gouty arthritis. Our patient was small at birth (5 lbs 6 oz at 38 weeks), had cleft palate repair, developed chronic joint pain at age 12, and has a history of mild leukopenia, mild thrombocytopenia, and Hashimoto's thyroiditis. She repeated the first grade and required speech therapy but currently attends a mainstream high school classroom. It is predicted that the clinical course of an individual with trisomy 3 depends on the proportion of cells with the supernumerary chromosome 3 and the tissue type where the aneuploidy exists. In our patient, both peripheral blood and buccal cells demonstrated mosaicism for trisomy 3. Although trisomy 3 has been seen as a somatic cytogenetic abnormality in lymphoma and renal cell carcinoma, these neoplasias have not been reported in previously reported patients with germline mosaicism for trisomy 3. Since it is not known whether germline mosaicism for trisomy 3 may increase the risk for neoplasia, a Hematology/Oncology consultation in addition to genetic counseling was provided for our patient. The oncologist's laboratory evaluation for malignancy was unremarkable. A bone marrow aspirate was not performed.

3210T

A de novo microduplication at 7q11.23 from Central Brazil detected by Chromosomal Microarray Analysis. *L.B. Minasi[1,2], I.P. Pinto[1,2], A.V. Melo[2,3], D.M.C. Cunha[1,2], C.L. Ribeiro[2], C.C. da Silva[1,2,3,5], D.M. e Silva[2,4], A.D. da Cruz[1,2,3,5].* 1) Pontifical Catholic University of Goias, Genetics Master's Program; 2) Pontifical Catholic University of Goias, Department of Biology, Replicon Research Group; 3) Federal University of Goias, Biotechnology and Biodiversity PhD Program; 4) Federal University of Goias, Genetics and Molecular Biology Master's and PhD Programs; 5) Human Cytogenetics and Molecular Genetics Laboratory, Secretary of Goias State for Public Health.

The Chromosome 7q11.23 Duplication Syndrome (MIM 609757) is a genomic disorder caused by the duplication of a common 1.5 Mb segment spanning 26 genes. This region is also associated with the Williams-Beuren Syndrome (WBS [MIM 194050]), and duplications of the WBS region should occur at the same frequency as deletions based on an interchromosomal nonallelic homologous recombination mechanism. The duplication at 7q11.23 shows variable clinical manifestations most common being speech delay, mild craniofacial anomalies, and neurocognitive and behavioral impairment, specifically, intellectual disability. Herein, we report a first case of 16 years old boy who presented moderate intellectual disability, mild facial dysmorphisms with broad forehead, high- broad nose, short philtrum, straight eyebrows, and speech delay with a *de novo* 7q11.23 microduplication in Central Brazil. Chromosomal analysis by GTG banding using the software IKAROS® (Metasystems Corporation, Germany) showed 46,XY and Chromosomal Microarray Analysis (CMA) with Affymetrix's CytoScan™ HD SNP array revealed a *de novo* 1.43 Mb microduplication at 7q11.23, encompassing 1,540 markers, and spanning over described 28 genes (*NSUN5, TRIM50, FKBP6, FZD9, BAZ1B, BCL7B, TBL2, MLXIPL, VPS37D, DNAJC30, WBSCR22, STX1A, MIR4284, ABHD11-AS1, ABHD11, CLDN3, CLDN4, WBSCR27, WBSCR28, ELN, LIMK1, EIF4H, MIR590, LAT2, RFC2, CLIP2, GTF2IRD1, GTF2I*). The progenitor's CMA confirmed *de novo* genomic imbalances in their child. Some genes included in that genomic region have been shown to be implicated in Williams Beuren Syndrome. However, deletion or duplication in that region produces different degrees of impairment with variable phenotypes. In addition, strategically CMA has proving to be a powerful tool to carry out genetic diagnostic and it is especially useful to investigate children with developmental and neurobehavioral delays, reducing the time to reach a clinically relevant diagnoses in which genetic rearrangement display an important etiological role such as the genetic evidence found for our proband. Although, CMA will not replace classical karyotyping, it is an efficient method to delineate phenotypic variations and to guide medical diagnosis.

3211T

The 9p trisomy due to maternal t(9;22) in four patients suggesting a palindorm mediated mechanism involving chromosome 9q12. *A. Mohamed[1], A.K. Kamel[1], M.S. Zaki[2], M. Mekkawy[1], N.M. Meguid[3], S.G. Abdallah[1], S.A. Temtamy[2].* 1) Human cytogenetics, National Research Centre, Cairo, Egypt; 2) clinical genetics; 3) research on children with special needs.

We reported on four patients with trisomy 9p resulting from maternal balanced translocation t(9;22)(q12,q11.2).The 1st patient aged 4year, she had severe intellectual disability (ID), delayed milestones, dysmorphic features and congenital heart . Her karyotype was 46,XX,der(22). FISH using WCP9 and 22, subtelomere 22, CATCH, and 9p21 revealed that the patient has trisomy 9p due to the balanced translocation in her mother. Patient two aged 9 months had ID, developmental delay, hypotonia, brachycephaly, hypertelorism. The cytogenetic revealed balanced translocation in the mother, the proband had a derivative 22. FISH showed the presence of 9p duplication. Patients 3 and 4 were brothers , they had ID, MCA. Their mother had 46,XX,der(22),der(22). Both patients had 47,XY,der(22),+marker. FISH indicated that the mother had involvement of both chromosomes 22 in double translocation to 9p and 9q.The balanced translocation in the mother produce trisomy 9p in her sons. In the 4 patients the break points were at 22q11.2 and 9q12. Chromosome 22q11.2 is one of the best studied example of palindromic AT-rich repeat (PATRR) which induced recurrent constitutional translocation. This is the first time to record non random occurrence of translocation involved 22q11.2 and 9q12. We suggest the presence of palindromic sequence at 9q12. The repeated nonrandom balanced translocation between 22q11.2 and 9q12 depends on the same mechanism of Emanuel syndrome and resulted in trisomy 9p.Further investigations of those patients and other similar finding may lead to diagnose a new syndrome or new subtype of trisomy 9p due to the palindromic region at 9p12.

3212T

An Investigation of Pediatricians' Use of Microarray. *N.A. Watkins[1,2], J. Carroll[3,4], D. Chitayat[1,2,5], S. Khattak[6], J. Starvopoulos[7,8], C. Shuman[1,2].* 1) Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Genetic Counselling Program, Department of Molecular Genetics, University of Toronto; 3) Dept. of Family & Community Medicine Mount Sinai Hospital, Toronto; 4) Department of Family and Community Medicine, University of Toronto; 5) Prenatal Diagnosis and Medical Genetics, The Hospital for Sick Children, Toronto; 6) Kids Clinic, Whitby ON; 7) Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto; 8) Cytogenetics, The Hospital for Sick Children, Toronto.

Objective: Chromosomal microarray (CMA) is used to detect microdeletions and microduplications throughout the genome. Both the Canadian and American Colleges of Medical Genetics recommend CMA as a first tier test for patients with developmental delay, multiple congenital anomalies, and autism. It is unclear how pediatricians are using CMAs and how comfortable they are incorporating this testing into their practices. Our objective was to determine pediatricians' use of and comfort with CMA. **Methods:** A self-complete questionnaire was mailed, with a reminder mailing, to all pediatricians listed on the website of The College of Physicians and Surgeons of Ontario in spring 2014. Descriptive results are reported. **Results:** Response rate was 23% (324/1419), 55% female, mean age 50. 68% of pediatricians reported they had ever ordered CMA with 25% ordering > 10 this past year. 52% reported increased use of CMA from the previous year. Commonest indications for ordering were: developmental delay (84%), intellectual disability (72%), dysmorphic features (92%), multiple congenital anomalies (89%). Of respondents currently using CMA, 10% reported above average knowledge and 62% reported average knowledge of CMA. Most pediatricians (82%) reported having an average or above average comfort level with ordering CMAs for their patients. When disclosing CMA results, respondents were least comfortable reporting a variant of unknown significance with 42% reporting below average comfort level. Barriers to ordering CMAs included: lack of sufficient genetic knowledge (23%), lack of clear clinical guidelines (20%), and questioning the benefit that the CMA will have for the patient (11%). 23% reported no barriers to ordering CMAs. Barriers to disclosing CMA results included lack of sufficient genetic knowledge (33%) and lack of clear clinical guidelines (14%). 27% listed no barriers to disclosing results. **Conclusions:** Our results suggest that many Ontario pediatricians have not ordered CMA testing. Those who are ordering CMAs are generally comfortable with ordering the tests and disclosing results, especially benign and pathogenic variants however pediatricians report barriers affect their utilization of CMAs. We recommend addressing these barriers with educational strategies to increase pediatricians' usage and comfort level with CMAs specifically focusing on lack of genetic knowledge, awareness of clinical guidelines and disclosure of variants of unknown significance.

3213T

Clinical and molecular cytogenetic characterisation of Williams syndrome. *B. Kar[1], S. Sivamani[1], G. Kalpana[2], K. Sivakumar[2].* 1) Center for Genetic Studies & Research; 2) Institute of Cardiovascular Diseases, The Madras Medical Mission, Chennai - 600037, Tamilnadu, India.

Williams syndrome (OMIM 194050) is a rare multisystem genetic disorder with an incidence of 1/75000 which usually occurs sporadically caused by the deletion of 26 contiguous genes, including elastin (*ELN*) (OMIM 130160) on chromosome 7q11.23. The Williams syndrome phenotype is characterized by facial dysmorphisms, cardiac abnormalities, behaviour characteristics, intellectual disability, elastin arteriopathy, short stature, connective tissue abnormalities, infantile hypercalcemia, and a unique personality and cognitive profile. We present here three cases of Williams syndrome who were referred to paediatric cardiologist because of a systolic murmur in upper sternal border. Case 1 aged 1 year had neonatal coarctation and arch hypoplasia with diffuse narrowing of ascending aorta. She underwent surgical correction of coarctation followed by balloon angioplasty for recoarctation. Her surgical findings were markedly thickened aortic wall. No aortic biopsies were done. Case 2 aged 9 months had mild supravalvar aortic stenosis (SVAS) and is on regular medical follow up. Case 3 aged 2 years was diagnosed to have large ventricular septal defect (VSD), bilateral peripheral pulmonary artery stenosis, mild SVAS had surgical fenestrated VSD patch closure, branch pulmonary artery pericardial patch repair with DOTY's repair of SVAS. With a triad of dysmorphic facies, cognitive disorder and characteristic congenital heart defect, a diagnosis of Williams syndrome was made clinically. Karyotyping and fluorescent in situ hybridization (FISH) techniques employing the elastin gene probe were performed. Karyotyping was normal in all the three patients whereas FISH result was positive for 7q11.23 microdeletion confirming Williams syndrome. Parents were screened for microdeletion by FISH and were found normal. In this paper we suggest a defined protocol with more attention while evaluating heart murmur in childhood period, especially when the patient has facial dysmorphisms or developmental delay.

3214T

Target-specific synthetic oligonucleotide libraries for use in Fluorescent In Situ Hybridization. *K.C. Semrau[1], Y.E. Murgha[1], E. Robinson[2], F.A. Ray[2], C. Proudhon[3], B. Hao[3], J. Skok[3,4], E. Gulari[1,5], J-M. Rouillard[1,5].* 1) MYcroarray, Ann Arbor, MI; 2) KromaTiD Inc., Fort Collins, CO; 3) Department of Pathology, New York University School of Medicine, New York, NY; 4) NYU Cancer Institute, New York University, New York, NY; 5) Department of Chemical Engineering, University of Michigan, Ann Arbor, MI.

Human cytogenetic applications rely on detecting the presence, position, and location of specific chromosomal regions within the nucleus as well as chromosomal abnormalities, mainly through the use of Fluorescent In Situ Hybridization (FISH). Conventionally, FISH probes have been generated from PCR amplification of genomic regions or BACs followed by nick translation to incorporate fluorophore(s). These probes are accompanied by limitations associated with the specificity of probes. We have developed a method to produce large quantities of customizable FISH probes from synthetic oligonucleotides that show improved specificity and coverage over traditional genome-derived probes. Here we present an algorithm to design complex probe libraries to target specific regions of the human genome (or any sequenced genome) optimized for reproducible hybridizations. The resulting probe sequences are synthesized on a microarray, cleaved, and linearly amplified by in vitro transcription before being converted to single-stranded labeled probes by reverse transcription. Through this IVT-RT procedure, we have shown improvement in the total yield of labeled probes over previously used methods. Here we present highly-specific detection of genomic regions ranging from a few kilobases to several megabases in size using probe densities ranging from 1 to several probes per kilobase. These libraries have been successfully used for FISH in mammalian, insect, and plant cells as well as for 3-D FISH.

3215T

Mapping breakpoints of a familial chromosome insertion (18:7) (q22.1; q36.2q21.11) to DPP6 and CACNA2D1 genes in an azoospermic male. *W. Fan[1], L. Li[2], H. Chen[3], C. Yin[1], C. Yang[3], B. Wang[1], S. Zheng[2], J. Zhang[2].* 1) Molecular genetics, Hebei University, Baoding, Hebei, China; 2) Institute of Medical Genetics, Linyi People's Hospital, Shandong 276003, China; 3) BGI, 11-2 Building, Northern Industry District, Shenzhen 518083, China.

It is widely accepted that the incidence of chromosomal aberration is 10-15.2% in the azoospermic male; however, the exact genetic damages are currently unknown for more than 40% of azoospermia. To elucidate the causative gene defects, we used the next generation sequencing (NGS) to map the breakpoints of a chromosome insertion from an azoospermic male who carries an balanced, maternally inherited karyotype 46, XY, inv ins (18;7) (q22.1; q36.2q21.11). The analysis revealed that the breakage in chromosome 7 disrupts two genes, dipeptidyl aminopeptidase-like protein 6 (DPP6) and contactin-associated protein-like 2 (CACNA2D1), the former participates in regulation of voltage-gated potassium channels, and the latter is one of the components in voltage-gated calcium channels. The deletion and duplication were not identified equal or beyond 100Kb, but 4 homologous DNA elements were verified proximal to the breakpoints. One of the proband's sisters inherited the same aberrant karyotype and experienced recurrent miscarriages and consecutive fetus death, while in contrast, another sister with a normal karyotype experienced normal labor and gave birth to healthy babies. The insertional translocation is confirmed with FISH and the Y-chromosome microdeletions were excluded by genetic testing. This is the first report describing chromosome insertion inv ins (18; 7) and attributes DPP6 and CACNA2D1 to azoospermia.

3216T

Refining 16p11.2 microdeletion region for Intellectual Disability/Developmental Delay (ID/DD). *P.S. Lai[1], Y. Rong[1], K.M. Eu[1], P.S. Low[1], E.C. Tan[2].* 1) Dept Pediatrics, National Univ Singapore, Singapore, Singapore; 2) KK Research Lab, KK Hospital, Singapore.

Copy number variations on chromosome 16p11.2 have been associated with extensive phenotypic variability and diverse phenotypes such as autism spectrum disorders (ASDs), intellectual and developmental disabilities (ID/DD), schizophrenia, etc. Within this region, a proximal 16q11.2 microdeletion region (≈ 593 kb spanning Chr16: 29.5 to 30.1 Mb) has been commonly associated with developmental delay, autism spectrum disorder, epilepsy and obesity. An atypical and less frequently reported microdeletion region (≈ 220 kb spanning Chr16:28.74 to 28.95 Mb) adjacent and distal to this has been reported in patients with developmental delays, behaviour problems and unusual dysmorphic features. We report a seven-year old child from non-consanguineous marriage, who presented with ID and DD. The child was diagnosed with severe learning impairment and psychomotor developmental delay. No dysmorphic features were observed with the exception of a bifid thumb. Array CGH analysis using the 180K oligo chip (Agilent) identified a de novo 1.8 Mb deletion at 16q11.2 (UCSC HG18, Chr16: 28,285094 - 30,098069). Further qPCR dosage analysis delineated the genomic deletion to a narrower region between 28.39 to 29.7 Mb with the breakpoints spanning EIF3CL gene to ZG16 genes. As this deletion spans the proximal 'typical' 16p11.2 microdeletion syndrome and distal 'atypical' 16p11.2 microdeletion syndrome, genotype-phenotype investigations can aid in delineating the critical regions underlying the clinical phenotypes. Our patient's phenotype most closely corresponds to that of proximal 16p11.2 microdeletion syndrome, with developmental delay as the predominant shared symptom, and without facial dysmorphism that is more closely associated with the distal 16p11.2 microdeletion syndrome. The deletion region in our patient overlaps with the microdeletion regions of both syndromes, harboring 19 annotated protein-coding genes, namely *ATXN2L*, *TUFM*, *SH2B1*, *ATP2A1*, *RABEP2*, *CD19*, *NFATC2IP*, *SPNS1*, *LAT*, *BOLA2*, *BOLA2B*, *SLX1B*, *SLX1A*, *SULT1A3*, *SULT1A4*, *SPN*, *QPRT*, *C16orf54* and *ZG16*. Recent gene dosage studies in animal models suggest that genes with deletion dosage sensor properties may contribute towards abnormal brain development and function leading to developmental and cognitive impairments. The clinical phenotype in our patient could most likely arise from the haploinsufficiency of one or more of the above genes, and contributes towards refining the critical 16p11.2 region underlying ID/DD.

3217T

A Five Year Retrospective Analysis of the Utility of Family Segregation Analysis in the Evaluation of the Clinical Significance of Variants of Uncertain Significance Detected by Chromosomal Microarray: The Greenwood Genetic Center Experience. *F. Bartel, A. Chaubey, B. DuPont.* Cytogenetics, Greenwood Gen Ctr, Greenwood, SC.

Chromosomal microarray (CMA) and massively parallel sequencing (MPS) technologies have been transformative to the practice of laboratory medical genetics over the past decade. These technologies allow the highest resolution interrogation of the entire genome in such a cost-effective manner that these techniques have become widely adopted by laboratories. In cytogenetics laboratories CMA has become a first tier test in the evaluation of nonsyndromal pediatric patients presenting with intellectual disability, autism or multiple congenital anomalies. CMA offers dramatic increases in the ability to detect genomic copy imbalances, but evaluation of the clinical significance of these imbalances presents a challenge for clinical diagnostic laboratories. Family segregation analysis (FSA) is among the best methods available to assist a clinical diagnostic laboratory in the evaluation of variants of uncertain significance (VUS) detected by whole-genome assays such as CMA and MPS. In order to evaluate the utility of FSA in attributing clinical significance to CMA detected VUS, we conducted a 5 year retrospective study of FSA performed for CMA detected VUS across all referral reasons at a single clinical cytogenetics laboratory. In total 2224 reported copy imbalances from 1682 consultands have been evaluated by quantitative PCR (qPCR) based FSA. Ten percent of VUS were found to be de novo in origin, with a 2.125:1 ratio of deletions to duplications. Deletions ranged from 50kb to 23Mb in size, with average and median of 2.4Mb, and 1.4Mb respectively. Duplications ranged in size from 95kb to 17.5Mb with average and median of 2.5Mb and 825kb, respectively. Among inherited cases, a 1.5:1 ratio of maternal to paternal transmission was detected. A clear inheritance pattern was only established for 64% of cases, largely due to the inability to obtain parental samples from one or both parents in 22.4% of cases. Analysis of de novo hot spots, and comparison of FSA results by referral reason will also be presented. Our studies underscore the value added by FSA in the evaluation of VUS, and highlight the inability to obtain relevant family members in a significant number of cases as a significant limitation of this approach.

3218T

Interstitial duplications of 19p13.3. *H. Risheg[1], R. Pasion[2], S. Schwartz[2], E. Prijoles[3], E.A. Keitges[1].* 1) Laboratory Corporation of America/ Dynacare, Department of Cytogenetics, Seattle, WA; 2) Laboratory Corporation of America, Center for Molecular Biology and Pathology, Department of Cytogenetics, Research Triangle Park, NC; 3) Greenwood Genetic Center, Columbia, SC.

Interstitial deletions involving 19p13.3 characterized by microarray analysis have recently been reported in the literature. However, there is little molecular cytogenetic information reported on interstitial copy number gains (i.e. duplications and triplications) of 19p13.3. We present 4 patients with overlapping copy number gains (3 duplications and 1 triplication) of 19p13.3 identified by high resolution SNP microarray analysis. Three of the 4 copy number gains with parental studies performed confirmed a *de novo* origin. Features described in these patients include intellectual disability, microcephaly, motor and speech delay, and short stature. Copy number gain sizes ranged from 0.613 Mb-2.45 Mb and included a common region arr[hg19]19p13.3(2,885,504-3,501,271). The patient phenotypes were compared with 7 additional patients with overlapping interstitial 19p13.3 duplications from DECIPHER and ISCA databases. All 7 patients from both ISCA and Decipher databases were reported as *de novo* 19p13.3 duplications. The common region of overlap of all 11 cases further narrowed the region of overlap to arr[hg19]19p13.3 (3,098,056-3,451,152) and includes 5 OMIM genes (*GNA11, GNA15, S1PR4, CELF5, NFIC*). The most common clinical features reported among all 11 were intellectual disability, microcephaly and short stature. Less common features were intrauterine growth retardation and fine motor and speech delays. Our results combined with those reported in the databases provide support for copy number gains within 19p13.3 as a cause of a constellation of findings, including intellectual disability/ developmental delay, microcephaly and poor growth. Identifying additional patients with imbalances within this region is important to further recognize associated clinical features.

3219T

Prenatal Chromosome Rearrangements and Markers: Normal SNP Microarray Analysis Associated with Favorable Pregnancy Outcome. *J.H. Tepperberg[1], R.M. Pasion[1], I. Gadi[1], R.D. Burnside[1], L. Kline[1], B. Williford[1], K. Phillips[2], E. Keitges[2], H. Risheg[1], A. Penton[1], J. Schleede[1], S. Schwartz[1], P. Papenhausen[1].* 1) Laboratory Corporation of America, Research Triangle Park, NC; 2) DynaCare/Laboratory Corporation of America, Seattle WA.

Prenatal microarray is effective in delineating copy number changes not detectable by cytogenetic analysis in and around chromosome rearrangements. The goals of this study were: 1. investigate how many prenatal cases with either G-banded chromosome rearrangement or marker chromosome analyzed by high resolution array were “balanced”, within the resolution of the 2.695,000 SNP/oligo microarray; 2. how many apparently balanced translocations were “balanced” by the array; 3. how many de novo balanced rearrangements had an unbalanced result on array. In order to understand the significance of chromosome rearrangements ascertained at the time of amniocentesis/CVS, we reviewed ~2675 prenatal cases from Oct 2011 to June 2012 by a SNP array. Of those referred for SNP microarray testing, 127 (4.75%) had a translocation, inversion, insertion or marker chromosome. Fifty of 127 G-banded chromosome rearrangments had a simple two-break translocation, 26 had an inversion, 3 had an insertion and 46 had a supernumerary marker. Of the 50 simple translocations, 19 cases (38%) were de novo. Of these, 17 (89.4%) had a normal SNP analysis and apparently normal ultrasound and/or pregnancy outcome. Two (10.5%) cases had copy number findings at the reported breaksite - one significant and one variant of unknown significance (VOUS). Two reported balanced translocations had pathogenic copy number changes identified by array that were presumed to be de novo. Of the 26 inversion cases, 22 (84.6%) had normal microarrays, two were VOUS findings within the inversion breakpoint and another two had secondary abnormal microarray findings unrelated to the inversion breakpoints. Of the 46 cases with supernumerary chromosomes, 16 (34%) were normal by array. In conclusion, the majority of apparently balanced chromosome rearrangements appeared to be balanced following microarray analysis. Including the two pathogenic cases that were presumed to be de novo, 19% (4/21) had a gain or loss at the reported rearrangement breaksite. Of the de novo rearrangements and normal microarray, only 1 of the cases was reported to have fetal anomalies. 34% of prenatal cases with supernumerary chromosome had a normal array result and only three of these had a reported abnormal ultrasound. One of these was UPD. While additional follow-up is pending, this preliminary data suggests that a normal microarray result after abnormal chromosome finding appears to reduce the risk for an abnormal pregnancy outcome.

3220T

Validation of an Ion AmpliSeq™ RNA Lung Fusion Panel, workflow, and analysis solution: an OncoNetwork collaborative research study. *J.G. Cienfuegos[1], K. Bramlett[1], C.P Vaughn[13], L. Lacroix[14], R. Petraroli[15], M. Budagyan[15], F. Hyland[15], R. Gottimukkala[15], O. Sheils[2], B. Tops[3], D. Le Corre[4], H. Kurth[5], H. Blons[4], E. Amato[6], A. Mafficini[6], A.M. Rachiglio[7], A. Reimann[8], C. Noppen[5], C. Ainali[15], P. Laurent-Puig[4], R. Franco[9], H. Feilotter[10], P. Park[10], J. Schageman[1], I. Cree[8], J.L. Costa[11], M. Ligtenberg[3], A. Scarpa[6], J.C. Machado[11], K. Nishio[12] OncoNetwork Consortia.* 1) Life Sciences Solution Group, Thermo Fisher Scientific, Austin, TX; 2) Trinity College, Dublin, Ireland; 3) Radboud University Medical Center, Nijmegen, Netherland; 4) University Paris Descartes, Paris, France; 5) Viollier AG, Basel, Switzerland; 6) ARC-NET University of Verona, Italy; 7) Centro Ricerche Oncologiche Mercogliano, Italy; 8) Warwick Medical School, United Kingdom; 9) Surgical Pathology, Instituto Nazionale Tumori "Fondazione Pascale", Napoli, Italy; 10) Queen's University, ON, Canada; 11) IPATIMUP, University of Porto, Portugal; 12) Kinki University Faculty of Medicine, Osaka, Japan; 13) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 14) Institut Gustave Roussy (IGR), Paris, France; 15) Thermo Fisher Scientific, Carlsbad, CA.

Fusion transcripts resulting from translocation events in the oncogenic driver genes ALK, RET, ROS1, and NTRK play an important role in lung adenocarcinoma. There is a need to detect these fusion transcripts with up to date technologies as they may serve as viable therapeutic targets. We have utilized a targeted sequencing approach and developed an Ion AmpliSeq™ RNA Lung Fusion panel, a workflow, and an Ion Reporter analysis solution to detect these known fusion events. The panel detects transcripts from 37 ALK, 9 RET, 15 ROS1, and 11 NTRK fusion variants along with 5 housekeeping genes to serve as internal controls. The workflow is FFPE compatible requiring an input of only 10 ng of total RNA with the capacity to multiplex up to 16 libraries on a single Ion 318™ chip. The panel was initially validated using 10ng of total RNA from a cocktail of 3 cell lines containing known lung cancer fusions (H2228 - EML4-ALK variant 3a and 3b, HCC78 - SLC34A2-ROS1 and LC-2/ad - CCDC6-RET). The library was sequenced using the Ion PGM™ system and analyzed with the AmpliSeq™ RNA Lung Fusion workflow in the Ion Reporter. Analysis showed that the positive control sample contained all expected fusions and control genes and reported zero false positives fusions. This multiplexed fusion transcript targeted sequencing solution is currently being validated by all members of the OncoNetwork Consortium who will test lung cancer tissue samples that have been well characterized by FISH, real-time PCR, IHC, and/or massarray. Initial results from OncoNetwork Consortia members reveal 100% concordance between the AmpliSeq™ RNA Lung Fusion panel and FISH in 25 lung tissue samples.

3221T

Variation in the Zinc Finger Binding Domain of PRDM9 is Associated with the Absence of Recombination on 21q. *T. Oliver[1], C. Middlebrooks[2], A. Harden[1], B. Johnson[1], C. Wilkerson[1], S. Saffold[1], N. Scott[1], S. Sherman[3].* 1) Department of Biology, Spelman College, Atlanta, GA; 2) Laboratory of Translational Genomics, The National Institutes of Health, Bethesda, Maryland; 3) Department of Human Genetics, Emory University, Atlanta, GA.

Proline Rich Domain Containing 9 (PRDM9) is a major determinant of meiotic recombination. Variation in the zinc finger-binding domain (ZFBD) of PRDM9 is linked to altered placement of recombination in the human genome. As altered recombination (both the absence and altered placement of recombination) is also observed among chromosomes 21 that nondisjoin, we genotyped the PRDM9 ZFBD among mothers of children with trisomy 21 to examine the relationship between variation in the ZFBD of PRDM9 and chromosome 21 nondisjunction. In our approach, PCR was used to amplify the ZFBD of PRDM9 and these PCR products were then subjected to bi-directional Sanger sequencing. In order to identify samples with PRDM9 minor alleles, DNA sequencing reads were aligned and compared to the sequence of the PRDM9 major allele previously identified by Berg et al. 2010. Chi-Square analysis was then used to compare the distribution of major and minor alleles between our cases (N=228, mothers of children with Trisomy 21) and controls (N=94). Among individuals exhibiting no recombination on 21q, the PRDM9 major allele was observed at a significantly lower frequency among our cases (0.63) when compared to that of controls (0.86). This suggests that variation in the PRDM9 ZFBD plays a role in the absence of recombination on chromosome 21 which is a major risk factor for its nondisjunction and trisomy 21.

3222T

Importance of cytogenetic and molecular characterization of patients with pigmentary mosaicism. *C. SALAS-LABADÍA[1], R. CRUZ-ALCÍVAR[1], V. ULLOA-AVILÉS[1], A. REYES-LEÓN[1], M.P. NAVARRETE-MENESES[1], S. GÓMEZ-CARMONA[1], C. DURÁN-McKINSTER[1], E. LIEBERMAN-HERNÁNDEZ[1], V. DEL CASTILLO-RUIZ[1], P. PÉREZ-VERA[1], D.E. CERVANTES-BARRAGÁN[1,2].* 1) INSTITUTO NACIONAL DE PEDIATRÍA, MEXICO CITY; 2) Hospital Central Sur de Alta Especialidad, PEMEX MEXICO, CITY.

Pigmentary mosaicism (PM) describes a heterogeneous group of skin anomalies and in association with multisystem involvement (central nervous, musculoskeletal and ocular), could be related with chromosomal abnormalities in 30-60% of PM patients. Many patients often reveal chromosomal mosaicism (>80%) in different proportions and tissues. Microdeletions or point mutations may be present. Therefore, it is very likely that individuals with PM and systemic abnormalities have genetic mosaicism. The aim of this study was: a) To establish a complete cytogenetic and molecular characterization of patients with PM. b) To establish the relationship between chromosomal mosaicism and the presence of PM. It has been suggested that PM should be evaluated to exclude chromosomal mosaicism, especially when follows a Blaschko lines pattern. We have 52 patients with PM and other systemic anomalies that have been analyzed with GTG-bands; we performed karyotype in lymphocytes, light and dark skin samples reviewing 50 metaphases per tissue, in an attempt to exclude low level mosaicism. Molecular analysis with SNP array and FISH were performed in some patients to finely characterize chromosomal abnormalities, and establish if exists an undetected low level mosaicism. Until now, 50 cases have been completely analyzed by cytogenetics, the results could be classified as follows: 1) Mosaics with 2 or more different cell lines and structural alterations n=10; 2) Balanced translocations X;autosomes n=2; 3) Miscellaneous abnormalities n=4; 4) Normal karyotypes, n= 34. In 5/10 patients from group 1, the molecular analysis delineated the structural alteration; in one case, it was able to identify a cell line present in low-level mosaicism (<5%), not detected by cytogenetics. In 1/4 patients of group 4, the molecular analysis became evident a structural alteration not observed by GTG bands. It is well described that chromosome analysis on stimulated cultures may not detect somatic chromosomal mosaic in individuals with PM and the number of metaphases counted may not be sufficient to detect it. Here, we screened a large number of metaphases from three different tissues applying in some cases SNP array/FISH. This plan of analysis could contribute to a better characterization of anomalies in patients with PM. The molecular tools identified cases with chromosomal mosaicism, allowing the delineation of the abnormality, to perform better associations with the phenotype. CONACYT 2012-01-1882277.

3223T

Chromosome Therapy: Correction of Large Chromosomal Aberrations by Inducing Ring Chromosomes in Induced Pluripotent Stem Cells (iPSCs). *T. Kim[1], M. Bershteyn[2], A. Wynshaw-Boris[1,2].* 1) Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH; 2) Institute for Human Genetics and Department of Pediatrics, University of California, San Francisco, CA, 94143.

Approximately 1 in 500 newborns are born with chromosomal abnormalities that include trisomies, translocations, large deletions and duplications. There is currently no therapeutic approach for correcting such chromosomal aberrations in vivo or in vitro. Recently, we attempted to produce induced pluripotent stem cell (iPSC) models from patients that contained ring chromosomes: one with a ring chromosome 17 (r17) and two patients with different ring chromosome 13s (r13). Surprisingly, while all three lines were reprogrammed to iPSCs efficiently, the ring chromosomes were eliminated and replaced by a duplicated normal copy of chromosome 17 in the r17 line and normal copies of chromosome 13 in the r13 lines (Bershteyn et al., 2014, Nature 507:99). This finding suggested a potential therapeutic strategy to correct large-scale chromosomal aberrations. We hypothesized that a chromosome with a large aberration could be corrected by producing a ring chromosome from the aberrant chromosome in iPSCs, which would then be eliminated and replaced by a normal chromosome. We are testing this hypothesis by inducing ring formation in patients with large deletions of chromosome 17 via a Cre/loxP approach. LoxP sites will be inserted by CRISPR/Cas9 mediated gene editing at sites near each end of chromosome 17. Once this is accomplished, we will infect these cells with Cre recombinase to induce the formation of a ring chromosome. To visualize cells that form a ring chromosome, we designed and cloned the partial tandem dimer Tomato (tdTomato) along with these loxP sites such that the cell will express tdTomato upon formation of a ring chromosome due to Cre-mediated recombination. If successful, we will have created a generalized system of “chromosome therapy” for the correction of large chromosomal aberrations by the induction of ring chromosomes through genome editing followed by duplication of the normal chromosome.

3224S

Targeted RNA sequencing of breast cancer genes using a genomic capture approach: cBROCA. *S. Casadei[1], S. Gulsuner[1], A.M. Thornton[1], J.B. Mandell[1], M.K. Lee[1], M.C. King[1,2], T. Walsh[1].* 1) Medical Genetics, University of Washington, Seattle, WA; 2) Genome Sciences, University of Washington, Seattle, WA.

RNA sequencing provides a powerful method for measuring expression of transcripts of interest. We have adapted BROCA, our targeted hybridization and multiplex sequencing approach for genomic DNA, for RNA sequencing. We call this approach cBROCA. The goals of cBROCA analysis are to detect two classes of mutations: those that alter transcript length, as the result of altered splicing, and those that alter relative expression of the allelic forms of a transcript, as the result of possibly cryptic regulatory mutations. As a test of cBROCA, we evaluated RNA from 16 subjects with known mutations in BRCA1, BRCA2, ATM, and PALB2, and from 17 subjects with normal transcripts of these genes. The mutations were conventional and complex changes at splice sites, exonification of intronic sequence, and a small exonic insertion. RNA was obtained from patients' lymphoblast cell lines treated with puromycin to arrest protein synthesis. RNA samples were digested with DNAse-I for removal of genomic DNA. Libraries prepared from double stranded cDNA were captured with biotinylated RNA oligomers targeted to the transcribed regions of 21 genes. Computational analysis of the sequenced libraries was performed using TopHat2 and Cufflinks. Of the 16 samples with known transcript altering mutations, all were correctly identified, and cBROCA results matched those from previous RT-PCR. Splice site changes were identified by sequence reads mapping to novel exon-exon junctions. For each mutation, the number of reads mapped to the novel junction corresponded to the predicted strength of the splice site change. For all complex splice mutations, multiple transcripts were identified, consistent with previous RT-PCR results. Degradation of nonsense-associated transcripts was partially inhibited by puromycin treatment. From all samples with mutations leading to truncations, the nonsense-associated transcripts were detected, proving the efficacy of the puromycin treatment and sensitivity of the targeted sequencing approach. For all samples, sequence reads spanning exon-exon junctions were clear of genomic contamination. In summary, targeted RNA sequencing enabled us to align sequencing reads across splice junctions, to identify and characterize both constitutive and alternate transcripts, and to measure different proportions of alternate transcripts. We will next apply cBROCA to genomically characterized samples for which a causal mutation has not been identified.

3225M

The 12p13.33/RAD52 and 13q13.1/BRCA2 loci and genetic susceptibility to squamous cell cancers of upper aerodigestive tract. *M. Delahaye-Sourdeix[1], J. Oliver[1], M.N. Timofeeva[2,3], V. Gaborieau[2], M. Johansson[2], A. Chabrier[1], M.B. Wozniak[2], D. Brenner[2], M.P. Vallée[1], D. Anantharaman[2], G. Byrnes[4], P. Brennan[2], J.D. McKay[1], ARCAGE, CEE, IARC MC oral cancer, Latin America, Polish, ACTREC, Rome&Japan Head&Neck cancer studies.* 1) Genetic Cancer Susceptibility group (GCS), International Agency for Research on Cancer (IARC), Lyon, France; 2) Genetic Epidemiology group (GEP), International Agency for Research on Cancer (IARC), Lyon, France; 3) Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh and Medical Research Council (MRC) Human Genetics Unit, Edinburgh, UK; 4) Biostatistics group (BST), International Agency for Research on Cancer (IARC), Lyon, France.

Recent meta-analysis of lung cancer GWAS have identified 13q13.1 and 12p13.33 regions, which encompass the BRCA2 and RAD52 genes respectively, as lung cancer susceptibility loci, and particularly lung squamous cell carcinomas (LUSC). Here, we report detailed investigation of two variants rs11571833 and rs10849605, located within the 13q13.1 and 12p13.33 regions respectively, for association with upper aerodigestive tract cancer squamous cell carcinoma (UADT). Using 5,942 UADT cases and 8,086 controls from 9 different studies, we demonstrate that rs11571833, a rare truncating variant in BRCA2 (K3326X), is also associated with risk of UADT (OR=2.53, 95% CI: 1.89-3.38, p=3×10-10). Similarly rs10849605, a common intronic variant in RAD52, is associated with UADT (OR=1.09, 95% CI: 1.04-1.15, p=6×10-4). There was little evidence for association between rs11571833 carrier status and the loss of wild type allele of the BRCA2 gene or a cis- expression quantitative trait locus (eQTL) effect in the Cancer Genome Atlas (TCGA) data. However, we identified rs10849605 as a RAD52 cis-eQTL in UADT (p=1×10-3) and LUSC (p=9×10-4), with the UADT/LUSC risk allele correlated with increased RAD52 tumor expression levels. The 12p13.33 locus, encompassing rs10849605/RAD52, was identified as a significant somatic focal copy number amplification in UADT (n=374, q-value=0.075) and LUSC (n=464, q-value=0.007) tumors and correlated with higher RAD52 tumor expression levels (p=6×10-48 and p=3×10-29 in UADT and LUSC, respectively). Within this 12p13.33 region, rs10849605 was a cis-eQTL for RAD52 only, making it the most plausible candidate gene at this locus. These results demonstrate that the 12p13.33/RAD52 and the 13q13.1/BRCA2 loci, linked by their roles in homologous recombination based DNA repair, are associated with genetic susceptibility of the UADT.

3226T

FOXA1 binding sites are predictive of breast cancer risk. *M. Ghoussaini[1], J. Allen[2], J. Simard[3], K. Michailidou[2], P. Soucy[3], J. Carroll[4], D.F. Easton[1,2], Breast Cancer Association Consortium (BCAC).* 1) Department of Oncology, University of Cambridge, Cambridge, UK; 2) Department of Public Health and Primary care, University of Cambridge, Cambridge, UK; 3) Genomics center, Centre Hospitalier Universitaire de Québec Research Center and Laval University, Quebec, Canada; 4) Cancer Research UK, Cambridge Institute. University of Cambridge, Robinson Way, Cambridge, UK.

FoxA1 acts as a transcription factor that is necessary for all estrogen receptor (ER) alpha-DNA interactions. When it attaches to its binding site, FOXA1 triggers the opening of condensed chromatin regions, the recruitment of ER and the activation of downstream genes crucial for cell division and proliferation. Recent fine-mapping of breast cancer susceptibility loci identified causal SNPs that affect FOXA1 binding. We hypothethised that FOXA1 binding sites could act as predictors for breast cancer susceptibility. To test this hypothesis, we analysed ~300,000 FOXA1 binding sites identified in three ER-positive breast cancer cell lines: MCF7, T47D and ZR751, and looked for enrichment of risk associated SNPs within them. We utilised association data for ~200,000 SNPs in more than 40,000 breast cancer cases and 40,000 control of European origin, from 41 studies participating in the Breast Cancer Association Consortium (BCAC). Associations (per-allele ORs and P-values) were generated for ~11.6M SNPs by imputation to the 1000 genomes project reference. SNPs associated with breast cancer at P<10-6 were ~4 fold enriched in FOXA1 sites in T47D and ZR751 cells and ~3.3 fold enriched in MCF7 cells. These data indicate that FOXA1 binding sites are a strong predictor of breast cancer susceptibility loci, consistent with the model that susceptibility to the disease is frequently mediated through differential FOXA1 binding.

3227S

Identification of Large Intergenic Non-coding genes as Candidate Targets for Prostate Cancer risk-SNPs Utilizing a Normal Prostate Tissue eQTL Dataset. *Y. Zhang[1], S. McDonnell[1], A.J. French[2], J. Cheville[2], S. Middha[1], S. Riska[1], S. Baheti[1], Z. Fogarty[1], L. Tillmans[2], M. Larson[1], N. Larson[1], A. Nair[1], D. O'Brien[1], J. Davila[1], L. Wang[3], J.M. Cunningham[2], D. Schaid[1], S.N. Thibodeau[2].* 1) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 3) Department of Pathology and MCW Cancer Center, Medical College of Wisconsin, Milwaukee, WI.

Prostate Cancer (PC) is the most frequently diagnosed solid tumor in men in the U.S. For PC, multiple genome-wide association studies (GWAS) have now been performed yielding a substantial number of well-validated SNPs that are associated with an increased risk. In general, only approximately 7% of disease-associated single nucleotide polymorphisms (SNPs) are located in protein-coding regions of the human genome, while the rest are located in gene regulatory regions or intergenic regions. Additionally, many GWAS have mapped disease-associated SNPs to the vicinity of noncoding regions, specifically long intergenic non-coding RNA (lincRNA) regions. Understanding how these genetic variations control the expression of lincRNAs in a tissue-specific manner is not well understood. In this study, we tested the association of SNP genotypes with expression levels of lincRNAs in 471 normal prostate tissue samples from men with PC using expression quantitative trait loci (eQTL) analysis. Genome-wide genotypes and genome-wide mRNA expression levels were obtained usingthe Illumina Human Omni 2.5M SNP array and RNA sequencing, respectively. We focused our analysis on 123 risk-SNPs previously identified by multiple GWAS studies for PC, as well as all SNPs (including imputed ones) that were in linkage disequilibrium (r2 larger than 0.5) with each risk-SNP. This resulted in 72 risk-intervals. Furthermore, we focused on cis-acting associations only where the lincRNA transcripts were located within a 2Mb region (+/-1Mb) of the risk-SNP interval. Of the 4,682 SNPs identified in the risk-intervals and 411 expressed known lincRNAs (median raw read counts larger than 10) within the cis regions, we identified 269 significant eQTL signals (p value with bonferroni correction less than 7.6e-7) in 11 risk-intervals associated with 16 genes. Of the 11 risk-intervals, 5 were associated with a lincRNA transcript only, whereas 6 were associated with both a lincRNA gene and a protein-coding gene. Our results suggest that lincRNA genes are important candidate targets for PC risk-SNPs, and, as such, may play an important role in PC susceptibility. Mapping of the causative risk-SNPs and their corresponding affected regulatory elements is currently in progress.

3228M

***BRCA1* and *BRCA2* mutational screening in 223 hereditary breast cancer patients in Chile: genotype-phenotype correlations.** *C. Alvarez[1], T. Tapia[1], E. Perez[1], D. Wiener[1], C. Ruiz[2], M. Ríos[3], I. Avendaño[4], C. Rodríguez[5], C. Cortés[5], L. Matamala[5], A. Cruz[4], C. Missarelli[3], M. Camus[2], P. Carvallo[1].* 1) Departamento de Biología Celular y Molecular, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile; 2) Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile; 3) Hospital Base de Valdivia, Valdivia, Chile; 4) Hospital Barros Luco Trudeau, Santiago, Chile; 5) Hospital Regional de Antofagasta, Antofagasta, Chile.

BRCA1 and *BRCA2* are the only genes described so far which mutations confer a high risk for breast/ovarian cancer. Frequency of families carrying a mutation in these genes varies across populations and in Chile account for almost 20% of all hereditary cases. Usually, patients are selected for *BRCA1* and *BRCA2* genetic screening by defined criteria related to family history of breast and/or ovarian cancer. More recently, other patients with no family history are being screened for these mutations. These patients may present bilateral breast cancer, age of onset for breast cancer before 40 years old, and breast and ovarian cancer in the same woman. Taking all these criteria in consideration we selected 223 Chilean patients for screening of *BRCA1* and *BRCA2* mutations. Screening was performed through Sanger and Next Generation Sequencing. Considering familial cases 34/179 (19%) carried a mutation, 13 in *BRCA1* and 21 in *BRCA2*. Among patients without family history, 6/44 carried a mutation (13.6%) all in *BRCA1*, being those diagnosed before age of 40 or with bilateral breast cancer the most prone to carry a mutation. Those six patients correspond to the 31.6% of *BRCA1* mutation carriers. In contrast, all *BRCA2* mutation carriers had family history. Average age of onset of index patients was 35 years old for *BRCA1* mutation carriers vs 44.4 years old for *BRCA2* mutation carriers. Considering all patients, average age of onset of mutation carriers was 42 years old vs 44 years old in patients without mutation. It has been described that several types of cancer are recurrent in BRCA1 and BRCA2 families. In our group of families, ovarian cancer was more frequent in BRCA1 families (54%) than BRCA2 (29%). Other types of cancer found were prostate and gastric, either in BRCA1 or BRCA2 carrier families. Interestingly uterus/cervix cancer was present in 26% of BRCA1 mutation families, and absent in BRCA2 carrier families. These observations support different and specific cancer susceptibilities among *BRCA1* and *BRCA2* mutation carriers. Acknowledgments to FONDEF ca12i10152 and FONIS sa12i2299.

3229T

Identification of new familial breast cancer susceptibility genes: are we there yet? *I. Campbell[1], E. Thompson[1], M. Doyle[1], P. James[2], A. Trainer[2], R. Scott[3], G. Mitchell[2], Lifepool.* 1) Research Div, Peter MacCallum Cancer Ctr, East Melbourne, Australia; 2) Familial Cancer Centre, Peter MacCallum Cancer Ctr, East Melbourne, Australia; 3) Hunter Area Pathology Service, Newcastle, Australia.

The genetic cause of the majority of multiple-case breast cancer families remains unresolved. Next generation sequencing has emerged as an efficient strategy for identifying predisposing mutations in individuals with inherited cancer. We are conducting whole exome sequence analysis of germline DNA from multiple affected relatives from breast cancer families, with the aim of identifying rare protein truncating and non-synonymous variants that are likely to include novel cancer predisposing mutations. Data from more than 200 exomes show that on average each individual carries 30-50 protein truncating mutations and 300-400 rare non-synonymous variants. Heterogeneity among our exome data strongly suggest that numerous moderate penetrance genes remain to be discovered, with each gene individually accounting for only a small fraction of families (~0.5%). This scenario marks validation of candidate breast cancer predisposing genes in large case-control studies as the rate-limiting step in resolving the missing heritability of breast cancer. The aim of this study is to screen genes that are recurrently mutated among our exome data in a larger cohort of cases and controls to assess the prevalence of inactivating mutations that may be associated with breast cancer risk. We are using the Agilent HaloPlex Target Enrichment System to screen the coding regions of 168 genes in 1,000 BRCA1/2 mutation-negative familial breast cancer cases and 1,000 cancer-naive controls. To date, our interim analysis has identified 21 genes which carry an excess of truncating mutations in multiple breast cancer families versus controls. Established breast cancer susceptibility gene PALB2 is the most frequently mutated gene (13/998 cases versus 0/1009 controls), but other interesting candidates include NPSR1, GSN, POLD2 and TOX3. These and other genes are being validated in a second cohort of 1,000 cases and controls. Our experience demonstrates that beyond PALB2, the prevalence of mutations in the remaining breast cancer predisposition genes is likely to be very low making definitive validation exceptionally challenging.

3230S

Significant evidence for linkage of cutaneous malignant melanoma to 1q41. *L.A. Cannon-Albright[1], J.M. Farnham[1], K.K. Thai[1], J.J. Zone[2], C.C. Teerlink[1].* 1) Division of Genetic Epidemiology, Internal Medicine, University of Utah School of Medicine., Salt Lake City, UT; 2) Dermatology, University of Utah School of Medicine, Salt Lake City, Utah.

A genome-wide linkage analysis using various Illumina SNP platforms was performed in 46 high risk extended Utah melanoma pedigrees including 237 genotyped cutaneous malignant melanoma cases. For linkage analysis a set of 25,437 genomewide SNPs with no LD was selected from the intersection of the 3 illumina high density SNP platforms (550k, 610k, 720K) used. General dominant and recessive models with low disease allele frequency and partial penetrance were used in an affecteds-only analysis. By-pedigree LODs were considered for these singly informative pedigrees using the TLOD statistic. In one pedigree significant linkage evidence was noted at chromosome band 1q25 (max TLOD = 3.6 at 232.6 cM, 1-LOD drop spanning 232.1 - 233.6 cM). An additional linked pedigree in this region achieved a TLOD = 2.7 at 232.4 cM, 1-LOD drop spanning 232.2-233.7 cM. A 2 recombinant region from these 2 pedigrees runs approximately from 232.1-233.7 cM, or 222.1-223.8 Mbp region (hg19). This 1.7 Mb region contains 11 genes, including: MIA3 (homo sapiens melanoma inhibitory activity family, member 3); HHIPL2 (hedgehog interacting protein like 2); TAF1A (encodes a subunit of the RNA polymerase I complex); AIDA (axin interactor, dorsalization associated); DISP1 (dispatched homolog 1; protein product required for normal Hedgehog signaling); TLR5 (toll-like receptor 5); SUSD4 (sushi domain containing 4; novel complement inhibitor-complement is a protease inhibitor belonging to the serpin superfamily); CAPN8 (calpain 8 - involved in membrane trafficking in the gastric surface mucus cells); BROX, FAM177B, and C1ORF65. Sequence analysis of distantly related carrier cases in these 2 linked pedigrees is underway.

3231M

Expression and insertion of MMTV/HMTV env gene sequences in human breast cancer. *A. Cedro-Tanda, A. Cordova-Solis, E. Pina-Jimenez, D. Arenas-Aranda, F. Salamanca-Gomez, N. García-Hernández.* Genetica Humana, Hospital de Pediatría, CMN S XXI, IMSS, México, D. F., Mexico.

Breast cancer is the leading cause of cancer death for women in Mexico. Recently, it was demonstrated that MMTV/HMTV retrovirus contributes with breast cancer progression. The aim of this study was to identify env gene expression and insertion sites of MMTV/HMTV retrovirus in breast tumours in Mexican women. DNA and RNA were extracted from a total of 73 tumours and normal tissues. In order to assess DNA quality, a 700bp GAPDH fragment was amplified for all samples. MMTV detection was performed by nested PCR for two fragments of env gene (660bp and 250bp), as a positive control MMTV env gene (C3H) in pBR322 was used. Presence of the retrovirus was validated by qPCR with TaqMan probes and quantification of the number of copies using the env gene inserted into pBR322. Expression was evaluated by RNA treatment with DNase and then a RT-qPCR with TaqMan probes to detect env gene. MMTV / HMTV retrovirus insertion sites in the tumour genome were determined by splinkerette PCR. The presence of MMTV was 10% in 73 tumours. The data was confirmed by qPCR with TaqMan probes designed for a specific region of the env gene. Sequence analysis shows that env gene fragments present an identity of 95% with HMTV and 97% with MMTV. The number MMTV / HMTV viral copies is variable for each sample (228-7729 copies). env gene expression was determined in 50% of the positive samples. Preliminary analysis reveals retroviral insertions throughout the genome of infected breast tumours. We are working to establish the participation of MMTV or its homolog, HMTV in tumour progression and breast cancer development.

3232T

Association of P2RX7 gene polymorphisms and cervical squamous cell carcinoma risk. *T. Chang[1], Y. Yang[2,3], Y. Lee[1,4,5], T. Chen[2], W. Lin[1], S. Chang[1].* 1) Medical Research Department, Mackay Memorial Hospital, New Taipei, Taiwan; 2) Department of Gynecology and Obstetrics, Mackay Memorial Hospital, Taipei, Taiwan; 3) Department of Gynecology and Obstetrics, Taipei Medical University, Taipei, Taiwan; 4) Pediatrics, Mackay Memorial Hospital, Taipei, Taiwan; 5) Pediatrics, Taipei Medical University, Taipei, Taiwan.

Cervical cancer is a multifactorial disease and increasing evidence suggests that host immunogenetic background may contribute to its pathogenesis. P2X7 receptor has been implicated in the regulation of immune response and associated with certain cancer development. The aim of this study is to investigate associations between the P2RX7 gene single nucleotide polymorphisms (SNPs) and cervical cancer susceptibility. We genotyped 4 functional SNPs (rs17525809 C/T, rs208294 C/T, rs1718119 A/G, and rs3751143 A/C) in 507 cervical squamous cell carcinoma (CSCC) patients and 430 age/sex matched healthy controls by using the Pre-Developed TaqMan Allelic Discrimination Assay. The presence and genotypes of HPV in CSCC patients were determined by PCR. We found no significant associations between the polymorphisms or haplotypes and CSCC. Stratified by the positivity of HPV-16 infection also did not find marked association. Our findings provide no support for the hypothesis that P2RX7 gene polymorphisms are associated with increased risk for CSCC in the Taiwanese population.

3233S

Functional Variants at The 21q22.3 Locus Involved in Breast Cancer Progression Identified by Screening of Genome-Wide Estrogen Response Elements. *H. Chu[1,2], C. Hsiung[1,2], Y. Huang[1], W. Chou[1], L. Hu[1], H. Hsu[3], P. Wu[1,2], M. Hou[4], J. Yu[3], C. Shen[1,2,5].* 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, 115, Taiwan; 2) Taiwan Biobank, Academia Sinica, Taipei, 115, Taiwan; 3) Department of Surgery, Tri-Service General Hospital, Taipei, 114, Taiwan; 4) Cancer Center and Department of Surgery, Kaohsiung Medical University Chung-Ho Memorial Hospital, Kaohsiung, 804, Taiwan; 5) College of Public Health, China Medical University, Taichung, 404, Taiwan.

Estrogen forms a complex with the estrogen receptor (ER) that binds to estrogen response elements (EREs) in the regulatory region of estrogen-responsive genes, and regulates their transcription. Since sequence variants in the regulatory regions have the potential to affect the transcription factor-regulatory sequence interaction, resulting in altered expression of target genes, this study explored the association between single-nucleotide-polymorphisms (SNPs) within the ERE-associated sequences and breast cancer progression. The ERE-associated sequences throughout the whole genome, demonstrated to bind ERα in vivo, were blasted against online information from SNP datasets, and 54 SNPs located adjacent to estrogen-responsive genes were selected for genotyping in two independent cohorts of breast cancer patients, 779 in the initial screening stage and another 888 in the validation stage. With death from breast cancer or recurrence of breast cancer being defined as the respective event of interest, the SNPs at 21q22.3 were significantly associated with overall survival and disease-free survival of patients. Furthermore, these SNPs (rs2839494 and rs1078272) could affect the binding of this ERE-associated sequence to ERα or Rad21 (an ERα coactivator), respectively, resulting in a difference in ERα-activated expression of the reporter gene. These findings support the idea that functional variants in the ERα-regulating sequence at 21q22.3 is important in determining breast cancer progression.

3234M

Post-GWAS functional characterization of the 12p11.23 renal cancer susceptibility locus. *L.M. Colli[1], P. Bigot[1,2], L. Jessop[1], M. Machiela[1], T. Myers[1], S. Chanock[1].* 1) Division of Cancer Epidemiology and Genetics, NCI - NIH, Gaithersburg, MD; 2) University Hospital of Angers, Department of Urology, Angers, France.

In previous GWAS, rs718314 and rs1049380 at 12p11.23 were associated with renal cell carcinoma (RCC). The aim of our study is to perform a functional analysis of the 12p11.23 region in relation to RCC risk. We performed an imputation analysis within 1 Mb of rs718314 in three different previously published RCC GWAS studies (4197 cases and 8527 controls). The genotyped and high-quality imputed SNPs were tested for association with RCC. After meta-analysis, 44 SNPs demonstrated nominally significant association with RCC risk ($p<5\times10\text{-}5$). The two initial GWAS SNPs, rs718314 and rs1049380, were strongly associated with RCC (Padj = $3.44\times10\text{-}6$ and Padj = $5.27\times10\text{-}6$). All nominally significant SNPs were in a non-coding region which contains the 3'-UTR of ITPR2. Six of the 44 variants were in regions enriched for H3K4me1 and H3K27ac, chromatin marks found in enhancers. Only rs7132434, which is highly correlated with the initial GWAS signal (rs718314, r2=1), showed allele specific regulatory activity in luciferase assays and allele specific differences in protein binding by EMSA. The RCC-associated variants were examined for an effect on nearby gene expression (ITPR2, SSPN, SHARP1), using the TCGA database. Five SNPs associated with RCC were present in the TCGA database and all of them were associated with SHARP1 expression in tumor tissues ($p<0.05$). The most significant association was found with rs12814794 ($p=1\times10\text{-}8$) which is in high linkage disequilibrium with rs718314 (r2=0.956) and rs7132434 (r2=0.956). There was no association between these SNPs and SSPN or ITPR2 expression. Recently, SHARP1 has been shown to be involved in the HIF pathway in breast cancer. Also, SHARP1 regulates adipogenic differentiation and could be a link between RCC and obesity, a known risk factor for RCC. Our results suggest rs7132434 is the functional SNP responsible for the GWAS signal and that this locus could act as an enhancer of SHARP1. Further investigations will be necessary to confirm the link between rs7132434 and SHARP1 and to understand the role of SHARP1 in renal carcinogenesis.

3235T

Targeted Gene Sequencing in Familial Colorectal Cancer Type X. *J. Cunningham[1], A. French[1], M. DeRycke[1], S. Riska[2], S. McDonnell[2], S. Gunawarden[1], Z. Fogarty[2], S. Middha[2], S. Baheti[2], D. Schaid[2], Y. Zhang[2], S. Gallinger[3], M. Cotterchio[4], R. Haile[5], G. Casey[6], M. Jenkins[7], J. Hopper[8], M. Woods[9], L. Le Marchand[10], J. Potter[11], P. Newcomb[11], D. Duggan[12], M. Clendenning[13], D. Buchanan[13], N. Lindor[14], E. Goode[2], S. Thibodeau[1] on behalf of the Colon Cancer Family Registry.* 1) Dept Lab Med & Pathology, Mayo Clinic & Fndn, Rochester, MN; 2) Department of Health Science Research, Mayo Clinic, Rochester, MN; 3) Samuel Lunenfield Research Institute, Toronto, Ontario, Canada; 4) Prevention and Cancer Control, Cancer Care Ontario, Toronto, Ontario, Canada; 5) Department of Medicine and Stanford Cancer Institute, Stanford, CA; 6) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, C; 7) Centre for Epidemiology & Biostatistics, Melbourne School of Population & Global Health, University of Melbourne, Melbourne, Victoria, Australia; 8) Centre for Molecular, Environmental, Genetic & Analytic Epidemiology, School of Population Health, The University of Melbourne, Melbourne, Australia; 9) Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada; 10) Epidemiology Program, University of Hawaii Cancer Center, University of Hawaii, Honolulu, HI; 11) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 12) Translational Genomics Research Institute (TGen), Phoenix, AZ; 13) Oncogenomics Group, Genetic Epidemiology, Department of Pathology, University of Melbourne, Australia; 14) Department of Health Science Research, Mayo Clinic, Scottsdale, AZ.

Much of the heritability of colorectal cancer (CRC) remains unexplained. Type X CRC cases are those that meet Amsterdam criteria for hereditary non-polyposis CRC but have microsatellite stable tumors and lack mutations in mismatch repair (MMR) genes. In our recent germline next generation sequencing custom capture project, 144 Type X CRC cases were sequenced. The capture covered 2.9Mb and included known hereditary colon cancer genes (HCC, n=18) and suspected HCC genes (sHCC, n=18). A minimum read depth of 20 and a genotype quality score of 30 were required for sample analyses. Regions containing pseudogenes and variants with < 90% total successful call rate were eliminated. Excluding synonymous SNVs, 64 variants in HCC genes and 64 variants in sHCC genes were detected in Type X CRC cases. Three were known pathogenic variants (*STK11* F354L, *MLH3* V741F and *APC* E1229Q). Three novel variants were splice variants (one each in *PALB2*, *NUDT1*, *RECQL5*), one was in an initiation codon (*NUDT1*), and 29 were predicted to be damaging in several in silico tools (*AXIN1*, *CTNNB1*, *FLCN, GALNT12, PALB2*, *RECQL5*, *APC*, *BMP4*, *CHEK2*, *MLH1*, *MSH2*, *MUTYH*, *PMS1*, *PTEN*, *SMAD4*). The occurrence of potentially deleterious MMR gene variants suggests possible misclassification of MMR status. Notably, among these 144 Type X CRC cases, 35 (24.3%) carried likely deleterious mutations in known CRC and suspected CRC genes, suggesting the value of a broad cancer gene panel. To uncover the remaining heritable factors in Type X CRC, however, more work remains. This work was supported by grant UM1 CA167551 from the National Cancer Institute.

3236S

***SF3B1* mutations in different cancer types cause recognition of sterically hindered cryptic splice-sites downstream of the branch point.** *C. DeBoever[1,2], E.M. Ghia[2], P.J. Shepard[3], L. Rassenti[4], K. Jepsen[5], C.H.M. Jamieson[2,4,5], D. Carson[2,5], T.J. Kipps[4], K.A. Frazer[2,3,6].* 1) Bioinformatics and Systems Biology, University of California San Deigo, La Jolla, CA; 2) Moores Cancer Center, University of California San Diego, La Jolla, CA; 3) Department of Pediatrics and Rady Children's Hospital, University of California San Diego, La Jolla, CA; 4) Department of Medicine, University of California San Diego, La Jolla, CA; 5) Institute for Genomic Medicine, University of California San Diego, La Jolla, CA; 6) California Institute for Regenerative Medicine, University of California San Diego, La Jolla, CA.

One of the biggest surprises to emerge from the growing catalog of somatic mutations in various cancer types is the recurrent mutation of genes encoding the RNA spliceosome. Recurrent mutations in the highly conserved HEAT 5-9 repeats of splicing factor 3B subunit 1 (*SF3B1*) have been reported in myelodysplastic syndrome (MDS), chronic lymphocytic leukemia (CLL), breast cancer, uveal melanoma (UM), and pancreatic cancer. Prior studies have shown that mutated *SF3B1* CLL samples use canonical 5' splice sites but cryptic 3' splice sites. However it is unknown whether *SF3B1* mutation causes the same 3' splicing defects in different cancers. The mechanism by which *SF3B1* mutations cause cryptic 3' splicing and the functional consequences thereof remain unresolved as well. Here we define the specific sequence requirements needed for cryptic 3' splicing in tumors with mutated *SF3B1*. We examined splice junction usage in transcriptome data from *SF3B1* mutant and unmutated CLL, UM and BRCA cases and found that *SF3B1* mutants use as cryptic acceptors AG dinucleotides ~13-17 bp downstream of the branch point that are likely sterically hindered when *SF3B1* is unmutated. The cryptic acceptors are also located >10 bp upstream of nearby canonical acceptors and thus avoid competing with them for splicing. In our genome-wide analysis only 617 AG dinucleotides met these specific sequence requirements and were used as cryptic acceptors. The same cryptic 3' splicing signature was observed in different cancers but only in samples with mutations in ~10 amino acid hotspots in the SF3B1 HEAT 5-9 repeats. We found that the cryptic acceptors used in the *SF3B1* mutants are typically spliced in at a low frequency (<10% relative to the canonical splice site) and are sometimes present in the SF3B1 unmutated tumors but at an even lower frequency (<0.5% relative to the canonical splice site). Nonetheless, we identified three genes previously implicated in cancer (TTI1, MAP3K7 and FXYD5) and four others (YIF1A, ORAI2, ZNF91, RP11-1280I22.1) with cryptic acceptors that were consistently preferred to the associated canonical acceptor in the CLL SF3B1 mutant samples. Our study suggests that cryptic 3' splicing in SF3B1 mutants results from altered sterics of SF3B1 and other proteins bound at the branch point allowing for the usage of acceptors that are normally hindered and provides a framework for understanding the effects of SF3B1 mutations on the pathophysiology of various cancers.

3237M

Targeted Germline Sequencing of Young Onset, Proficient Mismatch Repair Colorectal Cancer Genes. *M.S. DeRycke[1], S.M. Riska[1], S.K. McDonnell[1], A.J. French[1], Z.C. Fogarty[1], S. Middha[1], S. Baheti[1], S. Gunawardena[1], D.J. Schaid[1], Y. Zhang[1], D. Buchanan[2], M. Clendenning[2], G. Casey[3], M. Cotterchio[4], D.J. Duggan[5], S. Gallinger[6], R.W. Haile[7], J.L. Hopper[8], M.A. Jenkins[9], L. Le Marchand[10], J.D. Potter[11], P.A. Newcomb[11], M.O. Woods[12], N.M. Lindor[13], E.L. Goode[1], S.N. Thibodeau[1] on behalf of the Colon Cancer Family Registry.* 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Oncogenomics Group, Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Australia; 3) Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA; 4) Prevention and Cancer Control, Cancer Care Ontario, Ontario, Canada; 5) Translational Genomics Research Institute (TGen), Phoenix, AZ; 6) Lunenfeld Tanenbaum Research Institute, Toronto, Ontario, Canada; 7) Department of Medicine and Stanford Cancer Institute, Stanford, CA; 8) Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Melbourne, Victoria, Australia; 9) Centre for Epidemiology & Biostatistics, Melbourne School of Population & Global Health, University of Melbourne, Melbourne, Victoria, Australia; 10) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI; 11) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 12) Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada; 13) Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ.

Although the overall incidence of colorectal cancer (CRC) is decreasing, incidence in individuals diagnosed before the age of 50 has increased. While familial syndromes account for approximately 20% of young onset cases, the genetic underpinnings remain unknown for the remaining cases. We studied 370 young-onset (YO, <50 years) proficient mismatch repair (pMMR based on IHC and/or MSI testing) cases using targeted capture and sequencing to search for genes contributing to the susceptibility of YO pMMR CRC. Agilent SureSelect custom capture targeted 2.9Mb of germline DNA for sequencing genes on an Illumina HiSeq2000, including known and suspected hereditary CRC genes (HCC, n=36), several miRNAs (n=118), and genes identified by linkage (n=185) or whole exome sequencing (n=363) studies. Samples from 1,483 individuals (various CRC case groups and controls) were sequenced. Variants in the 370 YO pMMR cases were investigated and compared to 95 unaffected spousal controls. Following filtering (based on quality and minor allele frequency, MAF), 17,826 variants (6,979 coding, 52 splice, 10,795 non-coding) were identified in 702 genes and miRNAs in the cases; 142 deleterious variants (nonsense, splicing, initial codon changes, and stoploss) were found in 116 genes (present in 247 cases). Twenty-two genes harbored >1 deleterious variant, including *DNAH7* with four and *APC* and *TMC2* with three. Nine of the deleterious variants occurred in HCC genes, all affecting a single case, suggesting some misclassification of MMR status. Three nonsense variants were identified in *APC*, while *CHEK2*, *MLH1*, *MSH2*, *MSH6*, and *MUTYH* all harbored one nonsense variant; one splicing variant was found in *MLH1*. Two variants were identified in the suspected HCC gene group, both affecting one individual: a *RECQL5* splicing variant and a *NUDT1* missense variant affecting the initial codon. The majority of variants were present in only a single case (n=11,305, 63%) and most were not present in the 95 unaffected controls (n=13,727, 77%), highlighting the need for large control datasets. Of 4,099 intersecting variants, the MAF in cases was five-fold higher than controls for 15 variants, including three missense (in *FMO5*, *MLH3*, and *PROZ*), one synonymous, and 11 non-coding variants. In depth analysis and comparison of the genes and variants present in the YO pMMR cases and other CRC subtypes will be presented. This work was supported by grant UM1 CA167551 from the National Cancer Institute.

3238T

Characterization of *OLFML3* mutations in non-small cell lung cancer. *C. Drennan, M. Orloff.* Department of Epidemiology, University of Arkansas for Medical Sciences, 4301 W. Markham Street, #820 Little Rock, Arkansas 72205.

In Arkansas and the U.S., lung cancer (LC) accounts for more deaths each year than breast, colon, and prostate cancers combined. To date, the phenotypic heterogeneity associated with LC, particularly non-small cell LC (NSCLC), has been the main obstacle in the development of effective prevention and treatment. Recently, an integrative 'omics' analysis of significant regions of tracts of homozygosity combined with genome-wide expression array data was used to map and shortlist NSCLC-related genes (p-values <0.0001). Of these 9 significant genes, *OLFML3* [MIM 610088], which has been implicated in apoptosis, tumor growth, and cell cycle regulation in cancer cells, was reported to be significantly under-expressed in LC cases who were ever and never smokers (p-values <0.0001 and <0.006, respectively). As a follow-up, we postulate that mutations located in *OLFML3* exons, splice sites, and promoter regions contribute to either a risk or protective effect in NSCLC patients.

To address our hypothesis, NSCLC tumor and matched adjacent normal tissues were retrospectively obtained from the University of Arkansas for Medical Sciences (UAMS) Tissue Procurement Facility. From these tissues, we simultaneously extracted DNA, RNA, and protein. We screened for mutations using the Sanger sequencing method and in patients who were positive for *OLFML3* mutations, the transcript levels of *OLFML3* are being assessed. We explored predicted transcription factor (TF) and microRNA (miRNA) binding sites located on *OLFML3*. Our search identified multiple TFs (MZF1, C/EBPb, SRY, Oct-1, SP1, AML-1a, delta E, HNF-3b, v-Myb, XFD1, Nkx-2, CdxA) that have been previously described as important regulators of carcinogenesis. Additionally, we explored the SNPs located on *OLFML3* to identify predicted TF binding patterns that could be affected and thus potentially change the expression levels. On *OLFML3* exon 3, rs2055542, resides 8 base pairs from where transcription factor SRY was predicted to bind at 100% threshold. SRY transcription factors have been shown to regulate lung as well as many other cancers. One miRNA, miR-155, was predicted to bind to *OLFML3* 3' UTR in humans and has been validated in animal models. Similar to *OLFML3*, miR-155 has also been linked to angiogenesis and tumor proliferation. Our continued investigation and characterization of *OLFML3* mutations combined with *OLFML3*-specific functional analyses may unravel a biologically plausible explanation to NSCLC.

3239S

Hox pattern expression and non coding transcripts in the HOX locus are associated with adult medulloblastoma subtype. *A.M. Fontes[1], K.J. Abraham[2], D.G. Pinheiro[3], D.T. Covas[4], J.B. Veiga[1], D.P.C. Tirapelli[5], F.S. Ramalho[6], A.A. Cardoso[7], S.L. Gerson[8], C.G. Carlotti[5].* 1) Department of Genetics, School of Medicine of Ribeirão Preto, University of Sao Paulo, Brazil; 2) Department of Pediatrics, School of Medicine of Ribeirão Preto, University of Sao Paulo, , Brazil; 3) Department of Technology, FCAV-UNESP; 4) Blood Center of Ribeirão Preto - FMRP/USP; 5) Department of Surgery and Anatomy, School of Medicine of Ribeirão Preto, University of Sao Paulo, Brazil; 6) Department of Pathology and Legal Medicine, School of Medicine of Ribeirão Preto, University of Sao Paulo, Brazil; 7) Hermann B Wells Center for Pediatric Research, Indiana University School of Medicine, Indianapolis, USA; 8) Case Comprehensive Cancer Center, School of Medicine, CWRU, USA.

Introduction: Medulloblastoma (MB) represents a heterogeneous group of neuroepithelial primary tumors arising from the cerebellum. MB can be classified at the molecular level into at least four major subtypes: WNT, SHH, Group C and Group D. Although the current managment of MB patients involves specific treatment approaches to their molecular subtype, a significant proportion of patients do not respond to treatment suggesting heterogeneity whithin these subtypes. We investigate for differences in HOX gene expression and their regulators among MB subtypes to provide additional biomarkers for MBs and molecular mechanisms of HOX regulation involved in tumor progression. Methodology: In this study we analyzed 5 frozen tumor samples from 5 adult MB patients, 3 control cerebellar tissues, 4 MB cell lines and 3 control cerebellar primary cells. Two color microarray-based gene expression profilling was performed according to the manufacturer's instructions. R and Bioconductor were used for analysis. Results: The hierarchical clustering of our samples suggests the existence of three groups among the tumor specimens analyzed (WNT, SHH and Group D). Hierarchical clustering based on the expression level of 39 Hox genes showed that Hox pattern expression might be related to MB subtype whereas in normal cerebellum Hox genes are downregulated or not expressed, with the exception of CN5. We found that: HoxD1, D3, D4, D8 and D9, HoxC4, HoxA9 and A10 are overexpressed in WNT MB; HoxA2, A3, A4, A5, A6, A7, A9, A10, A11, HoxB2, B6, HoxC8, C10 and C11 are overexpressed in SHH MB and Group D MB. Since HOTAIR is located a distance of 5,726 bp from HoxC11, we look for 5'-ends and 3'-ends of long ncRNAs located 1bp - 10 kbp apart from each 39 HOX genes. The analysis shows the presence of 5 Hox-related long ncRNAs: HOTSIR-56 mapped at HoxB9; HOTSIR-24 mapped at HoxD3; HOTSIR-49, HOTSIR-69 and HOTNSIR-27 mapped at HoxC domain. HOTAIR is not expressed in MB. These long ncRNAs are differentially expressed among MB subtypes. In our analysis, HOX regulators (MLL2 and LSD1) and Otx-2 are overexpressed in MB-WNT. Conclusion: This study suggests that differences in expression pattern of HOX genes, MLL and long ncRNAs are present depending on MB subtype. Validation analyses and genetic studies are ongoing to elucidate the molecular mechanisms of HOX regulation according to MB subtype, which can provide useful insights for biomarkers and therapeutic intervention.

3240M

Associations between UGT1A polymorphisms and haplotypes and lung cancer risk. *C.J. Gallagher[1,2], A.Y. Angstadt[1], K.M. Schieffer[1], J. Zhu[1], P. Lazarus[3].* 1) Public Health Sciences, PSU College of Medicine, Hershey, PA; 2) Pharmacology, PSU College of Medicine, Hershey, PA; 3) College of Pharmacy, Washington State University, Spokane, WA.

The UDP-glucuronsyltransferases (UGTs) are Phase II drug metabolizing enzymes that catalyze the detoxification of several cigarette smoke carcinogens. Polymorphisms in these genes may affect an individual's ability to detoxify carcinogens and therefore alter risk of lung cancer. There are hundreds of genetic variants that span the UGT1A gene cluster, but the effect of UGT1A polymorphisms on lung cancer risk has not been well-studied. The present study sought to better understand the effect of genetic variation within the UGT1A loci on lung cancer risk by conducting a comprehensive association study of SNPs (tag and coding SNPs) and haplotypes with lung cancer risk by genotyping 407 lung cancer cases and 582 healthy controls on 96 SNPs. Known lung cancer risk factors, including age, sex, smoking status and pack-years of smoking, were adjusted for in logistic regression analysis and stratification analysis by sex, age, smoking status, pack-years of smoking, and lung cancer histology was also conducted. Many UGT1A SNPs showed association with lung cancer risk overall, and in each of the stratifications. Interestingly, one SNP in intron 1 of UGT1A10 was consistently associated with lung cancer risk in the overall analysis and had an even larger effect in cases with adenocarcinoma and in former smokers. Four additional missense SNPs yielded associations with lung cancer risk in the stratified analysis. Even after correcting for multiple testing, a significant association between two haplotypes and small cell carcinoma risk was observed. This study indicates the importance of UGT1A genetic variation in lung cancer risk. As the genes within the UGT1A loci are important in the detoxification of several tobacco-related carcinogens, identification of the effect of genetic variants in the UGT1A family on lung cancer risk will aid in defining high-risk smokers and help to develop targeted interventions.

3241T

Ptprj-interacting susceptibility genes for colorectal cancer. *M. Gerber, M. Cianciolo, A. Toland.* Department of Molecular Virology, Immunology, & Medical Genetics, The Ohio State University, Columbus, OH.

Colorectal cancer (CRC) causes over 50,000 deaths in the United States each year and is the third leading cause of cancer deaths. The identification of genetic risk factors underlying CRC will have immense value as a tool to identify individuals with increased predisposition to this cancer and to highlight new potential therapeutic targets. In the mouse, the gene *Ptprj* maps to a region of the genome linked to colon cancer susceptibility in mice. The *Ptprj* locus interacts synergistically with two other CRC susceptibility loci mapped in the mouse (*Scc5* and *Scc13*) to further increase CRC risk. The *Scc5* locus also independently interacts in a reciprocal manner with *Scc4* to augment susceptibility. Importantly, the susceptibility genes at *Scc4*, *Scc5*, and *Scc13* have not yet been identified. Our goals are (1) to identify which of the multiple genes present at these *Scc* loci are responsible for modifying CRC risk, and (2) to understand how the combination of the susceptibility genes at these regions increases CRC risk. To achieve these goals, RNA-Seq was used to identify genes with genetic variants or expression levels that differ between the CRC-resistant and CRC-susceptible mouse strains used to map the *Scc* loci. Next, SNPs in the human orthologs of these candidate susceptibility genes were tested in 194 pairs of normal and colon tumor DNA (from human CRC patients) for evidence of allele-specific gains or losses. SNPs that showed statistical trends of gains or losses were genotyped in 296 additional DNA pairs. These studies revealed that 75 of 950 tested SNPs reproducibly show preferential gains or losses in colon tumor DNA samples compared to matched healthy colon DNA samples (p-values<0.05). From these studies, *Epas1* (*Scc4*), *Csnk1a1* (*Scc5*), and *Prdm5* (*Scc13*) emerged as leading candidates at the loci of interest. Preliminary studies to uncover interactions among these candidate genes using *in vitro* models have thus far shown that the transcription factor Epas1 may regulate expression of *Csnk1a1* and *Ptprj*. The epigenetic regulator Prdm5 may also exert regulatory effects at these loci. Furthermore, Epas1 may promote Wnt signaling in colon epithelial cells and therefore function in an oncogenic capacity. Future work will delve deeper into the exploration of this complex network of gene-gene interactions by manipulating levels and isoforms of the candidate genes in cell lines and assessing effects on cell growth, death, and other cancer-relevant phenotypes.

3242S

Hereditary Acute Myelogenous Leukemia (AML) in a Druze family. *Y. Hadid[1,2], Y. Ofran[1], M. Hayun[1], A. Ghanayim[1], K. Skorecki[1].* 1) Technion, Haifa, Israel; 2) Bnai-Zion Medical Center.

Population genetics studies of the Druze communities by the applicant and colleagues show a genomic architecture which should greatly facilitate gene discovery in health and disease states with both Mendelian and non-Mendelian inheritance. In particular, we have previously shown the Druze to represent a sanctuary or refugium of the population genetic structure of the Levant in antiquity. In contrast to previous formulations, it is apparent, that the contemporary Druze population is comprised of numerous globally diverse founding lineages. The past one thousand year history as a transnational population isolate, characterized by high levels of consanguinity and endogamy has therefore resulted in a disproportionate prevalence and accompanying challenge of genetic disease with both Mendelian and non-Mendelian inheritance patterns. We have recently identified three sisters in a Druze family, two with AML and one with Myelodysplastic Syndrome (MDS). A detailed genealogical family tree going back five generations and including 120 persons (with known medical history) revealed an additional three family members who had been diagnosed with AML at ages younger than 40. The incidence of AML in people younger than 50 years old is 1:100,000, therefore a cluster of 6 young AML/MDS patients in a family of 120 (5%) covering several generations is strongly suggestive for a heritable contribution Research Objectives 1. Map AML/MDS in this family to a specific chromosomal locus in linkage with the disease phenotype.2. Discover a disease-causing abnormality underlying AML/MDS in a gene located within the presumably identified linked interval.3. Verify the pathogenic effect of a discovered mutation by developing a suitable experimental model for gene expression and functional testing.4. Provide genetic counseling and pre-leukemic testing for people from the same family and the Druze community for timely identification of people at risk. Preliminary Results Mutations in RUNX1 and CEBPα are known to be associated with familial AML syndromes and full genetic sequencing of these two genes did not identify mutations. Accordingly, we have proceeded to whole exome capture coupled with heterozygosity mapping in order to progress to localization of a disease-causing gene and identify specific mutations underlying the disease in this family.

3243M

A Novel Risk Variant at the 8q24 Cancer Susceptibility Locus in Men of African Ancestry. *Y. Han[1], K.A. Rand[1], D. Notani[2], D.V. Conti[1], N. Rohland[3], S.S. Strom[4], R.A. Kittles[5], B.A. Rybicki[6], J.L. Stanford[7], P.J. Goodman[8], S.I. Berndt[9], A.JM. Hennis[10], E.A. Klein[11], V. Stevens[12], S. Wu[10], J.S. Witte[13], J. Xu[14], W.B. Isaacs[15], S.A. Ingles[1], A.W. Hsing[16], S.J. Chanock[9], M.B. Cook[17], D. Reich[3], B. Nemesure[10], W.J. Blot[18], D.O. Stram[1], M.G. Rosenfeld[2], S. Watya[19], B.E. Henderson[1], C.A. Haiman[1], The African Ancestry Prostate Cancer Consortium.* 1) University of Southern California, Los Angeles, CA; 2) University of California, San Diego, La Jolla, CA; 3) Harvard Medical School, Boston, MA; 4) The University of Texas MD Anderson Cancer Center, Houston, TX; 5) University of Illinois at Chicago, Chicago, IL; 6) Henry Ford Hospital, Detroit, MI; 7) Fred Hutchinson Cancer Research Center, Seattle, WA; 8) SWOG Statistical Center, Seattle, WA; 9) National Institutes of Health, Bethesda, MD; 10) Stony Brook University, Stony Brook, NY; 11) Cleveland Clinic, Cleveland, OH; 12) American Cancer Society, Atlanta, GA; 13) University of California, San Francisco, San Francisco, CA; 14) Wake Forest University School of Medicine, Winston-Salem, NC; 15) Johns Hopkins Hospital and Medical Institutions, Baltimore, MD; 16) Cancer Prevention Institute of California, Fremont, CA; 17) National Cancer Institute, Bethesda, MD; 18) Vanderbilt University School of Medicine, Nashville, TN; 19) Makerere University, Kampala, Uganda.

Background: The 8q24 locus harbors multiple risk variants for distinct cancers, including >10 that are specific for prostate cancer. In African American men with prostate cancer, African ancestry is over-represented in this region, which is a strong indication that some of the underlying disease-relevant variants are more common in men of African than European ancestry. Therefore, we searched for novel associations with common and rare variation across the 8q24 risk locus in men of African ancestry. **Methods:** We examined the 8q24 locus (127.8-128.8 Mb) in 4,853 prostate cancer cases and 4,678 controls from the African Ancestry Prostate Cancer GWAS Consortium. Genotyping was conducted using Illumina Infinium 1M Duo with imputation to a cosmopolitan reference panel from the 1000 Genomes Project (1KGP; March, 2012). To identify independent risk variants, we performed stepwise conditional logistic regression adjusted for age, study, global ancestry (the first 10 principal components) and local ancestry. **Results:** We identified 199 variants at 8q24 associated with prostate cancer risk at $p<5.0\times10^{-8}$. Of these, the most significant SNP was rs114798100 (odds ratio per allele (OR)=2.32, $p=1.6\times10^{-33}$), which is correlated with a previously reported African-specific variant (rs116041037, $r^2=0.63$ in AFR 1KGP). Conditioning on rs114798100 revealed a second nearby signal, captured by marker rs72725879 (OR=1.28, $p=2.2\times10^{-14}$), which is the most strongly associated signal across the 8q24 region in Japanese men. Further conditioning on rs114798100 and rs72725879 revealed a third and novel signal, defined by variant rs111906932 (OR=1.76, $p=7.9\times10^{-11}$), which is uncommon and only found in African ancestry populations (risk allele frequency, 3%). The associations with these markers were replicated in additional studies from Ghana and Uganda. These three risk variants for prostate cancer explained 7% of the heritability and 7% of the genetic variance on log relative risk scale. They are located within or in close proximity to a number of long non-coding RNAs, including *PRNCR1*, which is prostate-specific, as well as *PCAT2*. **Conclusions:** We identified a novel variant for prostate cancer that is only found in men of African ancestry. These findings highlight ancestry-specific variation, allelic and effect heterogeneity, and, further implicate non-coding RNAs as underlying associations of genetic variation at 8q24 with cancer risk.

3244T

Molecular characterization of oncogenic properties of *S100A4* in pancreatic and lung cancers and identification and characterization of candidate downstream genes. *A. Horii, N. Chen, H. Sekine, N. Tsukamoto, T. Tabata, Y. Saiki, S. Fukushige, M. Sunamura.* Dept Molec Pathol, Tohoku Univ Sch Med, Sendai, Miyagi, Japan.

S100A4 is one of the members of the S100 family that characterized as a small calcium-binding protein, and frequent overexpression as well as positive association between overexpression and metastasis were reported. However, detailed mechanisms for such characteristics are not well understood. We analyzed S100A4 in pancreatic and lung cancers and observed frequently overexpression in both tumor types, irrespective of histological subtype. Methylation status in the CpG sites in intron 1 associated with expression of *S100A4*. In some tumors, however, not methylation at the CpG sites but acetylation of histone regulates expression. Then we performed knockdown and forced expression of *S100A4* in pancreatic and lung cancers. Results of both tumor types shared high similarities; specific knockdown of *S100A4* effectively suppressed cell proliferation, mainly by induction of apoptosis, only in cancer cells with S100A4-overexpression, and forced expression of *S100A4* accelerated cell motility only in S100A4 low-expressing cancer cells. Furthermore, S100A4 expression was significantly correlated with perineural invasion (P = 0.029) and invasion pattern (P = 0.001) in primary pancreatic cancer tissues. Microarray analyses using pancreatic cancer were performed both after specific knockdown and forced expression of *S100A4*, and we identified several candidate genes such as *PRDM2* and *VASH1* after knockdown, and *IFI27* and *NOV* after forced expression. It is notable that positive association between *IFI27* expression and perineural invasion (P = 0.0023) was observed. Our results suggest that S100A4 plays an important role in pancreatic and lung carcinogenesis, and that these two tumor types may share the same or similar pathway in cell proliferation and motility.

3245S

Estrogen Receptor Gene Polymorphisms and Lung Adenocarcinoma Risk in Never-smoking Women. *CF. Hsiao[1], KY. Chen[2], GC. Chang[3], YH. Tsai[4], WC. Su[5], YM. Chen[6], MS. Huang[7], C. Hsiung[1], CJ. Chen[8], PC. Yang[1].* 1) Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan; 2) Division of Pulmonary Medicine, Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan; 3) Division of Chest Medicine, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung; Department of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan; 4) Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Chang Gung Memorial Hospital, Taipei, Taiwan; 5) Division of Hematology/Oncology, Department of Internal Medicine, National Cheng Kung University, Tainan, Taiwan; 6) Chest Department, Taipei Veterans General Hospital; School of Medicine, National Yang-Ming University, Taipei, Taiwan; 7) Department of Internal Medicine, Kaohsiung Medical University Hospital; School of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan; 8) Genomics Research Center, Academia Sinica; and Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan.

Seldom is studied on the association between estrogen receptor gene polymorphism and lung cancer risk. This study is to explore the estrogen receptor gene polymorphisms associated with lung adenocarcinoma risk in never-smoking women. This case-control study included 532 never-smoking female patients with lung adenocarcinoma and 532 healthy controls. The ESR1 single nucleotide polymorphism (SNP) data were retrieved from a genome-wide association study. The associations of ESR1 SNPs with the lung adenocarcinoma risk were estimated by multivariate-adjusted logistic regression. For ESR1, a total of 98 SNPs were retrieved and analyzed. Among the retrieved SNPs, 7 tagged SNP associated with lung adenocarcinoma risk was identified. A stepwise forward selection logistic regression approach was performed, with adjustment for hormone replacement therapy (HRT), education level, passive smoking, cooking fume exposure, and other tagged SNPs. Two SNPs, rs7753153 and rs985192, were significantly associated with lung adenocarcinoma risk. (rs7753153: OR: 1.509, 95% CI: 1.168 - 1.950; rs985192: OR: 1.390, 95% CI: 1.001 - 1.712). Therefore, we may conclude that estrogen receptor gene SNPs may be associated with lung adenocarcinoma in never-smoking women.

3246M

***PALB2* mutations among unselected pancreatic cancer patients in the Czech Republic.** *M. Janatova[1], M. Borecka[1], J. Soukupova[1], P. Kleiblova[1], F. Lhota[1], J. Hojny[1], P. Soucek[2], Z. Kleibl[1], P. Pohlreich[1].* 1) Institute of Biochemistry and Experimetal Oncology, First Faculty of Medicine, Charles University in Prague, Prague 2, Czech Republic; 2) National Institute of Public Health, Prague 10, Czech Republic.

Pancreatic cancer is the fourth most common cancer-related cause of death in the Czech Republic and the incidence is still growing. It has one of the worst prognoses of any type of cancer. Early detection is needed for improving the survival. Because of a relatively low incidence of the disease, the screening should be limited for high-risk patients. Hereditary factors account for approximately 10% of patients with pancreatic cancer. The most prevalent gene responsible for familial pancreatic cancer is *BRCA2*. The *PALB2* gene was described as another gene predisposing to pancreatic cancer. The PALB2 protein has an important role in FA-BRCA pathway involved in DNA damage response and thus tumor suppression. Recently, we have described high frequency of the *PALB2* gene mutations in hereditary breast cancer patients in the Czech Republic (5.5%) that included large genomic rearrangements (LGRs). The aim of this study was to evaluate the frequency of *PALB2* mutations among Czech pancreatic cancer patients. We performed a mutation analysis of the *PALB2* gene in a set of 230 unselected patients with pancreatic ductal adenocarcinoma (PDAC). All exons and flanking intron-exon boundaries were analyzed by direct sequencing. Identified mutations were analyzed in a group of 1220 control samples by HRM analysis. We identified two pathogenic truncating mutations in the *PALB2* gene in 230 PDAC patients (2/230; 0.8%). One mutation is novel (c.1838delA, p.Q613Rfs*15) and one mutation is already described (c.509_510delGA; p.R170Ifs*14). This mutation was observed in one control sample (1/1220; 0.08%). Our findings show that the germline *PALB2* mutations are quite rare in unselected PDAC patients; however, frequency of LGRs needs to be determined. Our analysis will contribute to the improvement in estimation of the disease-associated risk in the *PALB2* mutation carriers. Moreover, unraveling the families with germline *PALB2* mutations can help to identify high-risk individuals who may benefit from early detection screening tests. Supported by grants: IGA MZCR NT 14006-3/2013, PRVOUK-P27/LF1/1, and SVV-UK 3362-2014.

3247T

Next-generation panel based characterisation of breast/ovarian cancer genetic predisposition. *R. Janavicius[1, 2], V. Rudaitis[3], L. Griskevicius[1].* 1) Hematology, oncology, transfusion medicine center, Vilnius University Hospital Santariskiu Clinics, Vilnius, VNO, Lithuania; 2) State Research Institute, Innovative Medicine Center, Vilnius, VNO, Lithuania; 3) Department of Gynecology, Centre of Women's Physiology and Pathology, Vilnius University Hospital Santariskiu Clinics, Vilnius, VNO, Lithuania.

BACKGROUND. Genetic predisposition to breast and/or ovarian cancer is largerly confined to mutations in BRCA1/2 genes, although rarer mutations in other known genes are also important. Massively parallel (or next-generation, NGS) resequencing technology is attractive for identifying cancer predisposing mutations in selected known genes (panels) and discover new associations. METHODS. We aimed to better characterise cancer predisposing landscape in clinically selected 127 breast (BC) and 88 epithelial ovarian (EOC) cancer cases from Lithuania (with family history or early age at diagnosis and negative for previously tested BRCA1/2 genes mutations) by performing NGS based targeted analysis (genomic DNA) of genes previously associated with both common (e.g., breast, ovarian) and rare cancers. Custom made TruSight Cancer Nextera Custom hybridization-based target enrichment method was used for the targeted hybridization and preparation of genomic libraries, which were sequenced on MiSeq (2x150-cycles; Illumina). VariantStudio software was used for annotation and filtering of genetic variants. Further analysis filter encompassed 32 genes previously associated with BC/OC predisposition and other mechanistically implicated cancer predisposition genes. RESULTS. Germline loss-of-function protein truncating mutations (PTM) were identified in 13,4% of BC (17/127) and 3,4% of OC (3/88) BRCA1/2 negative samples. Seven PTM were implicated in 6 genes previously not associated with BC/OC predisposition, which are under further investigation. Other PTM were in 6 known BC/OC genes (BRIP1, CHEK2, PALB2, RAD51C, MLH1, TP53). RESULTS. Potentially pathogenic missense mutations (MM) were identified in 10,2% of BC (13/127) and 3,4% of OC (3/88) BRCA1/2 negative samples. Twelve MM were implicated in 5 genes (ATM, CDH1, BRIP1, NBN, MUTYH) previously associated with BC predisposition. One compound HOXB13 p.V84E + NBN p.I171V and one MUTYH p.G396D + NBN p.I171V BC carrier was identified. Two MM were in 2 known OC susceptibility genes (PALB2, TP53) whereas CDH1 gene (1 MM) was not previously associated with predisposition to OC. Further functional characterisation of selected MM variants is currently ongoing. CONCLUSION. NGS panel based resequencing is effective way for better characterising of cancer predisposition landscape.

3248S

Identification of germline mutations in hereditary prostate cancer families satisfying clinical testing criteria for hereditary breast and ovarian cancer. *A.M. Johnson[1], K.A. Zuhlke[1], L.A. Okoth[1], K.A. Cooney[1,2].* 1) Department of Internal Medicine, University Michigan, Ann Arbor, MI; 2) Department of Urology, University Michigan, Ann Arbor, MI.

Co-clustering of breast cancer (BrCa) and prostate cancer (PCa) within families has been well described. Shared risk factors for hereditary BrCa and ovarian cancer (HBOC) and hereditary PCa (HPC) include increasing age, race, and family history. Epidemiological studies have suggested that mutations in *BRCA1*, *BRCA2*, and other BrCa susceptibility genes may account for some of the cancer risk within these families. Less than 1% of the general population is expected to have a *BRCA1/2* mutation; however, this number increases to 10-15% for those women diagnosed with BrCa. Overall, *BRCA1/2* pathogenic mutations are estimated to account for 20-25% of HBOC. The University of Michigan Prostate Cancer Genetics Project (PCGP) selectively enrolls families with 1) a living affected relative pair diagnosed with PCa at any age or 2) an individual diagnosed with PCa before age 56 regardless of family history. For this study, we restricted our sample to PCGP families with at least 1 male BrCa case or a female BrCa or ovarian cancer case in a first-degree relationship to a PCa case. With the exception of the families with male BrCa, we further enriched our sample set to increase our expectation of finding a mutation in *BRCA1/2* by requiring these families to meet at least two of the family history factors defined by the U.S. Preventive Services Task Force that would qualify an asymptomatic woman for clinical *BRCA1/2* testing (N=50). The PCa case with the earliest onset disease (Avg age dx = 54.9 years; range 41-77) , with some preference given for clinically significant PCa, was selected for sequencing of *BRCA1/2* and other candidate genes using a commercially available next-generation Comprehensive Cancer Panel (Qiagen, Valencia, CA). Investigation of the mutation status of 50 men from families enriched for HBOC and HPC resulted in the identification of four separate potentially deleterious (nonsense or frameshift) mutations in *BRCA1/2* amongst five (10%) of our highly selected cases. Two additional probands had truncating mutations in *ATM* and on proband had a truncating mutation in *TET2*. In conclusion, knowledge of positive family history of BrCa and/or ovarian cancer coupled with the application of clinical risk assessment guidelines has increased the likelihood of identifying mutations in *BRCA1/2* and other cancer genes in men with prostate cancer.

3249M

Fine-mapping of 67 prostate cancer GWAS regions identifies better and multiple association signals. *Z. Kote-Jarai[1], A. AminAl Olama[2], T. Dadaev[1], D. Leongamornlert[1], D.J. Hazelett[3], E. Saunders[1], D. Easton[2], G.A. Coetzee[2], D. Conti[3], R. Eeles[1], The PRACTICAL Consortium.* 1) Oncogenetics, Inst Cancer Research, Sutton, Surrey, United Kingdom; 2) Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, United Kingdom; 3) Keck School of Medicine, University of Southern CAlifornia, Los Angeles, CA, USA.

More than 70 common prostate cancer (PrCa) susceptibility loci have been reported in various genome wide association studies. However, the published associated SNPs are most often only tagging the genomic regions that include the possible causative variants and it remains challenging to identify these. We performed a comprehensive fine-mapping analysis to evaluate the association between genetic variation and risk across 67 PrCa regions. The genotyping data from a custom Illumina iSelect array, iCOGS, from PRACTICAL, a large multi-national collaboration and two UK GWAS studies were used for imputation using 1KG as a reference panel in 26,713 PrCa cases and 26,274 controls. We used the results of a meta-analysis from the above three studies and included all SNPs +/- 500 Kb around the previously reported top SNPs that reached $P< 10^{-4}$ significance. Stepwise logistic regression (SLR) was used to identify independent SNPs in each region and in 27 regions only one SNP remained in the model . In 9 of these the association did not reach genome wide significance, four regions were those identified in Asian population and five were identified as borderline significance previously in studies with a larger European sample sets than those used for this analysis. In 18 of 27 regions the top variant was either a novel SNP (16), or the original variant (2). Amongst these we could confirm 3 previously reported Asian ethnicity GWAS signals now in the European population. For the remaining 40 regions with more than one SNP remaining in the model in the first SLR, we analyzed all SNPs adjusted for the best signal and selected SNPs significant at $P<10^{-5}$ level for a second SLR. After this only one SNP remained in the model in 21 regions and 19 regions harbor multiple independent signals. Five of these are known complex regions at 5p15 (TERT), 8q24, 11q13, 17q12 and 19q13 (KLK), but within 14 regions we identified novel independent association signals. We used functional annotation using data from ENCODE filtered for PrCa cell lines and detailed account of possible protein disruption, microRNA target sequence disruption and regulatory response element disruption of all correlated SNPs at $r(2) \geq 0.80$ to support further selection of the most likely risk affecting variants within all studied regions. We also evaluated improvement in stratifying population risk using polygenic risk score after including the new signals identified in this fine-mapping study.

3250T

Localization and Expression Level of p16 Correlate with Patient's Survival and Human Papilloma Virus Status in Oropharyngeal Squamous Cell Carcinoma. *S. Lai[1,2], V. Sandulache[3,4], J. Zevallos[3,4].* 1) Department of Pathology, Michael E. DeBakey VA Medical Center, Houston, TX, USA; 2) Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, USA; 3) Department of Otolaryngology, Baylor College of Medicine, Houston, TX, USA; 4) Department of Otolaryngology, Michael E. DeBakey VA Medical Center, Houston, TX, USA.

p16 is a tumor suppressor gene. Its gene product has been reported to be overexpressed in many head and neck squamous cell carcinoma (SCC) during the last decade. The association between human papillomavirus (HPV) and head and neck SCC has also been recognized. HPV is thought to cause p16 overexpression by E7-induced inactivation of the retinoblastoma protein. p16 immunoreactivity is often considered as an excellent surrogate marker for HPV infection in head and neck SCC. However, the differential expression of p16 overexpression in nuclear and cytoplasm and correlation with HPV status of oropharyngeal SCC of veteran male patients has not been studied. Tumors in oropharyngeal area are difficult to be resected completely, often unresectable due to complexity of location. Study of the pattern of p16 overexpression, HPV infection, and patient's clinical outcomes plays an important role for therapeutic implication. We studied p16 expression by immunohistochemistry in 135 oropharyngeal SCC of veteran male patients. P16 expression pattern is divided into the five groups: low nuclei/low cytoplasm, high nuclei/low cytoplasm, low nuclei/high cytoplasm, high nuclei/high cytoplasm and no expression. Five year disease-free survival is 28.4%, 74.7%, 0%, 93.1% and 13.7% in each group, respectively. Five year overall survival is 23.5%, 74.2%, 0%, 88.7% and 24.2%, respectively. Nuclear and cytoplasmic expression of p16 correlates with HPV positivity detected by polymerase chain reaction. Our data suggests orophagyngeal SCC shows differential p16 overexpression. High nuclei/high cytoplasm p16 expression levels predict a better pathological response to chemoradiation therapy with longer disease free and overall survival. p16 immunoreactivity can be used as a prognostic marker for chemoradiation adjuvant therapy.

3251S

***HOXB13* G84E germline mutation and prostate cancer risk in the UK.** *D.A. Leongamornlert[1], C. Mikropoulos[1], T. Dadaev[1], M. Tymrakiewicz[1], E.J. Saunders[1], S. Jugurnauth-Little[1], K. Govindasami[1], M. Guy[1], R.A. Wilkinson[1], E.J. Sawyer[1], A. Morgan[1], D. Neal[2,3], F. Hamdy[4], J. Donovan[5], A.C. Antoniou[6], Z. Kote-Jarai[1], R.A. Eeles[1,7], The UK Genetic Prostate Cancer Study Collaborators.* 1) Cancer Genetics, The Institute of Cancer Research, Sutton, Surrey, United Kingdom; 2) Surgical Oncology, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK; 3) Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK; 4) Nuffield Department of Surgical Sciences, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK; 5) School of Social and Community Medicine, University of Bristol, Bristol BS8 2PS, UK; 6) Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge CB1 8RN, UK; 7) The Royal Marsden NHS Foundation Trust, London SM2 5PT, UK.

A rare recurrent missense variant in *HOXB13* G84E (rs138213197) was recently reported to be associated with hereditary prostate cancer. Follow-up population based studies established that the frequency of this variant varies between geographic regions and consequently the associated risk for prostate cancer (PrCa) is also highly variable. We investigated the association of *HOXB13* G84E with PrCa risk and prognosis using data from the largest case-control study from the UK. We screened self-identified white men using a custom TaqMan® assay; 8653 PrCa cases from the UK Genetic Prostate Cancer Study (UKGPCS) and 5251 population based controls from the Prostate testing for cancer and Treatment (ProtecT) study. *HOXB13* G84E carrier status was associated with an increased PrCa risk odds ratio (OR) 3.03, (95%CI=2.01-4.58, P= 1.5×10^{-7}) when all cases were compared with controls (carrier frequencies: *1.55% vs. 0.51%*). The association was stronger when cases with a 1st or 2nd degree PrCa family history were compared with controls (carrier frequencies: *2.4% vs. 0.51%*; OR: 4.76, 95%CI=3.03-7.47, P= $9.847\ 10^{-12}$). There was no association of *HOXB13* G84E carrier status with Gleason Score, presenting PSA, TNM stage and NCCN recurrence risk group or survival. We also investigated the combined effects of the G84E allele and a polygenic risk score (PRS) based on the combined effects of 71 common PrCa susceptibility alleles identified through GWAS. We found no evidence of interaction between the HOXB13 G84E allele and the PRS, suggesting that the PRS and G84E allele act multiplicatively on the risk of developing PrCa. Men in our study carrying the G84E variant who were also in the highest quintile of the PRS were estimated to be at a ~5-fold increased risk of developing PrCa, compared with non-carriers in the middle quintile of the PRS. *HOXB13* G84E is rare in the UK population and confers intermediate risks of developing PrCa. However, in combination with other known PrCa susceptibility alleles, carriers of the *HOXB13* G84E allele may be at a substantially higher risk of developing the disease which in the UK male population is linked to familial prostate cancer. Conversely G84E carrier status does not significantly associate with any clinicopathological variables or survival. Based on our study this rare variant explains ~1% of the familial risk of PrCa in the UK population.

3252M

Identification of hereditary alterations predisposing to breast cancer using Next-Gen Sequencing. *F. Lhota[1], V. Stranecky[2], P. Boudova[1], J. Soukupova[1], H. Hartmannova[2], K. Hodanova[2], P. Kleiblova[1], M. Janatova[1], P. Pohlreich[1], S. Kmoch[2], Z. Kleibl[1].* 1) Inst. of Biochemistry and Exp. Oncology, 1st Faculty of Medicine, Charles University in Prague, Prague, Czech Republic; 2) Inst. of Inherited Metabolic Disorders, 1st Faculty of Medicine, Charles University in Prague, Prague, Czech Republic.

The most common cancer not only in the Czech Republic is represented by Breast cancer (BC). Previously, we have tested 1035 high-risk breast/ovarian cancer patients and identified 269 carriers of pathogenic variants in eight BC susceptibility genes (BRCA1, BRCA2, p53, PALB2, CHEK2, ATM, NBN, and PPM1D). The most frequent alterations in identified mutation carriers were pathogenic variants in BRCA1 (N=188). Identification of some pathogenic variant failed in 74% of analyzed patients from high-risk families. With the aim to identify underlying BC susceptibility variants within hereditary breast cancer (HBC) patient subgroup (N=853), we launched next-gen sequencing (NGS) project targeting 594 genes that code for proteins involved in DNA repair or influencing BC pathogenesis using custom sequence capture panel and SOLiD sequencing. The NGS analysis in 314 high-risk patients and 101 non-cancer controls has been performed so far. We identified 84 (66 unique) protein truncating variants in 58 genes; 35 of these mutations were found in DNA repair genes in 33 out of 314 patients (10%). Totally, 387 missense variants (in 186 genes) were predicted as deleterious (in 15% of patients). Moreover, we found also 19 splicing site alterations. All truncating mutations were confirmed by Sanger sequencing. These variants affect several interesting genes including genes coding for Fanconi anemia proteins and proteins involved in DNA-repair pathways (e.g. WRN, BRIP1, RAD18). We also identified two BRCA1 and one BRCA2 mutations affecting the conservative splicing sites that has not been detected by mutation analysis performed previously. Our preliminary data indicates that targeted NGS, using panels of BC-susceptibility genes tailored to the analyzed geographical population is a rational, powerful, and economic strategy surpassing the classical strategies of mutation analysis. However, the implementation of exome-wide or genome-wide NGS approaches will be indispensable for the detection of population-specific rare variants or private mutations in individuals with higher risk of BC. With identification of truncating variants in potentially "actionable" genes, we are reducing the percentage of uncertainty in characterization of BC-susceptibility genes in high-risk patients from 75% down to 50 %. Supported by grants: IGA MZCR NT14054, IGA MZCR NT14006-3, PRVOUK P27, and SVV-UK 3362-2014.

3253T

Germline copy number variant analysis as a mechanism to identify novel high-risk endometrial cancer gene mutations. *F. Lose[1], G.L. Moir-Meyer[1,2], J.F. Pearson[3], M. Bowman[1], R.J. Scott[4,5], M. McEvoy[6], J. Attia[6,7], E. Holliday[5,8], P.D. Pharoah[9,10], A.M. Dunning[10], D.J. Thompson[9], D.F. Easton[9,10], L.C. Walker[2], A.B. Spurdle[1], HCS, SEARCH, ANECS.* 1) Molecular Cancer Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 2) Mackenzie Cancer Research Group, Department of Pathology, University of Otago, Christchurch, New Zealand; 3) Biostatistics Unit, University of Otago, Christchurch, New Zealand; 4) Centre for Information Based Medicine and the School of Biomedical Science and Pharmacy, University of Newcastle, Newcastle, New South Wales, Australia; 5) Hunter Medical Research Institute, Hunter Area Pathology Service, John Hunter Hospital, Newcastle, New South Wales, Australia; 6) Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia; 7) Department of General Medicine, Hunter Medical Research Institute, John Hunter Hospital, University of Newcastle, Newcastle, New South Wales, Australia; 8) Centre for Information Based Medicine and the School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia; 9) Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK; 10) Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK.

Current statistics indicate that 1 in 2 individuals will develop cancer in their lifetime. For 5-10% of patients - up to 25% for early onset/familial cases - cancer is shown to be due to "high risk" germline (inherited) mutation in a cancer predisposition gene. However, known cancer syndrome genes do not explain all familial cancer cases e.g. only 30% of breast cancer families are explained by mutations in the *BRCA1/2* "breast-ovarian" cancer genes, and our own data show that only 30% of endometrial cancer patients reporting multiple cancer-affected relatives have mutations in known endometrial-colorectal cancer mismatch repair (MMR) genes. It is increasingly recognised that, in addition to point mutations, rare copy number variations (CNVs) are a significant source of genetic susceptibility to cancer. In fact, pathogenic deletions are found in >40% of known familial cancer syndrome genes including "Lynch syndrome" colorectal-endometrial cancer MMR genes *MLH1, MSH2, MSH6, PMS2*, and also in *BRCA1, BRCA2, PTEN, PTCH1, STK11*. For some cancer syndromes/genes, deletions comprise such a high proportion of mutations that routine testing includes Multiplex Ligation-dependent Probe Amplification (MLPA) to detect structural aberrations. Although genome-wide association studies (GWAS) are principally aimed to assess the association between common single nucleotide polymorphisms and disease risk, the data generated from GWAS platforms can also be used to investigate the role of CNVs in cancer predisposition. While current findings from studies demonstrate CNVs have potential to drive discovery of novel cancer genes, there have been no CNV studies in endometrial cancer to date. We have utilised our published endometrial cancer GWAS of 1,190 cases to identify several rare patient-specific germline gene-disrupting CNVs that are recurrent (present in 4 or more cases and not in an ethnically-matched reference set of over 1,000 controls), which we have validated using quantitative Real-Time PCR. We propose that these CNVs identify novel genes involved in predisposition of endometrial cancer and potentially other cancers.

3254S

Fine-scale mapping of the 12q24 breast cancer susceptibility locus. *K. Michailidou[1], AM. Dunning[2], DF. Easton[1,2], Breast Cancer Association Consortium.* 1) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 2) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.

Genome-wide association studies (GWAS) and large scale replication studies have identified more than 80 common loci associated with breast cancer. However, for the majority of these loci the underlying causal genes and variants are unknown. We attempted to define the causal variants underlying one such locus, on 12q24, using dense genotyping of variants on a ~200K custom array (iCOGS), conducted by the Breast Cancer Association Consortium (BCAC). The region contains no RefSeq genes, and is flanked by TBX3, a gene known to be somatically mutated in breast tumours, and MD13L.

Across a 530kb region, 589 variants were genotyped on 48,155 cases and 43,612 controls of European ancestry from 41 studies and 6,269 cases and 6,624 controls of Asian ancestry from 9 studies. Genotypes for a further 4,799 variants were imputed using the 1000 genomes project data as a reference.

We identified 58 variants associated with breast cancer risk at $P<5\times10^{-8}$ in Europeans, the most strongly associated variant being rs1391721 (per-minor allele OR and 95% CI 0.92(0.9-0.94), $P=3.7\times10^{-17}$). rs1391721 is closely correlated with the original GWAS hit, rs1292011. Forward stepwise regression analysis in the region revealed two additional independent variants contributing to the breast cancer risk, at $P<5\times10^{-4}$: rs476780 which lies around 400kb centromeric to the original variant (adjusted OR and 95%CI 0.96(0.94-0.98), $P=8.2\times10^{-6}$), and a 4 base insertion rs200803242 (adjusted OR and 95%CI 1.06(1.03-1.08), $P=2.3\times10^{-4}$). rs1391721 was associated primarily with ER-positive disease (case-only P=0.00013), whereas rs476780 and rs200803242 were associated with both ER-positive and ER-negative disease. Analyses in the Asian samples revealed a significant association for rs1391721 (OR and 95%CI 0.89(0.84-0.94), $P=9.7\times10^{-5}$) but not the other two variants.

For the primary signal rs1391721, we defined a set of 8 variants, spanning a 1.1kb interval, that could not be excluded as causal (iCHAV1: based on a likelihood ratio of 100:1). Of these variants, only rs1391721 lies in in a predicted regulatory region, predicted to be binding site for multiple transcription factors including CTCF and FOXA1 in MCF7 cells based on ENCODE data. These results provide a basis for further analyses to define the causal gene(s) underlying this association.

3255M

Germline mutational analysis in mexican patients with Lynch Syndrome. *J.M. MORENO-ORTIZ[1,2], M.L AYALA-MADRIGAL[1], J.R. CORONA-RIVERA[1,3], M. CENTENO-FLORES[3], V. MACIEL-GUTIÉRREZ[3], R.A. FRANCO-TOPETE[3], J. ARMENDARIZ[4], E. HOTCHKISS[5], L. PEREZ-CARBONEL[5], J. RHEES[5], M. GUTIÉRREZ-ANGULO[1, 6], C.R. BOLAND[5].* 1) INSTITUTO DE GENETICA HUMANA, UNIVERSIDAD DE GUADALAJARA, GUADALAJARA, JALISCO, MEXICO; 2) TRAYECTORIA DE GENÓMICA ALIMENTARIA. UNIVERSIDAD DE LA CIÉNEGA DEL ESTADO DE MICHOACÁN. SAHUAYO MICHOACÁN MEXICO; 3) HOSPITAL CIVIL DE GUADALAJARA "DR. JUAN I. MENCHACA" GUADALAJARA, JALISCO MEXICO; 4) DEPARTAMENTO DE BIOLOGÍA MOLECULAR Y GENOMICA, GUADALAJARA JALISCO MEXICO; 5) GI CANCER RESEARCH LABORATORY, BAYLOR UNIVERSITY MEDICAL CENTER AT DALLAS, TEXAS USA; 6) DEPARTAMENTO DE CLINICAS, CUAltos, UNIVERSIDAD DE GUADALAJARA. TEPATITLAN JALISCO MEXICO.

BACKGROUND. Lynch Syndrome (LS) is characterized by germline mutations in the DNA mismatch repair (MMR) genes: MLH1, MSH2, MSH6 or PMS2. This syndrome is inherited in an autosomal dominant pattern and is characterized by early onset colorectal cancer (CRC) and extra-colonic tumors. The aim of this study was to identify mutations in MMR genes in four Mexican patients with LS. MATERIAL AND METHODS. The study comprised tumor and blood samples obtained from four patients diagnosed clinically with Lynch syndrome at the “Dr. Juan I Menchaca” Civil Hospital of Guadalajara, Jalisco, Mexico. All the patients signed an informed consent. The protein expression was evaluated by immunohistochemical analyses in tumoral tissues. Genomic DNA was extracted from peripheral blood samples for mutation analysis using dHPLC, sequencing and MLPA analysis. RESULTS. Immunohistochemical analyses; LS-23 and LS-41 samples showed the absence of nuclear staining for MLH1, and samples LS-3 and LS-52 showed loss of nuclear staining for MSH2. Germline mutation analysis; One was a substitution at splice donor site, located in first base of intron 18 of MLH1, c.2103+1G>C. Two different sequence variations in exon 16 MLH1 gene were found in the same patient. One allele had a deletion of a codon, c.1852_1854delAAG (p.K618del), and the other was a deletion/insertion of two bases, c.1852_1853delinsGC (K618A) which resulted in a Lysine to Alanine mutation at codon 618. The final sequence variation was a duplication c.638dupT resulting a frameshift in MSH2, exon 3, p.L213fs. No mutations were identified in patient LS-3. CONCLUSION. A compound heterozygous leading to mutation in MLH1 gene and a novel mutation for MSH2 gene are found in Mexican LS patients. REFERENCES 1.Boland CR, Lynch HT. The history of Lynch syndrome. Fam. Cancer. 2013, 12 (2):145-157. 2.Martín-López JV, Fisher R. The mechanism of mismatch repair and the functional analysis of mismatch repair defects in Lynch syndrome. Fam. Cancer. 2013, 12(2):159-168.

3256T

Deep intronic sequencing of mutation-negative Lynch Syndrom patients. *A.M. Nissen[1,2], M. Morak[1], A. Benet-Pages[2], A. Laner[2], M. Locher[2], E. Holinski-Feder[1,2].* 1) Medizinische Klinik und Poliklinik IV, Klinikum der Universität München, Ludwig-Maximilians-Universität, Munich, Germany; 2) Medical Genetics Center Munich, Munich, Germany.

Lynch Syndrome (LS) is caused by germline mutations in genes involved in the DNA mismatch repair (MMR) pathway, and frequently indicated by a high microsatellite-instability and a protein loss in tumor tissue. The mutation detection rate of pathogenic mutations in our patients with MMR-defects is 65%, in 17% uncertain variants (VUS) were detected, and in 18% no variation was found in routine diagnostics. Hence, for a considerable number of patients with MMR-defects in the tumor and familial tumor clustering in many cases the causative genetic predisposition cannot be identified. We set up a cohort of 71 index patients with MMR-defect with no pathogenic variants/VUS or deletions/duplications in the corresponding gene (MSH2/EPCAM, MSH6, MLH1 or PMS2) detectable by DHPLC mutation screening or Sanger sequencing and MLPA analysis. In order to uncover further pathomechanisms in this cohort we performed deep intronic sequencing of eight genes associated with LS or involved in the MMR pathway (MLH1, MLH3, PMS1, PMS2, MSH2, MSH3, MSH6 and EPCAM). All exons, introns and chromosomal regions far upstream/downstream of the genes were included in the target region, allowing for detection of variants in regulatory regions and chromosomal rearrangements. Library preparation was done with the NEBNext Ultra DNA kit and capture enrichment with the Illumina TruSeq kit. We used paired-end sequencing on an Illumina MiSeq system. Data were analyzed using a bioinformatics pipeline consisting of BWA, Stampy, GATK, SAMtools, Pindel and snpEff. Deep intronic sequencing of the patients revealed no clear pathogenic variants within coding exons and flanking intronic regions of the pre-analyzed genes, except for a single nucleotide frame-shift deletion in MSH2 in one patient which was previously missed by DHPLC. No significant coding variants were found in the additional genes PMS1, MLH3, MSH3 and EPCAM. SNP analysis of the intronic and regulatory regions revealed 19 rare VUS, located in either TF binding sites, miRNAs or may introduce potential intronic splice sites, that are now under investigation for pathogenicity. Rearrangement analysis was first validated on control patients with known rearrangements. We confirmed these results and could reliably detect the exact breakpoints of an inversion in MLH1, a duplication in MSH2 and four deletions in MLH1, PMS2 and MSH6. Rearrangement analysis of the patient cohort with the validated pipeline is currently being performed.

3257S

Ephrin Receptor Genotypes Modify Chemotherapy-Induced Peripheral Neuropathy Symptoms: A Candidate Gene Study in Breast Cancer Patients. *K.N.H. Nudelman[1,4], N.R. Zanville[2,4], B.C. McDonald[3], D.J. Smith[3], J.D. West[3], B. Schneider[1,5], V.L. Champion[2,4,5], A.J. Saykin[1,3,5].* 1) Department of Medical and Molecular Genetics, Indiana University-Purdue University Indianapolis, Indianapolis, IN; 2) School of Nursing, Indiana University-Purdue University Indianapolis, Indianapolis, IN; 3) Department of Radiology and Imaging Science, Indiana University-Purdue University Indianapolis, Indianapolis, IN; 4) Training In Research for Behavioral Oncology and Cancer Control Program, Indiana University-Purdue University Indianapolis, Indianapolis, IN; 5) Simon Cancer Center, Indiana University-Purdue University Indianapolis, Indianapolis, IN.

Chemotherapy-induced peripheral neuropathy (CIPN) affects up to 60% of breast cancer patients undergoing chemotherapy. However, work remains to elucidate the etiology of CIPN, including the role that genetic modifiers may play in CIPN symptom (CIPN-sx) onset and severity. Leandro-Garcia et al., 2013, found that SNPs in Ephrin A receptor genes *EPHA4*, *EPHA5*, and *EPHA6* were associated with CIPN in cancer patients treated with taxanes. In addition to replicating this finding, we investigated the association of significant SNPs with different types of CIPN-sx to further clarify the underlying biological mechanism and functional implications. This study utilized data from female breast cancer patients treated with (Ctx+, n=26) and without (Ctx-, n=26) taxanes, assessed at baseline (post-surgery, pre-treatment) and one month post chemotherapy completion. CIPN-sx type and severity was collected with a validated 11-item self-report subscale (FACT/GOG-Ntx). In additional to total score, items are divided into sensory (numbness and tingling), motor, hearing, and functional domains. Candidate SNPs were tested for association with CIPN-sx using PLINK and SPSS. Total CIPN-sx increased in Ctx+ compared to Ctx- (p=0.001); multivariate analysis showed that this signal was driven by significant increases in sensory (p=0.001) and motor (p=0.024) domains. Analysis of 18 Ctx+ white non-Hispanic individuals indicated that SNPs in *EPHA5* (rs1159057, p= 0.020, and rs7349683, p=0.035; in LD, r^2=0.79) were dominantly associated with pre/post change in total CIPN-sx, while *EPHA4* (rs17348202, p= 0.059) showed a trend level association, and *EPHA6* rs301927 (p=0.897) was not associated. Multivariate analysis indicated that six Ctx+ with the rs1159057 common genotype GG had increased scores for sensory (p= 0.022) and motor (p=0.040) CIPN-sx, as well as a trend towards functional CIPN-sx (p=0.079), compared to Ctx+ with GA and AA genotypes. This study replicated the previously demonstrated association of SNPs in *EPHA5* with CIPN, and additionally, rs1159057 was shown to be protective against sensory and motor symptoms. Given the function of Ephrin receptor genes in axonal guidance and repair following injury, and the potential impact of sensory and motor neuropathic symptoms on ability to work post-treatment, these genes may be important targets for future therapeutic development.

3258M

Identification of familial Wilms tumor predisposition genes using whole genome sequencing. *T.B. Palculict[1], E.C. Ruteshouser[1], Y. Fan[2], W. Wang[2], V. Huff[1].* 1) Genetics, MD Anderson Cancer Center, Houston, TX; 2) Bioinformatics and Computational Biology, MD Anderson Cancer Center, Houston, TX.

Wilms tumor (WT), a childhood tumor arising from undifferentiated renal mesenchyme, is diagnosed in North America at a frequency of 1 in 10,000 live births and accounts for 5% of all childhood cancers. The etiology of WT is heterogeneous with multiple genes known to contribute to the development of sporadic WT cases; however, these genes are rarely associated with familial Wilms tumor (FWT). Approximately 2% of WT patients have a family history of WT, which displays an autosomal dominant inheritance pattern and incomplete penetrance, estimated to be 30%. To identify genes that, when mutated predispose to FWT, we used whole genome sequencing of peripheral blood DNA from five individuals in two WT families. Intra-family shared variants were identified and prioritized based on minor allele frequency, expression in fetal kidney and the predicted functional significance of the alteration to protein structure. Our whole genome sequence analysis of three individuals in Family 1 resulted in the identification of a novel germline *DICER1* missense mutation that co-segregates with the affected status. Sanger sequencing data revealed loss of heterozygosity (LOH) in a tumor from one individual in this family. Mutational analysis of this exon in blood samples from members of 47 WT families revealed novel germline *DICER1* mutations that cosegregate in 2/47 families (G803R, R800fs). *DICER1* mutations have previously been reported in three families where WT, as well as other phenotypes of DICER1 Syndrome, are observed. Whole genome sequencing of two members from Family 2 revealed the presence of a 570kb duplication on chr2p, encompassing *NBAS*, *DDX1*, *MYCN* and *MYCNOS*. This duplication was found in 2/47 WT families, co-segregated with the affected status and was absent in all spousal controls. A similar constitutional microduplication was recently reported in a WT family and our data provides strong support for the hypothesis that this germline duplication predisposes to the development of WT. Our results underscore the utility of whole genome sequencing to identify familial predisposition genes, as well as the involvement of *DICER1* and a 570kb chr2p microduplication in the development of FWT.

3259T

The FANCM c.5791C>T nonsense mutation (rs144567652) induces exon skipping and is a risk factor for familial breast cancer. *P. Peterlongo[1,2], I. Catucci[1,2], M. Colombo[2], L. Caleca[2], E. Mucaki[3], F. Damiola[4], L. Bernard[5,6], V. Pensotti[1,6], S. Volorio[1,6], V. Dall'Olio[1,6], A. Kvist[7], H. Ehrencrona[8], S. Mazoyer[4], S. Pizzamiglio[9], P. Verderio[9], P.K. Rogan[3], P. Radice[1,2].* 1) IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy; 2) Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; 3) Department of Biochemistry, Schulich School of Medicine and Dentistry, University of Western Ontario, London Ontario, Canada; 4) Cancer Research Centre of Lyon, CNRS UMR5286, INSERM U1052, Université Claude Bernard Lyon 1, Centre Léon Bérard, Lyon, France; 5) Department of Experimental Oncology, Istituto Europeo di Oncologia, Milan, Italy; 6) Cogentech, Cancer Genetic Test Laboratory, Milan, Italy; 7) Department of Clinical Sciences, Division of Oncology, Lund University, Lund, Sweden; 8) Department of Clinical Genetics, University and Regional Laboratories, Lund University and Skane University Hospital, Lund, Sweden; 9) Unit of Medical Statistics, Biometry and Bioinformatics, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.

Several genetic factors conferring breast cancer risk are known. However, approximately two thirds of the overall familial risk remains unexplained. It is speculated that a fraction of this missing hereditability is due to rare variants with high to moderate risk. We tested the c.5791C>T nonsense mutation in exon 22 of FANCM, one of the genes involved in the Fanconi Anemia pathway, in 8,635 familial cases and 6,625 controls from different populations and found association with breast cancer risk (OR = 3.93, 95%CI = 1.28-12.11, P = 0.017). Following information theory-based prediction, we showed by transcript analysis that the mutation causes the deletion of exon 22, producing the incorporation of 11 frameshifted residues and the loss of 132 amino acids from the FANCM C-terminus (p.Gly1906Alafs12*). We also showed that this exon skipping is most likely due to the creation of a binding site for the pre-mRNA processing protein hnRNP A1. This is the first report providing robust evidence of a heterozygous loss-of-function mutation of the FANCM gene acting as a risk factor for familial breast cancer.

3260S

The *CDH1* gene as a susceptibility locus for lobular breast carcinoma. *C. Petridis[1,2], I. Shinomiya[1], S. Nowinski[1], I. Tomlinson[3], R. Roylance[4], M.A. Simpson[2], E.J. Sawyer[1].* 1) Research Oncology, Division of Cancer Studies, King's College London, Guy's Hospital, Great Maze Pond, London, SE1 9RT; 2) Medical and Molecular Genetics, King's College London, Guy's Hospital, Great Maze Pond, London, SE1 9RT; 3) The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN; 4) Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary, University of London, John Vane Science Centre, Charterhouse Square, London, EC1M 6BQ.

Introduction: Both Invasive lobular breast cancer (ILC) and lobular carcinoma *in situ* (LCIS) are characterised by loss of E-cadherin expression, an adhesion molecule encoded by the *CDH1* gene. While germline *CDH1* mutations have been described in cases of familial ILC associated with diffuse gastric carcinoma (DGC), they are rare in sporadic lobular breast cancer and have not been described in women with LCIS. Common variation at the *CDH1* locus has been associated with colorectal cancer but this association has not been observed in breast cancer to date. **Rare variants:** We sought to address the role of rare loss of function (LoF) *CDH1* variants in ILC/LCIS. We screened *CDH1* in 50 cases with bilateral lobular carcinoma (ILC and/or LCIS) using exome sequencing, Sanger sequencing and MLPA. We found four LoF germline mutations (c.48+1G>A, c.1465insC, c.1942G>T, c.2398delC). None of the four carriers had family history of gastric cancer. We were unable to detect any larger rearrangement in the gene and found no abnormality in any of the remaining cases using MLPA. **Common variants:** Further to the identification of the rare LoF mutations, we evaluated the role of common variants at this locus on ILC/LCIS using genotypic data from the iCOGS platform on 2500 cases and 5000 controls. No association was found in the *CDH1* locus using data from 55 genotyped SNPs on the iCOGs chip. These SNPs were selected to capture the majority of the known common genetic variation at *CDH1*. We have also generated genotypic data for 2500 cases and 1500 controls for rs35187787, which is relatively rare (MAF<1%), not genotyped on the iCOGs chip and predicted to be potentially pathogenic, but found no association (Fisher's exact, p= 0.49). **Conclusion:** We have described *CDH1* germline mutations in women with LCIS for the first time. We have shown that germline *CDH1* mutations are associated with early onset bilateral LCIS with or without ILC in women without a family history of gastric cancer. We have found no common variants in the *CDH1* locus associated with the disease based upon 55 genotyped SNPs. Imputing the 1000 genomes data in the *CDH1* locus, will allow capturing more common variation in *CDH1*. Our data suggest that *CDH1* screening should be considered in early onset (< 50 years) cases of bilateral ILC/LCIS and no family history of DGC in order to identify individuals who may benefit from breast MRI screening and endoscopic surveillance for DGC.

3261M

Melanoma Profiler Web Tool for Integrative Genomic Analysis of Melanoma. *K. Qaadri[1], C. Olsen[1], S. Stones-Havas[2], B. Ammundsen[2], A. Trevarton[3], C. Print[3].* 1) Biomatters, Inc., Newark, NJ; 2) Biomatters Ltd. Auckland, New Zealand; 3) University of Auckland, Auckland, New Zealand.

Despite ongoing research, metastatic melanoma five-year survival rates remain low and treatment options limited. Researchers can access a rapidly growing amount of molecular and clinical information about melanoma that may be critical to understanding this disease and making clinically relevant treatment decisions. However, this information is becoming difficult to collate and clinically interpret due to its dispersed nature. Presented here is Melanoma Profiler, a new cloud-based web application for clinically relevant metastatic melanoma genomic research. It performs a tumor/normal comparative variant analysis, comparing proband modified genes and pathways to those in a curated set of characterized melanoma tumor samples. To provide clinical decision support, modified genes and pathways are cross-referenced with clinical and molecular data. These data are incorporated from publicly available sources, including: associations between gene expression and patient survival, data concerning drug targets, biomarkers, and druggability as well as past clinical trials. The resultant MelanomaDB cross-reference database integrates data from cutting edge research on a continuous basis. The interface is presented in two tiers: a front-panel summary tier and an evidential detail tier. For the busy clinician, immediately presented in the summary view is a dynamic report clearly highlighting the relevant actionable information, potential drug targets, and a summary of affected genes and pathways. For researchers that wish to dig deeper, they may browse annotated pathway data, and view a comprehensive comparative summary of affected genes in waterfall and heatmap reports. Melanoma Profiler is a free research tool hosted by Biomatters, produced in collaboration with the University of Auckland, New Zealand. A recent Frontiers in Oncology methods article describing the research behind the application can be found at http://bit.ly/1bUilqS. The Melanoma Profiler web application may be found at https://apps.biomatters.com/melanoma-profiler/.

3262T

Development of a Next Generation Sequencing Panel for Clinical Diagnostic Analysis of Breast and Ovarian Cancer. *C. Rapp, A. Nissen, A. Benet-Pages, A. Laner, M. Locher, E. Holinski-Feder.* Medizinisch Genetisches Zentrum - MGZ, Muenchen, Germany.

Breast and Ovarian Cancer (BC, OC) are by far the most common cancers affecting women. Approximately 5-7% of BC and 11-15% of OC are estimated to be caused by germline DNA mutations mainly in the BRCA1 and BRCA2 genes, but mutations in additional genes have been also demonstrated to account for significant cancer risk. Knowing the underlying molecular defect can be very valuable for diagnosis, guiding treatment and estimating recurrence risks. Next generation sequencing (NGS) enables development of high-quality clinical tests for disease diagnosis with fast turnaround times. We developed a comprehensive NGS diagnostic panel including 38 genes relevant for breast and ovarian cancer. In order to increase the diagnostic yield we divided these genes into two subpanels according to both phenotypes. Our NGS diagnostic criteria require that sequencing reaches a medium quality of more than 30 per sequence cycle (detection precision of 99.9%) and that coverage of more than 20 sequences per base pair in at least 98% of the analyzed regions is assured. In addition, for BRCA1 and BRCA2 we achieve the highest level of exactitude for NGS that a lab can offer at the current stage (i.e. 100% sequencing of the coding region and flanking intronic sequences with more than 20 sequences per base pair, and deletion/duplication analysis with MLPA). The NGS platform at MGZ has been validated for diagnostic routing in terms of accuracy, analytical sensitivity, analytical specificity and precision. Up to 24 samples of patients with suspected hereditary breast and/or ovarian cancer are processed weekly on an automated workflow which combines capture enrichment sample preparation (TruSightTM Cancer Panel) with massive parallel sequencing (Illumina MiSeq). Data analysis is performed with a bioinformatics pipeline consisting of BWA, SAMtools, and snpEff. Our preliminary results show that 7.4% of the patients had BRCA1/2 pathogenic mutations and 92.6% were negative or inconclusive. Further analysis of these patients revealed in 13.9% of the cases pathogenic variants (CDH1, CHEK2 and RAD51D genes) and 25% presented inconclusive VOUS (ATM, BRIP1, FANCI, MSH2, PALB2, PMS2, PTEN, RECQL4 and RHBDF2 genes). Still 61.1% of the cases remained unsolved. Some of the genes were identified as causative only once, emphasizing the advantage of diagnostic panels for breast and ovarian cancer testing.

3263S

***ABRAXAS* (*FAM175A*) and breast cancer susceptibility: no evidence of association in the Breast Cancer Family Registry.** *A. Renault[1,2], F. Lesueur[1], P. Soucy[2], Y. Hamdi[2], Y. Coulombe[3], S. Gobeil[4], F. Le Calvez-Kelm[5], M. Vallée[5], J.L. Hopper[6], I.L. Andrulis[7], M.C. Southey[8], E.M. John[9,10], J.Y. Masson[3], S.V. Tavtigian[11], J. Simard[2], The Breast Cancer Family Registry.* 1) Curie Institute, Paris, France; 2) Cancer Genomics Laboratory, CHUQ Research Center, Quebec City, Canada; 3) Genome Stability Laboratory, Laval University Cancer Research Center, Quebec City, Canada; 4) CHUQ Research Center, Faculty of Medicine, Laval University, Quebec City, Canada; 5) Genetic Cancer Susceptibility, International Agency for Research on Cancer, Lyon, France; 6) Center for Molecular, Environmental, Genetic and Analytical Epidemiology, School of Population Health, EGA The University of Melbourne, Victoria, Australia; 7) Departement of Molecular Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada; 8) Genetic Epidemiology Laboratory, The University of Melbourne, Victoria, Australia; 9) Cancer Prevention Institute of California, Fremont, USA; 10) Stanford University School of Medicine and Stanford Cancer Institute, Stanford, USA; 11) Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA.

Background: Currently, less than 20% of the familial breast cancer can be explained by highly penetrant mutations in the *BRCA1* and *BRCA2* genes. Other breast cancer susceptibility genes conferring a moderate increase in risk have been recently identified. However, approximately 50% of the heritability is still unexplained, suggesting than other susceptibility genes remain to be discovered. DNA double-strand breaks are one of the most damaging events occurring in a cell, as they can disrupt the integrity and stability of the genome. The majority of breast cancer susceptibility genes already identified, including *BRCA1, BRCA2, CHECK2, ATM, BRIP1* and *PALB2*, are involved in the DNA double-strand breaks repair pathway. *ABRAXAS* is involved in this pathway as a member of the “A complex” which leads BRCA1 to DNA damage sites during homologous recombination repair. Mutations in *ABRAXAS* impair BRCA1 recruitment to DNA damage foci and increase cell sensitivity to ionizing radiation. Moreover, a recurrent germline mutation has been reported in Finnish high-risk breast cancer families. To determine if *ABRAXAS* could be a breast cancer susceptibility gene in other populations, we conducted a population-based case-control mutation screening study in the Breast Cancer Family Registry. **Methodology/Principal finding:** The nine coding exons of *ABRAXAS* were screened by High Resolution Melting curve analysis in 1,332 early-onset breast cancer cases and 1,123 controls. Sixteen distinct rare variants were identified, 15 in cases and 11 in controls: of these one was an in-frame deletion, eight were non-synonymous, four were synonymous, two were intronic and one was located in the 5’UTR of the gene. The two variants p.Thr141Ile (found in cases and controls) and p.Gly39Val (found in one case), were predicted to affect protein function, and we confirmed experimentally that both of them diminish phosphorylation of gamma-H2AX, an important DNA damage signaling event. **Conclusion:** Overall, likely damaging or neutral variants were evenly represented among cases and controls suggesting that rare variants in *ABRAXAS* may explain at most a small proportion of hereditary breast cancer.

3264M

Association between rare and common variants in DNA repair genes and prostate cancer using the iCOGS genotyping array. *E. Saunders[1], T. Dadaev[1], D. Leongamornlert[1], M. Tymrakiewicz[1], S. Jugurnauth-Little[1], A. Amin Al Olama[2], S. Benlloch[2], R. Eeles[1], Z. Kote-Jarai[1], The PRACTICAL Consortium.* 1) Oncogenetics, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, United Kingdom; 2) Centre for Cancer Genetic Epidemiology, University of Cambridge, Strangeways Research Laboratory, Cambridge, CB1 8RN, United Kingdom.

Prostate cancer (PrCa [MIM 176807]) is the most frequently diagnosed male cancer in developed countries and amongst the leading causes of cancer related death. It is known to have a strong heritable component and so far 77 common, low penetrance susceptibility variants have been identified by genome-wide association studies (GWAS). In addition, a small number of rare variants have been found to give rise to greater risk; the majority of which are in DNA repair genes. We previously genotyped 211,155 SNPs on a custom chip (iCOGS) in blood DNA from 21,780 PrCa cases and 21,727 controls of European ancestry from the international PRACTICAL consortium. These SNPs were selected predominantly to evaluate previous suggestive low penetrance associations, to finemap validated associations, or to evaluate candidate genes and pathways. In this study we have analysed the association between common and rarer variants in DNA repair genes and PrCa. The iCOGS chip contained 9,192 SNPs within 10kb of a widely recognised DNA repair gene and at 7 of the 77 known PrCa susceptibility loci these were also significantly associated with risk (Chr3 - *RUVBL1*; Chr5 - *FGF10*; Chr9 - *RAD23B*; Chr11 - *CCND1*; Chr14 - *RAD51B*; Chr20 - *RTEL1*; Chr22 - *BIK*). These may therefore represent potential candidate genes for the aetiology of prostate cancer risk. No statistically significant associations with risk were observed for any of the rare variants (predominantly BRCA1/2 missense SNPs) submitted on the iCOGS chip. We also investigated association between DNA repair gene variants on the iCOGS array and phenotypic characteristics of PrCa; to examine whether any of these variants could modulate risk towards developing aggressive, younger onset or familial disease, rather than increasing the risk of developing PrCa *per se*. We observed several potential associations between low frequency variants in multiple genes and the likelihood of nodal spread or metastases (*ATM*, *FANCL*, *RAD51B*, *RAD54L*, *XRCC3*) but found limited evidence for association with other variables such as age of diagnosis, family history of PrCa, PSA levels or tumour stage.

3265T

Frequency of novel and known deleterious germline variants in rhabdomyosarcoma and neuroblastoma by next-generation sequencing. *D.R. Stewart[1], R. Patidar[2], J. Shern[2], S. Zhang[2], T. Pugh[3], S.J. Diskin[4], S. Sindiri[2], Y.K. Song[2], H. Liao[2], J. Wang[2], S.X. Skapek[5], F.G. Barr[6], R.C. Seeger[7], J.M. Maris[4], D. Hawkins[8], J. Khan[2], J.S. Wei[2].* 1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 2) Oncogenomics Section, Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD; 3) Clinical Genomics Research Program, Princess Margaret Cancer Centre, Toronto, Canada; 4) Children's Hospital of Philadelphia, Philadelphia, PA; 5) UT Southwestern Medical Center, Dallas, TX; 6) Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, MD; 7) Children's Hospital Los Angeles, Los Angeles, CA; 8) Seattle Children's Hospital, Seattle, WA.

INTRO Outside familial syndromes, the pathogenesis of rhabdomyosarcoma (RMS) and neuroblastoma (NB) remains obscure. We sought to identify rare deleterious germline variants that may play a role in the initiation of sporadic RMS and NB. **METHODS/RESULTS** We investigated two cohorts consisting of RMS (n=133) and NB (n=222) patients; the NB data is from the TARGET initiative. We first called high-quality protein-coding changing single nucleotide variants (SNVs) in both paired germline and tumor DNA from the Illumina, SOLiD and CGI platforms (exome and whole-genome sequencing). We excluded variants with minor allele frequency (MAF) > 0.01 in 1000G or NHLBI ESP to yield 56,336 germline variants. To find the most deleterious variants, we filtered for a C-score >= 20 (top 1%) using CADD, a method to integrate many diverse bioinformatics predictions (SIFT, PolyPhen, conservation) into a single score. We then binned variants in 3 categories: 1) variants in known cancer genes and select human syndromes (n = 177), 2) variants reported in >= 3 tumors from TCGA and our somatic NB and RMS data (n = 116), and 3) variants not known to be cancer-associated, have ESP MAF = 0 and a germline variant allele count of 2-8 in our data (n = 159). Of the resulting 446 SNVs (6 variants overlapped categories), we picked the 20 most-frequent and 198 random SNVs for Sanger verification sequencing. Out of the 305 verification sequencing reactions, no DNA was available for 121 sequencing reactions. Of the 184 sequencing reactions, 2 failed PCR, and the verification rate equaled 97%; none of the TARGET NB samples had a false positive. Of the 446 SNVs, there were 2/133 (1.5%) germline variants in 2 known RMS-associated genes (*TP53*, *CTCF*) and 10/222 (4.5%) germline variants in 4 known NB-associated genes (*ALK*, *NF1*, *PINK1*, *BARD*). We also observed 3/133 (2%) germline *ALK* variants in 3 patients with RMS and 1/222 (each) germline *TP53* and *BUB1B* variants in 2 NB patients. Known reported somatic mutations in NB with germline variants in NB patients included *NRAS* (1/222, 0.5%) and *VANGL1* (4/222, 2%). Known reported somatic mutations in RMS with germline variants in RMS patients included *FGFR4* (1/133, 1%) and *BCOR* (1/133, 1%). We also observed predicted deleterious germline variants in *SDHB* (NB), *DICER1* (NB), *ERCC2/3/4* (NB and RMS) and *TSC1/2* (NB and RMS). **CONCLUSION** In this high-quality dataset using stringent filtering, we found multiple novel germline variants associated with NB and RMS.

3266S

A recurrent germline mutation in the splicing factor *SRRM2* gene is implicated in papillary thyroid cancer predisposition. *J. Tomsic[1], H. He[1], K. Akagi[1], S. Liyanarachchi[1], Q. Pan[2], B. Bertani[1], R. Nagy[3], D. Symer[1,3,4], B. Blencowe[2,5], A. de la Chapelle[1].* 1) Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, OH, USA; 2) Banting and Best Department of Medical Research, University of Toronto, Toronto, Canada; 3) Department of Internal Medicine, The Ohio State University, Columbus, OH, USA; 4) Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA; 5) Department of Molecular Genetics, University of Toronto, Toronto, Canada.

Purpose: Papillary thyroid carcinoma (PTC) is highly heritable, but the few predisposing genes identified until now account for relatively few cases. In this study we identified a candidate gene for PTC risk via next generation sequencing in PTC families. **Patients and Methods:** Next generation sequencing was used on two affected individuals from each of seven PTC families. Whole exome sequencing was carried out using HiSeq 2000 platform and the Illumina Genome Network (IGN; http://www.illumina.com) after exome enrichment. The Illumina short reads were aligned using Burrows-Wheeler Transform and genomic variants were called using Samtools/BCFtools. Several candidate DNA variants were tested in further PTC patients and controls. Linkage analysis was performed in the same families. Functionality of the most highly ranked candidate DNA variant was tested. **Results:** Whole exome sequencing, linkage analysis and population studies identified the serine/arginine repetitive matrix 2 gene (*SRRM2*) as a strong-appearing candidate gene in one family. The heterozygous missense mutation c.1037C→T (Ser346Phe) cosegregated with PTC in the family and was not found in 132 other families. The variant occurred in 7/1170 sporadic PTC cases and in 0/1404 controls (p=0.004). *SRRM2*, also known as *SRm300*, is part of the splicing machinery. The effect of the missense mutation on alternative splicing (AS) was tested using RNA-Seq on RNA extracted from blood cells. Significant differences in AS occurred in numerous genes between three unrelated mutation carriers and three controls. In a limited number of genes the RNA-Seq results were verified experimentally using RT-PCR demonstrating a higher ratio of inclusion of exons in RNA from carriers of the S346F mutation. **Conclusions:** Our data suggest that *SRRM2* is a candidate gene for PTC susceptibility. SRRM2 S346F is a rare, medium-penetrance mutation affecting alternative splicing.

3267M

BCL7B functions as a tumor suppressor in the Wnt signaling pathway. *T. Uehara[1], E. Kage-Nakadai[1], S. Yoshina[1], R. Imae[1], S. Mitani[1,2].* 1) Physiology, Tokyo Women's Medical University, Tokyo, Japan; 2) Tokyo Women's Medical University Institute for Integrated Medical Sciences, Tokyo, Japan.

Introduction; Cytogenetic abnormalities of chromosome 7 occur frequently in patients with cancer. Although some of the genes located on chromosome 7 are thought to act as tumor-related genes, little is known about the details of them. Williams-Beuren syndrome (WBS) is one of the most well-known diseases related to chromosome 7 microdeletions and thought to be at an increased risk of malignant transformation due to aberrations in candidate genes, such as *BCL7B*. However, there is a little knowledge regarding the specific functional roles of *BCL7B*. Here, we analyzed the functional significance of *BCL7B* in KATOIII human gastric cancer cells and its homolog in *Caenorhabditis elegans* (*C. elegans*). **Methods**; First, to analyze the function of *bcl-7*, which is the homolog of *BCL7B* in *C. elegans*, we generated a *bcl-7* deletion mutant, and observed and analyzed its phenotype. As a result, we found that *bcl-7* is required for the asymmetric differentiation of epithelial "stem-like" seam cells through the WNT pathway. This result suggested that BCL-7 functions as a negative regulator of the WNT pathway in *C. elegans*. In addition, *bcl-7*-deletion mutants exhibited nuclear enlargement, which was reminiscent of the anaplastic features of malignant cells. Second, to analyze the function of *BCL7B* in human cell line, we used KATOIII cells, which were derived from gastric signet-ring cell cancer and expressed only *BCL7B* of the *BCL*7 family members. As a result, *BCL7B*-knockdown induced nuclear enlargement, as observed in *C. elegans*, and promoted the multinuclei phenotype. Moreover, we showed that RNA levels were increased in the cells with enlarged nuclei. *BCL7B* knockdown also suppressed cell death by inhibiting the apoptotic pathway. In addition, our study showed that *BCL7B* negatively regulates the WNT-signaling pathway in KATOIII cells similar to that in *C. elegans*. **Results and Discussion**; In conclusion, we investigated the functional role of *bcl-7*/*BCL7B* genes in WNT signaling in *C. elegans* and in human gastric cancer cells. Additionally, human *BCL7B* may function as a negative regulator of Wnt signaling and a positive regulator of apoptosis in KATOIII cells. Furthermore, this study revealed that BCL-7/BCL7B is also involved in the mechanisms of nuclear enlargement, which is an important signature of malignancy. Collectively, our data suggests that the members of the BCL7 family and their homologous proteins may function as tumor suppressors by affecting multiple pathways.

3268T

Investigating the genetic basis of multiple primary tumors. Clinical and gene panel analyses. *J. Whitworth[1,2], J. Hoffman[1], AB. Skytte[3], ER. Maher[1,2,4].* 1) Clinical Genetics, Birmingham Women's Hosptial, Birmingham, United Kingdom; 2) Medical Molecular Genetics, University of Birmingham, Birmingham, United Kingdom; 3) Clinical Genetics, Aarhus University Hospital, Aarhus, Denmark; 4) Medical Genetics, University of Cambridge, Cambridge, United Kingdom.

Multiple primary malignant tumors (MPMT) are frequently taken as an indicator of potential inherited cancer susceptibility and occur at appreciable frequency both among unselected cancer patients and referrals to cancer genetics services. Analysis of a referral based series of 212 MPMT cases showed that only around 40% of those who underwent genetic testing and 20% of referrals overall were identified as having a pathogenic germline variant conferring predisposition to malignancy. Comparison of individuals who tested positive and negative revealed considerable overlap between the two groups with respect to clinical characteristics indicative of an inherited cancer syndrome, suggesting that many of the latter group also have a genetic basis. Analysis of PTEN and TP53 by Sanger sequencing, however, did not reveal any significant variants. Failure to detect a causative mutation may result from mosaicism for a mutation in a known inherited cancer gene, an unusual phenotype that leads to the relevant gene being overlooked or mutation in a novel inherited cancer gene. To address these possible explanations, next generation sequencing techniques are being applied to blood samples from an expanded series of mutation negative MPMT cases (≥2 cancers diagnosed before 60 years) ascertained through clinical genetics services. Initial analysis is being performed using a panel of 94 known inherited cancer genes. Results from the first 63 samples produced 309 variants, 10 of which could be assigned as pathogenic in the heterozygous state based on literature review and predicted effect. These 10 variants were identified among 8 cases (12.7%) and all but one was associated with a characteristic tumor. One case harboured three separate mutations in SDHC, FLCN and NF1, all of which were penetrant. In silico predictions revealed multiple further variants as potentially causative of the patient's phenotype. Individuals not identified with explanatory variants through panel analysis are being entered into a study conducting whole genome sequencing with the aim of identifying potential novel inherited cancer genes with subsequent functional validation. A sequencing dataset from large heterogeneous series such as this should provide opportunity for analysis of multiple phenotypic subsets and allow for potential pleiotropic effects of unidentified genes relevant to cancer predisposition.

3269S

Characterization of T gene sequence variants and germline duplications in familial and sporadic chordoma. *R.X. Yang[1], M.J. Kelley[2,3], J. Shi[1], B. Ballew[1], P.L. Hyland[1], W-Q. Li[1], M. Rotunno[1], D.A. Alcorta[2], N.J. Liebsch[4], J. Mitchell[5], S. Bass[5], D. Roberson[5], J. Boland[5], M. Cullen[5], J. He[5], L. Burdette[5], M. Yeager[5], S.J. Chanock[1], D.M. Parry[1], A.M. Goldstein[1].* 1) Division of Cancer Epidemiology & Genetics, NCI/NIH, Bethesda, MD; 2) Department of Medicine, Duke University Medical Center, Durham, NC; 3) Durham Veterans Affairs Medical Center, Durham, NC; 4) Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA; 5) Cancer Genomics Research Laboratory, SAIC-Frederick, Inc., Bethesda, MD.

Background Chordoma is a rare bone cancer that is believed to originate from notochordal remnants. We previously identified germline T duplication as a major susceptibility mechanism in several chordoma families. Recently, a common genetic variant in T (rs2305089) was significantly associated with the risk of sporadic chordoma. Methods We sequenced all T exons in 24 familial cases and 54 unaffected family members from eight chordoma families (three with T duplications), 103 sporadic cases, and 160 unrelated controls. We also measured T copy number variation in all sporadic cases. Results We confirmed the association between the previously reported variant rs2305089 and risk of familial (odds ratio [OR]=2.6, 95% confidence interval [CI]=0.93, 7.25, P=0.067) and sporadic chordoma (OR=2.85, 95%CI=1.89, 4.29, P<0.0001). We also identified a second common variant, rs1056048, that was strongly associated with chordoma in families (OR=4.14, 95% CI=1.43, 11.92, P=0.0086). Among sporadic cases, another common variant (rs3816300) was significantly associated with risk when jointly analyzed with rs2305089. The association with rs3816300 was significantly stronger in cases with early age-onset. In addition, we identified three rare variants that were only observed among sporadic chordoma cases, all of which have potential functional relevance based on in silico predictions. Finally, we did not observe T duplication in any sporadic chordoma case. Conclusions Our findings further highlight the importance of the T gene in the pathogenesis of both familial and sporadic chordoma and suggest a complex susceptibility related to T.

3270M

Exome sequencing identified potential causative candidate genes for hyperplastic polyposis syndrome. *S. Aretz[1], C. Trueck[1], J. Altmueller[2,3], S. Horpaopan[1], P. Hoffmann[1,4,5], H. Thiele[2,3], I. Spier[1].* 1) Institute of Human Genetics, University Hospital Bonn, Bonn, Germany; 2) Cologne Center for Genomics, University of Cologne, Germany; 3) Institute of Human Genetics, University of Cologne, Germany; 4) Department of Genomics, Life & Brain Center, University of Bonn, Germany; 5) Division of Medical Genetics, University Hospital Basel; Department of Biomedicine, University of Basel, Basel, Switzerland.

Purpose: Hyperplastic polyposis syndrome (HPS), also known as serrated polyposis syndrome (SPS), is a yet poorly defined colorectal cancer (CRC) predisposition characterized by the occurrence of multiple and/or large serrated lesions throughout the colon. A serrated polyp-CRC sequence (serrated pathway) has been postulated, however, to date, only few molecular signatures of serrated neoplasia were described and neither the etiology of the syndrome nor the distinct genetic alterations during tumorigenesis have been identified. Methods: To uncover predisposing causative genes, the exomes of 11 unrelated and clinically well characterized HPS patients with sporadic appearance were sequenced (Illumina HiSeq platform). The variants were filtered for rare truncating germline mutations assuming a monogenic disease model. For data analysis and variant filtering the GATK software and in-house tools (VARBANK pipeline) were applied. Results: Altogether, 261 rare truncating germline variants were found. After stringent filtering steps including quality scores, the comparison with large datasets from population-based controls, detailed manual investigations of the variants and data mining according to functions and pathways, 139 unique variants in 136 genes remained. Of those, six genes were affected by biallelic variants (recessive model) in at least one patient and 19 genes by heterozygous variants (dominant model) in at least two patients. The majority of these genes is supposed to be associated with cancer or is involved in molecular and cellular functions related to tumorigenesis such as DNA repair, apoptosis, or cell proliferation. Another 53 genes, which are affected by heterozygous variants in only one of the patients, are regarded as interesting candidates according to functional scores and known somatic mutations in cancer. Conclusions: Using exome sequencing we identified several potentially causative genes for HPS. The further work-up includes the validation of all variants by Sanger sequencing and pathway analyses. The clinical relevance of the genes will be clarified in a cohort of another 20 HPS patients.

3271T

Identification of the frequent hereditary cancer mutations in high-risk non-BRCA breast cancer patients from Puerto Rico. *J. Dutil[1], J.K. Teer[2], S.J. Yoder[2], J.L. Matta[3], A. Monteiro[4].* 1) Biochemistry Dpt, Ponce School of Medicine & Health Sciences, Ponce, PR; 2) Dpt of Biomedical Informatics, Moffitt cancer Center, Tampa, FL; 3) Physiology and Pharmacology Dpt, Ponce School of Medicine & Health Sciences, Ponce, PR; 4) Dpt of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL.

Approximately 5-10% of all breast cancers are caused by mutations in highly penetrant susceptibility genes, such as BRCA1 and BRCA2. Deleterious mutations in genes such as PTEN, p53, CHEK2, ATM, NBS1, RAD50, BRIP1 and PALB2 also confer moderate to high risk of breast cancer, and together with BRCA1 and BRCA2, they may explain up to 50% of breast cancer risk. Hispanics make up 16.3% of the population, and accounted for 56% of the national growth between 2000 and 2010. Yet, these populations are often underrepresented in genetics studies. Our previous work has shown that the BRCA mutation spectrum underlying hereditary breast cancers in Puerto Rico is distinct from that of other Hispanic populations such as Mexican Americans. Our objective was to identify the common germline mutations underlying non-BRCA breast cancer in the Hispanic population from Puerto Rico. Thirty-two breast cancer patients that met the clinical criteria for BRCA testing but had received a negative BRCA1/2 result were recruited in this study. Exons and intron-exons junctions of 40 tumor suppressor genes that have been previously implicated in hereditary cancer predisposition were captured using the BROCA cancer risk panel. Massively parallel sequencing was performed on an Illumina MySeq platform. A total of 5,846 variants were identified within the targeted gene regions. Each patient carried on average 72 coding variants, which included 41 synonymous variants and 31 non-synonymous variants. Work is underway to prioritize variants according to their likelihood of being a causative deleterious mutation. Our findings may uncover genetic changes specific to this population that will allow for the development of custom-designed predisposition tests, taking into account natural differences in populations to maximize the identification of individuals at risk.

3272S

Deleterious mutations in multiple cancer-risk genes in individuals from a high-risk cancer genetics clinic. *C.M. Laukaitis[1,2], C. Mauss[2], M. Urquijo[2,3], A. Chaudhury[1], K. Maher[1,2], J. Jeter[1,2].* 1) Department of Medicine, University of Arizona College of Medicine, Tucson, AZ; 2) University of Arizona Cancer Center, Tucson, AZ; 3) Departamento de Medicina, Universidad de Sonora, Hermosillo, Sonora, Mexico.

The widespread use of massively-parallel next generation sequencing (NGS) technology has led to surprising findings. We are using NGS to identify inherited genetic changes predisposing to cancer in people cared for by the University of Arizona Cancer Center high-risk cancer genetics clinic. Under an IRB-approved research protocol, we sequenced a custom panel of 104 cancer-related genes in blood samples from 48 individuals. This was a heterogeneous population containing both patients who were diagnosed with cancer and people without a personal history of cancer but with a strong cancer family history. Surprisingly, 6 individuals (12.5 percent) carried clearly deleterious mutations in multiple genes linked to increased cancer risk. While 35 percent of the cohort has been diagnosed with cancer, five of six people with multiple deleterious mutations have had cancer ($p = 0.017$). Individual mutations are being validated by Sanger sequencing. Additional clinical characteristics (participant age, function of mutated genes, type of cancer, family history of cancer, type of cancer treatment) are being evaluated to identify associated characteristics in order to determine whether this evidence of genomic instability is the cause or an effect of the cancer diagnosis.

3273M

Whole exome sequencing to identify novel breast cancer susceptibility genes. *K.N. Maxwell[1], J. Vijai[2], K. Schrader[2], L. Guidugli[3], S. Hart[3], T. Thomas[2], X. Wang[3], B. Wubbenhorst[4], S.M. Domchek[1], M.E. Robson[5], S. Neuhausen[6], K. Offit[2], C. Szabo[7], J. Weitzel[6], K.L. Nathanson[4], F.J. Couch[3].* 1) Department of Medicine, Division of Hematology-Oncology, University of Pennsylvania, Philadelphia, PA; 2) Clinical Genetics Research Lab, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 3) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 4) Department of Medicine Division of Translational Medicine and Genetics, University of Pennsylvania, PA; 5) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 6) Beckman Research Institute of City of Hope, Duarte, CA; 7) National Institutes of Health.

Over 50% of the familial relative risk of breast cancer is unexplained by currently identified common low risk polymorphisms and rare moderate to high risk alleles in genes such as BRCA1/2. Early identification of the individuals with an inherited susceptibility to breast cancer may lead to enhanced screening and prevention strategies and the opportunity for targeted therapies of associated cancers. In order to identify novel genes which confer an inherited risk to breast cancer, whole exome sequencing was undertaken in 333 BRCA1/2 negative individuals with at least one primary breast and ovarian cancer or with high-risk familial breast cancer, defined as a proband and at least two first or second degree relatives with breast cancer. Samples included 256 individuals from 109 families and 77 unrelated individuals, for a total of 186 independent cases. Data were first analyzed for identification of rare mutations in 49 cancer susceptibility genes. Eight independent cases (4.3%) were found to carry deleterious mutations in high penetrance cancer susceptibility genes, namely TP53, CDH1, MSH6, and CDKN2A. In addition, sixteen of 186 independent cases (12.3%) were found to have deleterious mutations in other breast and/or ovarian cancer predisposition genes, namely ATM, CHEK2, PALB2, BARD1, RAD51D, RAD50, and MRE11A. For novel cancer susceptibility gene discovery, whole exome data was harmonized from three separate sites into a single vcf file for variant filtering and analysis. Two major methodologies were used to select genes for further study, a candidate gene approach and an agnostic gene ranking approach. In the candidate gene approach, genes matching either specific gene ontologies or those identified by published somatic tumor studies were queried for the presence of rare deleterious mutations in the dataset. In the agnostic gene ranking approach, genes were ranked based on the frequency of specific types of variants in the dataset versus population estimates and the level of segregation of these rare variants in families. These methodologies identified 636 candidate cancer susceptibility genes in the discovery exome set. These genes will be evaluated by a targeted sequencing approach in a validation set of over 2000 individuals with familial breast cancer. Overall, our data demonstrate a significant level of heterogeneity in the genetic basis of familial breast cancer in BRCA1/2 negative individuals.

3274T

Germline epigenetic inactivation of BAP1 in a subset of patients with uveal melanoma. *R. Pilarski[1], G. Boru[2], O.H. Saqr[2], K. Rai[1], J.B. Massengill[2], C.M. Cebulla[2], M.H. Abdel-Rahman[1,2,3].* 1) Dept of Internal Medicine and Comprehensive Cancer Center, Ohio State Univ, Columbus, OH; 2) Dept of Ophthalmology, Ohio State Univ, Columbus, OH; 3) Dept of Pathology, Menoufiya Univ, Egypt.

Objective: To identify the frequency of germline deletions and/or epigenetic inactivation in BAP1 in uveal melanoma (UM) patients with high risk for BAP1 tumor predisposition syndrome (TPDS) but no detectable mutation. Methods: Twenty three UM patients with high-risk for BAP1 TPDS, including 12 UM patients with family history of UM, one UM patient with bilateral disease and 10 UM patients with personal or family history of renal cell carcinoma (RCC), were studied. Constitutional decrease in BAP1 mRNA expression was assessed in an additional 17 UM patients with available non-tumor choroidal tissue. Germline copy number variation was studied by Multiplex Ligation-dependent Probe Amplification (MLPA). Methylation specific PCR and pyrosequencing was used to assess germline promoter hypermethylation. The expression of BAP1 mRNA was carried out utilizing quantitative RT-PCR. Results: No germline CNV or significant promoter methylation was detected in any of the 23 patients with high-risk for BAP1 TPDS. Two out of the 17 non-tumor choroidal tissues showed significant decreases in BAP1 expression compared to normal controls, but no germline CNV or aberrant promoter methylation were identified. Conclusion: Epigenetic inactivation of BAP1, through mechanisms other than promoter hypermethylation, could be responsible for germline constitutional inactivation of BAP1 in a subset of UM patients. Copy number variation is not a major mechanism for germline inactivation of BAP1. Our results have important implications for designing clinical assays for detection of BAP1 germline inactivation in patients at risk for the BAP1 TPDS.

3275S

Integrating Whole Genome and Exome Sequencing with Structural Variation Analysis to Identify Potential Causative Mutations in Patients with Cancer Phenotypes Suggestive of Li-Fraumeni Syndrome. *D.I. Ritter[1,2], B. Powell[2], J. Bojadzieva[3], D.A. Wheeler[1,2], R. Gibbs[1,2], L.C. Strong[3], S.E. Plon[2,4].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Cancer Genetics, University of Texas MD Anderson Cancer Center, Houston, TX; 4) Department of Pediatrics, Texas Children's Cancer Center, Baylor College of Medicine, Houston, TX.

While mutations in *TP53* explain ~70% of the inherited cancer condition Li-Fraumeni Syndrome (LFS), a sizeable fraction of families or individuals with cancer phenotypes suggestive of *TP53* mutations remains unexplained. To identify additional candidate genes underlying cancer phenotypes suggestive of LFS, we applied an integrative analysis of whole exome rare coding variants with whole genome structural variants (SV) and copy number variation (CNV) for 27 subjects who were negative for mutations in *TP53*, including family members when available (four kindreds). Probands demonstrated (1) childhood sarcoma and at least one additional primary malignancy by age 40 or (2) a rare childhood cancer with at least one first-degree relative with a rare cancer. Whole exome sequence data (VCRome2.1) were analyzed for rare heterozygous single nucleotide variants (SNVs) and insertions/deletions (INDELs), encoding missense, nonsense, frameshift and splice variants. We eliminated variants >1.5% in 1079 unaffected normal samples, and >1% dbSNP minor allele frequency. Variants were limited to those shared among affected family members, when available. We applied a proportions test to variants in genes and exons against 100 random controls. To identify structural variants, we applied complementary algorithms of insert-size (Breakdancer) and soft-clip stacking (CREST). We used an empirical annotative filtering strategy against normal, unaffected whole genomes. Structural variants were progressively clustered to identify recurrence and CNVs. We prioritized SV events in genes harboring rare variants, interrupting coding regions, introns or <=10kb of genes. We identified 110 unique genes harboring rare coding variants that affect >=4 independent families. The top genes of interest are *TEP1* (gene: p=3.6E-3, exon 30: p=0.02) and *ATR* (gene: p=0.02, exon38 p=4.9E-4). We found no rare coding SNVs or INDELS within *TP53*, nor in *CHEK2*, a gene known to be mutated in a subset of LFS phenotypes. We found no evidence for SV within 1MB of either of these genes. We have identified multiple additional regions of rare SVs and CNVs affecting alternative genes, and are combining these results with exome variant data. With this effort, we aim to identify additional genetic causes for pediatric inherited cancer syndromes with phenotypes suggestive of Li-Fraumeni. Supported by RP10189 from CPRIT and R01-CA138836 to SEP, and K12GM084897 from IRACDA to DIR.

3276M

Causative novel POLE mutations in hereditary colorectal cancer syndromes. *A.M. Rohlin[1], T. Zagoras[1], F. Eiengård[1], Y. Engwall[1], S. Nilsson[2], U. Lundstam[3], J. Björk[4], G. Karlsson[5], M. Nordling[1].* 1) ,Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden; 2) Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden; 3) Department of Surgery, Sahlgrenska Academy at University of Gothenburg, Sahlgrenska University Hospital/Östra, Gothenburg, Sweden,; 4) The Swedish Polyposis Registry, Department of Medicine, Karolinska Institute, Stockholm, Sweden,; 5) The Swedish NMR-centre, University of Gothenburg, Gothenburg, Sweden.

Background: In Familial Adenomatous Polyposis (FAP) it is today possible to find almost all of the disease-causing mutations. However, in patients with a more attenuated phenotype (AFAP) with less than 100 polyps only a fraction of the disease-causing mutations can be identified. The low detection rate implicates the probable presence of additional disease-causing genes still to be identified. FAP is caused by autosomal dominantly inherited mutations in the APC (Adenomatous polyposis coli) gene. In AFAP families 20%-30% of the patients have a germline APC mutation. Recently a new CRC syndrome, polymerase proof reading associated polyposis (PPAP) was described. This syndrome is characterized by a dominantly inherited predisposition to the development of a variable number of colorectal adenomas and carcinomas [1] The aim of this study was to sequence the exonuclease domain of POLE in 88 index patients with a familial history of polyposis or non-polyposis and/or early onset CRC that had previously tested negative for mutations in APC, MUTYH and/or mismatch-repair genes MSH2, MLH1, MSH6 and PMS2. Method: In one large family exome sequencing was performed on four family members and for the remaining 87 index patients mutation screening by Sanger sequencing of POLE exonuclease domain (ex 3-14) was conducted. Results and Conclusion: We have identified two novel mutations in the exonuclease domain of POLE. The first mutation was identified from exome sequencing in a large Swedish family with CRC. The POLE: c.1089C>A, p.Asn363Lys mutation is directly involved in DNA binding. Family members carrying this mutation demonstrate a high penetrant predisposition not only to CRC but also to extra-intestinal tumours such as ovarian, endometrial and brain tumours [2]. The second mutation located in POLE: c.1274A>G, p.Lys425Arg was found in a patient with early onset CRC. Theoretical prediction of the amino acid substitution suggests a profound effect of the substrate binding capability and a severe impairment of the catalytic activity for both these mutations, which strongly suggest a pathogenic nature of these mutations. Screening the proofreading domains of POLE should be considered in routine genetic diagnostics in families with hereditary CRC. 1.Palles C, Nature genetics 2013, 45(2):136-144. 2.Rohlin A, International journal of oncology 2014, 45(1):77-81.

3277T

Parental Inheritance and *WT1* Abnormality Types May Affect the Penetrance Rate of Hereditary Wilms Tumor. *Y. Kaneko[1], H. Okita[2], M. Haruta[1], Y. Arai[3], T. Oue[4], T. Koshinaga[5], M. Fukuzawa[6], Japan Wilms Tumor Study Group.* 1) Research Institute for Clinical Oncology, Saitama Cancer Center, Ina, Saitama; 362-0806 Japan; 2) National Institute for Child Health and Development; Hematology/Oncology; 2-10-1, Ohkura, Setagaya-ku, Tokyo; 157-0074; 3) National Cancer Center Research Institute; 5-1-1 Tsukiji, Chuo-ku, Tokyo; 104-0045; 4) Osaka University Graduate School of Medicine; 2-15 Yamadaoka, Suita, Osaka; 565-0871; 5) Nihon University; 30-1 Ohyaguchi-Kamimachi, Itabashi-ku, Tokyo 173-8610; 6) Osaka Medical Center and Research Institute for Maternal and Child Health; 840 Murodo-machi, Izumi, Osaka 594-1101.

Background: Wilms tumor (WT) is a genetically heterogeneous disease which arises from developmental kidney. The percentages of hereditary WT are unknown, and only some hereditary WTs are thought to be caused by a germline mutation in *WT1*. *IGF2* is an imprinted gene expressed by the paternal allele. Both *WT1* and *IGF2* genes are located on the short arm of chromosome 11 (11p), and uniparental disomy (UPD) on 11p was found in 30-40% of WT, and regularly accompanied by maternal allele loss and paternal allele duplication. The inheritance of *WT1* mutations has been poorly studied in familial WTs, although carriers with *WT1* mutations are now increasing because multidisciplinary therapies have improved the survival rates of patients with bilateral WTs and those with a unilateral WT with a germline *WT1* mutation. We examined the statuses of *WT1* and *IGF2* in bilateral and familial WTs in Japan, and summarized the present and previous findings on the penetrance rate for children who inherited various types of *WT1* abnormalities from their fathers or mothers, or had *de novo WT1* abnormalities that occurred in the paternal or maternal germ cell. **Results:** We detected *WT1* abnormalities in 25 (81%) of 31 patients with bilateral WTs and 2 of 2 families with hereditary WT. Of 35 WTs from the 25 patients, 31 had duplications of a small *WT1* mutation and paternal *IGF2* caused by 11p UPD, while 4 had large deletions in one *WT1* allele and small mutations or deletions in the other with the retention of 11p heterozygosity. The high incidence of *WT1* abnormalities in bilateral WTs in Japan sharply contrasts with the lower incidence in America (27%). The penetrance rate was shown to be 100% if children inherited small *WT1* mutations from their fathers, and 67% if inherited the small mutations from their mothers, or inherited 11p13 deletions or had *de novo* 11p13 deletions irrespective of parental origin (P=0.057). The duplication was identified in most WTs with small *WT1* mutations of paternal origin, but not in WTs with large 11p13 deletions irrespective of parental origin. **Conclusion:** We hypothesize that individuals who inherited small *WT1* mutations from their fathers may be more likely to develop WT due to the expressing *IGF2* than those who inherited small mutations from their mothers or inherited large 11p13 deletions or had *de novo* large deletions irrespective of parental origin. These findings may be useful for the genetic counseling of individuals who may inherit *WT1* mutations.

3278S

Functional characterization of the 19p13 breast and ovarian cancer risk locus identifies *ABHD8* as a novel candidate breast-ovarian cancer susceptibility gene. *J. Beesley[1], K. Lawrenson[2], K. Kuchenbaecker[3], S. Kar[3], Q. Li[4], S. Edwards[1], J. French[1], H. Shen[2], S.J. Ramus[2], J.M. Lee[2], T.J. Spindler[2], D. Hazelett[2], S.K. Rhie[2], J. Simard[5], F. Couch[6], A. Dunning[3], G. Coetzee[2], M. Freedman[4], D. Easton[3], G. Chenevix-Trench[1], P.P. Pharoah[3], A.C. Antoniou[3], S.A. Gayther[2] on behalf of CIMBA, BCAC and OCAC.* 1) Department of Genetics and Computational Biology, QIMR Berghofer, Herston, QLD, Australia; 2) University of Southern California, Los Angeles, CA, USA; 3) University of Cambridge, Strangeways Research Laboratory, Cambridge, UK; 4) Dana-Farber Cancer Institute, Boston, MA, USA; 5) Centre Hospitalier Universitaire de Québec (CHUQ), Québec, Canada; 6) Mayo Clinic, Rochester, MN, USA.

Multiple independent genome-wide association studies (GWASs) have identified variation at the 19p13 locus that predisposes to estrogen receptor-negative breast cancer and high-grade serous ovarian cancer (HGSOC). Fine-mapping of the region using the iCOGS customized Illumina Infinium genotyping array, followed by imputation, limited the risk associated region to a 20 kb window encompassing two protein coding genes - *ANKLE1*, which may be involved in DNA repair, and *ABHD8* of unknown function - and just distal to the *BRCA1*-interacting gene *BABAM1*. The strongest association signal is represented by 13 highly correlated single nucleotide polymorphisms ($P \sim 10^{-22}$). By integrating genotyping data with epigenetic data generated from normal ovarian and breast epithelial cells we identified risk SNPs intersecting known or putative enhancers, and ELF1, ELK4, GABP and GATA3 transcription factor binding sites. Expression quantitative trait locus (eQTL) analyses identified significant genotype-gene expression associations for *ANKLE1* in normal ovarian epithelial cells (P = 0.02), and for *ABHD8* in both HGSOCs ($P = 3.0\times10^{-5}$) and normal breast tissues ($P = 2.8\times10^{-3}$), but not for *BABAM1*. Risk associated SNPs were also significantly associated with allele specific expression of *ABHD8* in breast cancer ($P = 2.5\times10^{-5}$). *ABHD8* (but not *ANKLE1* or *BABAM1*) was prioritized in a co-expression network analysis based on expression data in primary HGSOCs. Genes in the 19p13 region are frequently overexpressed in both breast cancers and HGSOCs; however, stable overexpression of neither *BABAM1* nor *ANKLE1* had a significant impact on the neoplastic phenotype of normal breast and ovarian epithelial cells. Using chromosome-conformation-capture (3C) assays in these same cell types, we identified interactions between a putative regulatory region and the promoter of *ABHD8*, as well as in breast and ovarian cancer cell lines. Taken together, these data suggest that the same functional mechanism at this locus underlies breast and ovarian cancer development, and that *ABHD8* may represent a novel breast/ovarian cancer susceptibility gene.

3279M

Identification of germline mutations in *TEP1* among familial and sporadic pediatric acute myelogenous leukemia cohorts. *N.R. Oak[1], D.I. Ritter[3], B.C. Powell[1], H.C. Cheung[1], M.M. Gramatges[2], D.A. Wheeler[1,3], S.E. Plon[1,2,3].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

Identification of cancer susceptibility gene mutations can impact surveillance for second primary malignancies and risk in family members. Germline mutations in genes related to telomere maintenance are associated with bone marrow failure and hematologic malignancies. To explore genetic predisposition to pediatric cancers, we performed whole exome sequencing of constitutive DNA from childhood cancer patients including seven kindreds with familial leukemia/lymphoma in at least two first degree relatives. We filtered for rare variants (<1% in dbSNP or in another non-cancer database, Atherosclerosis Risk in Communities (ARIC)) that are shared between affected individuals within a family and were predicted to exert deleterious effects on the protein. We limited our analysis to truncating (nonsense and frameshift) mutations. In one of the families, we found 7 frameshift and 4 nonsense mutations. The most interesting candidate among these was the novel variant p.R314X in *TEP1*. Loss-of-function mutations in *TEP1* are very rare in sequence databases including the NHLBI exome sequencing project. This rare mutation was shared by a parent-child pair affected by acute lymphocytic leukemia (ALL) and acute myeloid leukemia (AML) respectively. We further explored whether germline mutations in *TEP1* are found in larger cohort of apparently sporadic AML patients. AmpliSeq™ sequencing of 43 genes related to telomere maintenance and DNA repair pathways was performed on DNA samples from 82 pediatric AML patients. This analysis revealed 10 rare missense germline mutations in *TEP1*. Four of these 10 mutations predicted to have damaging effect on protein function according to Combined Annotation Dependent Depletion (CADD) tool (CADD scaled score of 15 or more). The *TEP1* familial p.R314X truncating mutation in addition to germline *TEP1* missense mutations in sporadic AML patients warrant further functional studies. These studies will include assessing the role of these *TEP1* mutations on the telomere length phenotype of cells from the patients carrying the mutation. The results of these studies are designed to determine whether *TEP1* represents another member of the telomere maintenance genes to be associated with AML predisposition. This work was supported by research grant RP10189 from the Cancer Prevention and Research Institute of Texas and R01-CA138836 to Plon, SE and training grant T32 GM007526 to Powell, BC.

3280T

***Hdac9* Intronic Enhancer Variants as Candidates for Skin Cancer Risk.** *A. Toland[1,3], T. Siekmann[2,3].* 1) Human Cancer Genetics; 2) Biomedical Sciences Program; 3) The Ohio State University, Columbus, OH.

The susceptibility to skin tumorigenesis 5 (*Skts5*) locus on mouse chromosome 12 was mapped through linkage analysis of skin tumor susceptible Mus musculus (NIH/Ola) and skin tumor resistant outbred Mus spretus (SPRET/Out) mice. Expression and sequence analysis of genes at *Skts5* led to the identification of *Hdac9* as a potential candidate for *Skts5*; *Hdac9* contains both amino acid variations and differential expression in skin between the strains. Furthermore, we found that variants in human *HDAC9* show allele-specific imbalance in human cutaneous squamous cell carcinomas (cSCC), suggesting a role for this gene in human cSCC. Interestingly, studies by others identified an exonic/intronic enhancer in *HDAC9* that impacted expression of *TWIST1*. From these data we hypothesized that mouse *Hdac9* might also contain an enhancer element and that variants in this region might contribute to differential expression of the oncogene, *Twist1*. To test this hypothesis we performed sequencing analysis and identified 45 sequence variants between NIH/Ola and SPRET/out mice from the orthologous region of the human *HDAC9* enhancer. We subcloned this region into nine segments; two of these segments differentially impacted luciferase expression in vitro. NIH/Ola clones showed 2-fold increased luciferase expression relative to vector alone or the similar region in SPRET/Out. Furthermore, cells transfected with this segment of the NIH/Ola intron 17 led to a 2.2 fold increase in *Twist1* expression, but the same region in SPRET/Out resulted in no up-regulation of *Twist1*. *In silico* transcription factor analyses identified a number of transcription factors that were predicted to differentially bind NIH/Ola and SPRET/Out variants. Chromatin immunoprecipitation studies of two transcription factors, Gata3 and Oct1, demonstrated differential binding between NIH/Ola and SPRET/Out DNA that fit the in silico predictions. Together these studies show evidence that the mouse orthologous region to a human *HDAC9* enhancer, also acts as an enhancer for *Twist1*. As *Hdac9* intron 17 sequence variants between NIH/Ola and SPRET/Out differentially impacted luciferase expression, *Twist1* expression and Gata3 and Oct1 binding, they are candidates for differences in skin tumor susceptibility locus *Skts5*.

3281S

Allelic imbalance in gene expression as a mechanism in breast cancer development. *I. Pulyakhina[1], M.P.G. Vreeswijk[1], J.F.J. Laros[1,2], C.M. Meijers[1], J.T. den Dunnen[1,2], P. Devilee[1], P.A.C. 't Hoen[1].* 1) Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands; 2) Leiden Genome Technology Center, Leiden University Medical Center, Leiden, the Netherlands.

Differential allelic expression (DAE) refers to differences in expression levels between the two alleles of a gene. DAE can be caused by genetic variation and is known to affect a considerable fraction of genes in the human genome. DAE has been suggested to play an important role in human phenotypic variability, including complex traits and diseases. We hypothesize that DAE contributes to the pathogenesis of breast cancer. Loss of the highest expressed allele of a tumor suppressor gene would lead to substantially lower expression levels of the gene than loss of the lowest expressed allele. This may promote tumorigenesis. We performed whole transcriptome sequencing of paired normal-tumor samples from breast cancer patients and analyzed the allelic ratios for heterozygous, coding SNPs. We focused specifically on SNPs that showed DAE in the normal samples and a switch or loss of DAE in the tumor. We separated genomic variants from RNA editing events and investigated the behavior of allelic imbalance and the underlying molecular mechanism in a subset of these SNPs. 102 SNPs in 60 genes demonstrated significant changes in allelic ratios between normal and tumor tissue. A decrease in allelic imbalance in tumor was most frequently observed. 20% of SNPs showed a complete switch in ratios, where the allele with higher expression in normal cells was lower expressed in tumors and *vice versa*. Multiple coding SNPs in the same gene and the same sample showed consistent changes in allelic imbalance. Genes showed recurrent and consistent changes in allelic imbalance across multiple samples, although sometimes measured by different SNPs. Some of the affected genes have been shown to play a role in cancer development, but contribution of their allelic imbalance has not been described so far. The validation of the findings in a larger cohort and investigation of a possible role for DAE in other cancer types is ongoing.

3282M

Germline Mutations in Men with Multiple Primary Malignancies from a Hereditary Prostate Cancer Cohort. *P.G. Pilie, K. Zuhlke, A. Johnson, K. Cooney.* Internal Medicine, University of Michigan, Ann Arbor, MI.

There exists a variety of well-known familial cancer syndromes originating from germline mutations that lead to multiple primary malignancies in a single individual. Cases of men with early-onset and/or familial prostate cancer in addition to other primary cancers, even including other urogenital cancers, are relatively rare. Furthermore, the molecular underpinnings are not well understood. High-throughput DNA sequencing encompassing large comprehensive cancer whole-gene exome panels allows for the rapid identification of novel, rare germline mutations in cancer patients. To search for these mutations, germline DNA from eight men with prostate cancer who also had a history of pathologically-confirmed multiple primary cancers involving other sites were selected for the study. Each individual provided a cancer family history which was pathologically-confirmed where possible. Blood DNA from the proband was screened using a PCR-based Comprehensive Cancer Panel (Qiagen) DNA sequencing platform. Two of eight probands (25%) from the prostate cancer cohort with multiple primary cancers harbored deleterious germline mutations, one a frameshift mutation in *FGFR3* gene and the other a truncating variant in *ATM* gene. The *FGFR3* mutation was discovered in a man with a personal history of thyroid cancer diagnosed at age 44, prostate cancer diagnosed at age 55, renal cell cancer diagnosed at age 57 as well as significant family history of prostate cancer. Interestingly, multiple first and second degree relatives with only prostate cancer tested negative for *FGFR3* mutation. In summary, our strategy of selecting men with prostate cancer and multiple other primary cancers led to the identification of novel deleterious mutations in *ATM* and *FGFR3* which likely contribute to cancer risk in these individuals and their families. Whereas somatic *FGFR3* mutations have been described in a number of cancers, this is the first report of a *FGFR3* germline mutation that may increase cancer risk and studies are ongoing to understand the biologic implications.

3283T

Genetic analysis of the chromosome 15q25.1 region identifies IREB2 variants associated with lung cancer. *C. Amos[1], I. Gorlov[1], Y. Han[1], L. LeMarchand[2], X. Ji[1], D. Christiani[3], M. Frazier[4], C. Wei[4], J. McKay[5], P. Brennan[5], J. Field[6], Y. Li[1], R. Hung[7].* 1) Dartmouth, Lebanon, NH; 2) University of Hawaii Cancer Center, Honolulu, HI; 3) Harvard University, Boston, MA; 4) U.T. M.D. Anderson Cancer Center, Houston, TX; 5) International Agency for Research on Cancer, Lyon, France; 6) University of LIverpool, Liverpool, UK; 7) University of Toronto, Toronto, CA.

Genome-wide association studies of lung cancer identified the region of chromosome 15q25.1 that includes a nicotinic acetylcholine receptor cluster as being the most strongly associated with lung cancer risk. To characterize the impact that specific functional variants in this region have upon risk for lung cancer development, we performed fine mapping selecting all currently known SNPs influencing lung cancer risk along with all coding SNPs in the 200 megabase region surrounding CHRNA5, a gene known to influence smoking behavior in this region. SNPs were initially identified by Sanger sequencing of 96 individuals for CHRNA5, CHRNA3, CHRNB4, and PSMA4. Additional markers were selected from dbSNP. Markers that were selected for genotyping were chosen based upon the following criteria: known functional effect on activity of genes in the region, validation in African or European populations, position across the region, predicted effect on function, r-square value for LD with respect to other markers being less than 80%. We carried out fine mapping of the 15q25.1 region by genotyping 1395 SNPs extending from the gene CRABP1 to ADAMTS7 from position 79103132 to position 79103132 using a custom Affymetrix Axiom array in 3063 cases and 2940 controls of European ancestry from 5 studies: MSH-PMH, EPIC, MEC, LLPC, HPFS & NHS. Odds ratios (OR) adjusted for age, sex, the first two principal components and site of analysis were estimated using logistic regression. Across this region, 268 markers had p-values less than 0.05 and 101 SNPs met the multiple testing corrected threshold ($p<3.5\times10\text{-}5$). The most significant SNPs lie in a region including of IREB2, with the most significantly associated variant being rs17483686 (OR=1.26, p=8.93 x 10-12). The previously well characterized SNP in CHRNA5, rs16969968, which causes reduced signaling, yielded a less significant association (OR=1.24, p=8 x 10-10) than many variants in IREB2, and iron sensing protein. These findings support an independent effect of IREB2 on lung cancer risk beyond the role of nicotinic receptor variants in this region. Findings will be further evaluated as results from studying additional samples become available in the next month.

3284S

Genome-wide analyses identify gene interaction between SMAD7 and body mass index with risk of colorectal cancer. *P.T. Campbell[1], C. Hutter[2], Y. Lin[3], J. Gong[3], W.J. Gauderman[4], S. Berndt[5], H. Brenner[6,7], G. Casey[4], A.T. Chan[8,9], J. Chang-Claude[10], C. Edlund[4], J. Figueiredo[4], G. Giles[11], L. Le Marchand[12], M. Lemire[13], L. Li[14], P. Newcomb[3,15], F. Schumacher[4], M. Slattery[16], D. Thomas[4], E. White[3,15], M.O. Woods[17], S. Gruber[18], U. Peters[3,15], V. Moreno[19], L. Hsu[3,20].* 1) Epidemiology Research Program, American Cancer Society, Atlanta, GA., USA; 2) Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA; 3) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 4) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 5) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA; 6) Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ) Heidelberg, Germany; 7) German Cancer Consortium (DKTK), Heidelberg, Germany; 8) Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 9) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 10) Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; 11) Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia; 12) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA; 13) Ontario Institute for Cancer Research, Toronto, ON, Canada; 14) Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, USA; 15) Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA; 16) Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT, USA; 17) Discipline of Genetics, Memorial University of Newfoundland, St. John's, NL, Canada; 18) Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 19) Catalan Institute of Oncology and University of Barcelona, Barcelona, Spain; 20) Department of Biostatistics, University of Washington, Seattle, WA, USA.

Body mass index (BMI) is a complex phenotype that is consistently associated with risk of colorectal cancer, although associations are often stronger for men than for women. There is limited information, however, on whether genetic variants interact with BMI to modify colorectal cancer risk. We tested interactions between approximately 2.7M single nucleotide polymorphisms (SNPs) and BMI with colorectal cancer risk among 17,122 colorectal cancer cases and 17,609 controls. We used comprehensive statistical methods with individual-level data to evaluate sex-specific multiplicative interactions between each SNP and BMI with colorectal cancer risk. BMI per 5 kg/m^2 was associated with higher risk of colorectal cancer (men, odds ratio (OR): 1.26; p-value: 4.2×10^{-25}; women, OR: 1.14; p-value: 2.1×10^{-16}). Among men, from traditional case-control logistic regression models, we identified a suggestive interaction between BMI and a locus in *adiponectin receptor 2* (*ADIPOR2*, rs9805042, OR: 0.77; p-value: 9.2×10^{-7}; p-threshold: 5×10^{-8}). Among women, we identified a statistically significant interaction between BMI and a locus in *SMAD7*. Specifically, when using the Cocktail method or the EDGxE method, which each involves two-step screening and testing methods and corrects for multiple testing using weighted hypothesis testing, an interaction was detected between BMI and rs4939827/*SMAD7* (p-observed: 0.0009; p-threshold: 0.005). A statistically significant interaction was also detected between BMI and rs4939827/*SMAD7* using a 2-d.f. joint test (p-value: 2.4×10^{-10}). ORs for the association between BMI per 5 kg/m^2 and colorectal cancer risk differed according to strata of rs4939827 genotype (T/T, OR: 1.08; p-value: 0.02; G/T, OR: 1.13; p-value: 5.4×10^{-6}; G/G, OR: 1.25; p-value: 7.6×10^{-9}). rs4939827/*SMAD7* was previously identified in GWA studies as a colorectal cancer susceptibility locus. Herein, we showed that a common susceptibility locus in *SMAD7* modifies the association between BMI and colorectal cancer risk for women. Further functional analyses are needed to understand this interaction. Future work with larger study samples is warranted to explore the borderline interaction between *ADIPOR2* and BMI in men, especially given the strong biological plausibility for a connection between obesity, circulating adiponectin and colorectal cancer. Our results identify a novel gene-BMI interaction for colorectal cancer risk, which may have implications for disease prevention.

3285M

Estimation of Whole Genome Variations of Hepatocellular Carcinoma among Chronic Hepatitis C Patients. *Y. Chang[1], M. Lee[2], C. Liu[1], H. Yang[3], H. Chen[1], C. Chen[3].* 1) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan; 2) Institute of Clinical Medicine, National Yang Ming University, Taipei, Taiwan; 3) Genomics Research Center, Academia Sinica, Taipei, Taiwan.

Background & Aims: Host genetic susceptibility may be associated with the occurrence of hepatocellular carcinoma (HCC) among patients with chronic infection of hepatitis C virus (HCV). This study aimed to discover genomic variations associated with hepatocellular carcinoma risk through the genome-wide association study (GWAS) and imputation analysis. Methods: There were 472 HCC cases and 806 unaffected controls. All study subjects were adults seropositive for antibodies against HCV and seronegative for HBsAg. The demographic characteristics and serum markers for liver functions were also evaluated. High quality human genomic DNA was extracted from each blood sample to perform genotyping by AxiomTM Genome-Wide CHB Array. Then, imputation algorithm was applied to get whole genome variations in patients based on genotyping data from SNP microarrays. The reference genomes of imputation were Han Chinese population in 1000 genome project. The logistic regression was used to evaluate association between disease and genotype based on four different genetic models (allelic, dominant, additive and recessive). P values $<10^{-5}$ were considered significant. Results: A total of 36,175,343 SNPs were obtained in the samples after imputation. There were 765, 134, 612, and 855 SNPs significantly associated with HCC based on additive, recessive, dominant, and allelic genetic models, respectively. Among the SNPs discovered in our analysis, less than 4% could be detected by microarray chips. In addition, 18 nonsynonymous SNPs were found to be significantly associated with HCC under different genetic models. Only 3 out of 18 nonsynonymous SNPs could be detected in microarray chips. Interestingly, 7 nonsynonymous SNPs clustered on the human leukocyte antigen (HLA) complex region. It indicated that HLA region play a important role in host susceptibility to hepatocellular carcinoma within chronic hepatitis C patients. Conclusion: The SNPs associated with HCV-related HCC were identified in this study. It provide insight for identification of high risk population in HCV after clinical validation.

3286T

Functional characterization of *PARP1* melanoma-associated locus. *J. Choi[1], M. Makowski[1], M. Xu[1], T. Zhang[1], M. Law[2], W. Kim[1], M. Kovacs[1], H. Parikh[1], L. Aoude[2], M. Gartside[2], H. Yin[3], J. Trent[3], S. Macgregor[2], N. Hayward[2], K. Brown[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) Queensland Institute of Medical Research, Brisbane, QLD, Australia; 3) Translational Genomics Research Institute, Phoenix, AZ.

Recent genome wide association studies (GWAS) identified several new loci for melanoma susceptibility. While these results highlight potential pathways predisposing to melanoma, functional risk variants in these regions as well as the mechanism by which they influence risk have yet to be elucidated. To nominate functional variants we performed expression quantitative trait loci (eQTL) analysis in 62 melanoma cell lines. Transcript levels were measured using Affymetrix U133Plus2 expression microarray. SNPs (>700K) were then typed on Illumina OmniExpress arrays and the ones in GWAS loci were further imputed using 1000 genomes (1KG) data. Among 16 GWAS loci tested two chromosome 1 loci exhibited significant *cis*-eQTL. Subsequent validation using Taqman quantitative PCR (qPCR) demonstrated that Poly [ADP-ribose] polymerase 1 (*PARP1*) expression is significantly associated with the lead SNP (p=0.03, genomic copy number adjusted). Namely, the risk allele is correlated with an increased *PARP1* transcript levels after adjusting for genomic copy number (p=0.03). We then further interrogated the genotype-expression correlation by Taqman allele discrimination qPCR in 21 melanoma cell lines heterozygous for the GWAS lead SNP. The results demonstrated significantly higher proportion for the risk allele in *PARP1* transcripts (p=0.0001). To identify functional risk variants mediating these effects we annotated the *PARP1* locus using ENCODE database. Among 56 SNPs of strong linkage disequilibrium (LD) with the lead SNP (r^2>0.5 using 1KG genotypes), six exhibited strong evidence as potential transcriptional enhancers in melanoma relevant cell types. One of them is a six-base pair indel in GC-rich region poorly covered by 1KG. Genotype reassessment of this indel using gel-based fragment analysis in 745 healthy Europeans from DCEG imputation reference set resulted in remarkably enhanced LD with the lead SNP (r^2=0.94 from 0.67) supporting direct link to melanoma susceptibility. Subsequent Electro Mobility Shift Assays and luciferase assays for this indel demonstrated allele-specific protein binding and differential transcriptional activities in melanoma cell lines. Chromatin immunoprecipitation indicated that enhancer-binding proteins are enriched in melanoma-associated deletion allele in melanoma cell lines. Identification of proteins directly mediating this function will further elucidate PARP1 contribution to melanoma susceptibility.

3287S

Heritable missense variant rs3731249 underlies the *CDKN2A* association with childhood ALL and is preferentially retained by tumors harboring somatic *CDKN2A* loss. *A.J. de Smith[1], K.M. Walsh[2], H.M. Hansen[2], L.F. Barcellos[3], A.P. Chokkalingam[3], R.B. Jenkins[4], M.R. Wrensch[2], J.K. Wiencke[2], C. Metayer[3], J.L. Wiemels[1].* 1) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; 2) Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA; 3) Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA; 4) Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Rochester, MN.

Little is known about the interaction between constitutive and somatic genetic variation in carcinogenesis. Genome-wide association studies (GWAS) of childhood acute lymphoblastic leukemia (ALL [MIM 613065]) have identified 6 genetic associations, including rs3731217, located between the *CDKN2A* [MIM 600160] and *CDKN2B* [MIM 600431] genes. Somatic deletions of this 9p21.3 region are common in cancer, and occur in ~30% of childhood ALL. We replicated the rs3731217 association with childhood ALL in Hispanics (p=0.021) via a GWAS of 323 childhood ALL cases and 454 controls, using the Illumina OmniExpress SNP array. SNP imputation analysis across 9p21.3 revealed a stronger association at rs3731249 ($p=6.4\times10^{-4}$; OR = 2.77), a low-frequency missense variant in exon 2 (p.Ala148Thr) of *CDKN2A* ($MAF_{controls}$=2.1%). These associations were replicated in an additional 378 ALL cases and 536 controls using Taqman assays, where the missense variant again had a greater magnitude of effect ($P_{rs3731249}=6.5\times10^{-3}$; $OR_{rs3731249}$ = 2.00 versus $P_{rs3731217}$=0.029; $OR_{rs3731217}$ = 1.38). We assessed the effect of rs3731249 on risk of other cancers with known chr9p21.3 GWAS associations, but observed no association between this variant and either adult glioblastoma or melanoma risk, suggesting an ALL-specific role for rs3731249. We hypothesized that cases heterozygous for rs3731249 may show allelic imbalance, with preferential loss of the wildtype (WT) allele when hemizygous chr9p21.3 deletions occur in leukemic cells. To test this, we developed a novel method termed "SMART-ddPCR" (Somatic Mutation Allelic Ratio Test using Droplet Digital PCR). Concentration of risk and WT alleles was measured by ddPCR using the rs3731249 Taqman assay in diagnostic bone marrow (*i.e.* tumor) DNA from 35 known heterozygotes. Allelic imbalance due to somatic *CDKN2A* alterations was identified in 17 cases, of which 14 had higher risk:WT allele ratio (p=0.006, 1-sided binomial significance test), thus demonstrating preferential loss of WT rs3731249 and retention of the missense allele. Subclonal heterogeneity in *CDKN2A* copy number was also evident, suggesting that this is a late event in leukemogenesis. Preferential allelic imbalance towards the rs3731249 risk allele in ALL tumor DNA provides strong evidence of an important role for this missense variant in leukemogenesis and shows, for the first time, a direct relationship between the constitutive and somatic genetic variants underlying ALL development.

3288M

Leveraging Sequence and Phenotype-Specific Information to Design Custom Genotyping Arrays: Example from Prostate Cancer. *N. Emami[1], T. Hoffman[1], J. Mefford[1], K. Lindquist[1], C. Cario[1], E. Wan[1], S. Wong[1], J. Gollub[2], A. Finn[2], P. Kwok[1], N. Risch[1,3], S. Van Den Eeden[3], J. Witte[1].* 1) University of California, San Francisco, San Francisco, CA; 2) Affymetrix, Inc., Santa Clara, CA; 3) Kaiser Permanente Northern California Division of Research, Oakland, CA.

Association studies are increasingly focused on the contribution of rare genetic variants to phenotypic variation. Detecting rare variants via sequencing large numbers of samples remains expensive. Moreover, there is a limit to the statistical power for detecting associations with extremely rare variants. These issues have resulted in the continued used of DNA microarrays. While various array designs are commercially available off-the-shelf, these may not adequately reflect rare variation in the target population and may misallocate space to uninformative, monomorphic markers. Instead, recent large-scale studies of human genetic variation, as well as sequencing studies for disease, have provided a repository of information upon which custom array solutions can be tailored to particular phenotypes. We illustrate the potential of such custom arrays for studying rare- and putative trait-specific variation in prostate cancer. First, we sequenced tumor and normal DNA from a limited number of African-American men with prostate cancer, and obtained prostate cancer germline exomes from the Cancer Genome Atlas and dbGaP; these data were then filtered to select rare variants for inclusion on the array. Second, we included the following additional custom content in our target set: pleiotropic loci previously associated with multiple cancers, loci with measured DNA hypersensitivity in prostate cancer cell lines, and regions of dense genotyping centered at loci previously associated with prostate cancer, with the purpose of discovering potentially causal variation. Third, we selected sets of target markers recently distinguished as assaying informative rare and putatively functional variation. Finally, the custom array was designed to complement existing GWAS data on our study samples by using a greedy SNP-selection algorithm to minimize redundancy of typed markers while optimizing coverage of our target set. The final array includes 415,664 assayed markers, and we have typed it on a population of 8,000 prostate cancer cases and 8,000 controls nested within the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH). In preliminary results, we replicated the strong association between the rare HOXB13 G84E mutation and prostate cancer (p=0.002, OR=2.74, 95% CI 1.45-5.22). Our approach and results provide a framework for designing highly informative custom arrays to study rare and trait-specific genetic variation and complement existing GWAS data.

3289T

Genome-wide scan identifies variants in 2q12.3 and 9q22.33 associated with risk of multiple myeloma. *S.W. Erickson[1], V.R. Raj[1], O.W. Stephens[1], I. Dhakal[1], S.S. Chavan[1], N. Sanathkumar[1], E.A. Coleman[1], J.Y. Lee[1], J.A. Goodwin[1], S. Apewokin[1], D. Zhou[1], C.J. Heuck[1], A.J. Vangsted[2].* 1) UAMS, Little Rock, AR; 2) Roskilde Hospital, Copenhagen, DK.

Despite recent studies which have identified multiple susceptibility loci for multiple myeloma (MM), the genetic contributions to risk of the disease are only partially understood. In the current study, we performed a genome-wide association study (GWAS) of MM risk on 972 patients who underwent melphalan-based high-dose treatment supported by autologous stem cell transplant (HD-ASCT), using genotype data from 1064 cancer-free adults as controls. We identified two novel susceptibility loci, one in 2q12.3 upstream of the gene ST6GAL2, and one in 9q22.33 between the genes FOXE1 and XPA. Risk genotypes in 9q22.33 were also associated with lower expression of FOXE1. All three of these genes have been previously associated with risk of cancer or apoptosis. Our study also replicated six out of seven previously published germline associations to MM risk.

3290S

Genetic determinants of Breslow tumor thickness and their impact on melanoma progression. *S. Fang[1], Y. Wang[1], D. Deng[1], Q. Liu[2], R. Feng[1], K. Xu[1], H. Liu[1], M. Ross[1], J. Gershenwald[1], J. Cormier[1], R. Royal[1], A. Lucci[1], J. Wargo[1], C. Schacherer[1], J. Reveille[3], D. Sui[1], R. Bassett Jr[1], F. Demenais[4], L. Wang[1], M. Lu[1], Q. Wei[5], C. Amos[6], J. Lee[1], GenoMEL group.* 1) University of Texas MD Anderson Cancer Center, Houston, TX; 2) The fifth affiliated Hospital of ZhengZhou University, Zhengzhou, Henan; 3) The University of Texas Health Science Center at Houston, Houston, TX; 4) Université Paris Diderot Paris; 5) Duke University School of Medicine, Durham, NC; 6) Dartmouth College, Lebanon, NH.

Breslow thickness of the primary melanoma tumor is one of the most important clinicopathologic factors predicting risk of melanoma progression. Genetic mechanisms determining Breslow thickness are not yet known. We performed a genome-association study of Breslow thickness using 1547 patients in MD Anderson (MDA) in discovery analysis and further evaluated their association using GenoMEL (UK and European) and French melanoma datasets. None of the SNPs identified in 4 candidate gene regions (MUC2/MUC5AC/CALR/RAD23A) in the MDA dataset was replicated in the GenoMEL or French datasets. However, gene-based VEGAS testing confirmed that two MDA candidate genes, MUC2 and MUC5AC, were significantly related to tumor thickness in the MDA dataset (both P-values<2.8×10-6). Additionally, both CALR and RAD23A genes were associated with melanoma disease-free and overall survival (P-value<0.05) in the MDA dataset. Finally, immunohistochemistry of primary melanoma tumors demonstrated that both MUC2 and CALR expression levels increased with tumor thickness; expression of MUC2 was also elevated in regional nodal metastases from melanoma patients. These data suggest that MUC2 and CALR genes may determine melanoma tumor thickness and influence melanoma disease severity.

3291M

Association of *2R3R* polymorphism of the Thymidylate synthase gene with toxicity in breast cancer patients treated with FEC chemotherapy. *M.P. Gallegos-Arreola[1 y 2], R. Ramírez[1 y 3], IA. Gutíerrez[1 y 3], OM. Soto[1 y 2], DI. Carrillo[1 y 3], AM. Puebla[6], GM. Zúñiga[4], LE. Figuera[5].* 1) Lab. de Genética Molécular. Div Med Mol, CIBO, IMSS, Guadalajara, Jal,. Mex; 2) Doctorado en Farmacología, CUCS, U de G; 3) Doctorado en Genética Humana, CUCS, U de G,; 4) Lab. Mutagénesis, CIBO, IMSS; 5) División de Genética, CIBO, IMSS; 6) Laboratorio de Inmunofarmacología. Departamento de farmacología, CUCEI, U de G.

Background: The influence of polymorphisms in the TYMS gene involved in metabolism of chemotherapeutic agents has been studied in different cancers. Thymidylate synthase (TS) plays a major role in folate metabolism and consequently could be an important factor for the efficacy of a treatment with FEC (fluorouracil-epirubicin-cyclophosphamide) in breast cancer patients. Our aim was to evaluate the association of 2R3R polymorphism with toxicity effects in breast cancer patient's neoadjuvantly treated with FEC chemotherapy. Methods: DNA genomic samples from 525 patients (UMAE gynecology and obstetrician Hospital, CMNO, IMSS), that received FEC neoadjuvant chemotherapy; were included in the study. Protocol was support by FIS/IMSS/PROT/G13/1231. The 2R2R polymorphism was determinate by polyacrylamide gels electrophoresis, previously PCR amplification analysis. The association was determinate by odds ratio. Results: The genotype heterozygous (2R3R) was associated with hepatotoxicity (high levels of GGT enzyme) in breast cancer patient with response at FEC chemotherapy [2.7(IC95% 1.07-3.4), p=0.026], as well as with hematological toxicity (neutropenia) [2.7(IC95% 1.3-5.5), p=0.024] and mucositis toxicity [2.9(IC95% 1.03-8.2), p=0.027] in breast cancer patients non responder at FEC chemotherapy. Conclusion: The polymorphism 2R3R could be a good marker of toxicity in breast cancer patients treated with FEC chemotherapy in the analyzed sample.

3292T

Regulatory polymorphisms in lymphoma and chronic lymphocytic leukemia risk. *J. Hayes[1,2], G. Trynka[3,4], V. Joseph[1], K. Offit[1], S. Raychaudhuri[3,4,5,6], R. Klein[7].* 1) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 2) Biochemistry, Cellular, and Molecular Biology Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA; 3) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 4) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 5) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 6) NIHR Manchester Musculoskeletal Biomedical, Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK; 7) Icahn Institute for Genomics and Multiscale Biology and Department of Genetics and Genomic Sciences, Icahan School of Medicine at Mt. Sinai, New York NY USA.

The first genome-wide association study (GWAS) was published in 2005. Since then, there have been over 1,300 published GWAS papers spanning over 200 diseases and more than 3,000 disease susceptibility loci identified. However, one of the major challenges of the field is that over 90% of these identified risk variants do not result in an amino acid change, thus failing to explain the mechanism for increased risk. We hypothesize that the GWAS-identified risk-SNPs, or those variants highly correlated with them, increase risk by altering transcriptional regulatory elements.

To test this hypothesis, we generated a flexible and customizable computational pipeline to ask if a set of SNPs associated with a phenotype are enriched in a particular functional annotation of the genome. To demonstrate the power of this approach, we tested risk-SNPs for both lymphoma and chronic lymphocytic leukemia (CLL) for enrichment in regions of the genome reported to be biochemically functional by the ENCODE project. Using this pipeline we were able to identify numerous SNPs that fell into putative regulatory regions by several measures, including DNase-hypersensitivity regions, ChIP-Seq peaks, and integrative analysis of histone methylation data.

The computational pipeline utilizes a Monte Carlo approach to test for enrichment, taking linkage disequilibrium patterns into account when compared to feature-matched random SNP sets. We observed a statistically significant enrichment of the lymphoma and CLL risk SNPs in a lymphoid-lineage specific manner for numerous regulatory marks, including DNase peaks and segmentation defined “strong enhancers” at $p<0.001$. However, even though the LCLs tested are of the same tissue of origin, we were able to detect inter-individual variability among the cell lines. These results suggest that regulatory variants can explain many of the known high-frequency genetic risk factors for lymphoma and chronic lymphocytic leukemia.

3293S

Role of polymorphic fibroblast growth factor receptor (FGFR) Gene and Breast Cancer Risk. *M. Hosseini[1], M. Houshmand[2].* 1) Biological Science, Islamic Azad Univ, Islamshahr Branch, Islamshahr, Tehran, Iran; 2) National Institute for Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran.

Background: Genetic factors related to cancer have been extensively studied and several polymorphisms have been associated to breast cancer. Breast cancer (BC) is one of the most common causes of death among women, and second in Iran. The objectives of this study were to determine the frequency of the fibroblast growth factor receptor (FGFR) Gene polymorphism in patients with breast cancer. Methods: For the first time, we evaluated these polymorphisms and effects on the breast cancer risk association in an Iranian sporadic population-based case-control study of 126 breast cancer cases and 160 controls using a PCR-RFLP-based assay. Results: Analyses of affected and controls show that homozygote genotype FGFR4 Gly/Gly has the highest frequency in patients and control groups (30.4 and 18.9%). Genotype FGFR4 Gly/Gly most risk factor were in our population: ArgGly /GlyGly, OR= 2.359, 95% CI= 0.208 - 4.621, p=0.001; ArgArg / ArgGly, OR=0.412, 95% CL=0.082 - 0.547, p=0.078 , ArgArg /GlyGly, OR= 0.076, 95% CI=0.030 - 0.189, p=0.26. Conclusions: There was a significant association of breast cancer risk with FGFR4 GlyGly and ArgGly polymorphism. Keywords: FGFR4 gene, polymorphism, breast cancer, BstNI , PCR-RFLP.

3294M

GWAS meta-analysis identifies three novel risk loci for melanoma at 6p22, 7q21 and 9q31. *M.H. Law[1], D.T. Bishop[2], C.I. Amos[3], M. Brossard[4], S.V. Ward[5], N.K. Hayward[6], N.G. Martin[7], R. Kumar[8], A.E. Cust[9], G. Radford-Smith[10,11], G.W. Montgomery[12], D.C. Whiteman[13], J.A. Newton-Bishop[2], D.L. Duffy[7], J.H. Barrett[2], J.E. Craig[22], J. Han[14,15], J.E. Lee[16], A. Qureshi[17], A.M. Dunning[18], F. Song[19], G.J. Mann[20], P.D.P. Pharoah[18], E.K. Moses[5], F. Demenais[4], Q. Wei[21], S. Macgregor[1], M.M. Iles[2], Contributing consortia: GenoMEL, QMEGA/QTWIN, MELARISK, AMFS, IBD and SDH.* 1) Statistical Genetics Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia; 2) Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, Leeds Cancer Research UK Centre, University of Leeds, Leeds, UK; 3) Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College; 4) INSERM, UMR-946, Genetic Variation and Human Diseases Unit, Paris, France; Université Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie, Paris, France; 5) Centre for Genetic Origins of Health and Disease, Faculty of Medicine, Dentistry and Health Sciences, The University of Western Australia, Crawley, WA 6009, Australia; 6) Oncogenomics, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia; 7) Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia; 8) Division of Molecular Genetic Epidemiology, German Cancer Research Center, 69120 Heidelberg, Germany; 9) Cancer Epidemiology and Services Research, Sydney School of Public Health, The University of Sydney, Australia; 10) Inflammatory Bowel Disease, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia; 11) University of Queensland School of Medicine, Herston Campus, Brisbane, Australia; 12) Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia; 13) Cancer Control Group, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia; 14) Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN, USA; 15) Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, IN, USA; 16) Department of Surgery, The University of Texas MD Anderson Cancer Center, Houston, Texas; 17) Department of Dermatology Warren Alpert Medical School, Brown University; 18) Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK; 19) Departments of Epidemiology and Biostatistics, Key Laboratory of Cancer Prevention and Therapy, Tianjin, National Clinical Research Center of Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin, P. R. China; 20) Westmead Institute of Cancer Research, University of Sydney at Westmead Millennium Institute and Melanoma Institute Australia, Sydney, Australia; 21) Duke Cancer Institute, Duke University Medical Center, Durham, NC, USA; 22) Department of Ophthalmology, Flinders University, Adelaide, SA 5042, Australia.

Melanoma is the deadliest form of skin cancer, representing a significant burden in terms of mortality and healthcare costs. Prior to the advent of genome-wide association studies (GWASs), our understanding of the melanoma genetics was limited to rare familial mutations in cell cycle check point genes and single nucleotide polymorphisms (SNPs) that impacted pigmentation. Our first wave of melanoma GWASs in 2009 confirmed the importance of pigmentation genes in melanoma, and our second wave in 2011 identified additional pathways including DNA repair genes, greatly increasing our understanding of melanoma aetiology.

We build on these findings through meta-analysis with genomic imputation of 11 melanoma GWASs totalling 12,814 cases and 23,204 controls, representing the largest genetic study of melanoma to date. In addition to 14 established melanoma risk loci reaching genome-wide significance, we have confirmed that genetic variants at 15q13.1 near the eye colour gene *oculocutaneous albinism 2* (*OCA2*) are associated with melanoma (rs4778138 $p = 3.95 \times 10^{-9}$, $I^2=8.33$).

Further, we have identified three novel genome-wide significant loci ($p<5 \times 10^{-8}$) with homogenous effects across studies ($I^2=0$). First, rs6914598 ($p=3.86 \times 10^{-8}$) at 6p22.3 lies in the *CDK5 regulatory subunit associated protein 1-like 1* gene (*CDKAL1*). CDKAL1 methylthiolates tRNAs, modulating expression of a range of genes including proinsulin. Additional SNPs in *CDKAL1* have been associated with numerous diseases including type-2 diabetes and bladder cancer; these SNPs are not obviously associated with melanoma (rs9348440 $p=0.024$ with melanoma, all others $p>0.2$), indicating that the rs6914598 signal is independent. Second, rs1636744 ($p=2.79 \times 10^{-9}$) in 7p21.1 is near the *anterior gradient 3 homolog* gene (*AGR3*). This SNP is reported as an eQTL for *AGR3* in lung and thyroid tissue, and *AGR3* expression has been associated with breast cancer risk and poor survival in ovarian cancer. rs1636744 is not in a conserved region, nor is there reported evidence of functional regulation at rs1636744 or SNPs in strong LD. Third, rs10739221 ($p=1.38 \times 10^{-8}$) in 9q31.2 is intergenic, with nearby genes including *transmembrane protein 38B* (*TMEM38B*) and the nuclear excision repair gene, *RAD23 homolog B* (*RAD23B*). The success of this meta-analysis indicates the power of combining GWASs to further uncover the biological architecture of melanoma susceptibility.

3295T

Detection of *trans* and *cis* splicing QTLs through large scale cancer genome analysis. *K. Lehmann[1], A. Káhles[1], C. Kandoth[1], W. Lee[1], N. Schultz[1], O. Stegle[2], G. Rätsch[1].* 1) Memorial Sloan-Kettering Cancer Center, Computational Biology, New York, NY; 2) European Bioinformatics Institute; Hinxton; Cambridge; CB10 1SD; United Kingdom.

The comprehensive survey of molecular characteristics provided by The Cancer Genome Atlas (TCGA) enables large scale analyses across multiple cancers. However sophisticated tools for the joint analysis of the thousands of samples and cancer specific challenges are needed. In an effort to enable joint analysis, we have re-aligned and re-analyzed RNA and whole exome sequencing data of ~4.000 individuals across 12 cancer types in a uniform manner. We used a newly developed open source SplAdder pipeline to count gene expression as well as annotate and quantify a comprehensive set of alternative splicing events. We identified threefold more high confidence alternative splicing events than annotated in the GENCODE annotation which reflect cancer-specific and tissue-specific splicing variation. Comparisons to matching tissue normal samples confirm a ~20% increase of splicing complexity in tumor samples. We have identified 22 genes with splicing changes that recurrently occur in tumor samples (>10%) but are virtually never observed in normal samples or ENCODE cell lines (<0.5%) and could be possible targets for new drugs. While population structure is one of the most severe confounding factors in QTL analysis, tumor samples open up many new additional challenges. Tumor specific somatic mutations and recurrence patterns as well as sample heterogeneity can lead to spurious associations. Thus, we have developed a new strategy to perform a common variant association study using mixed models on tumor samples enabling us to account for tumor specific genotypic and phenotypic heterogeneity in addition to population structure. Due to sample size constraints, many previous QTL studies have been limited to the analysis of *cis*-associated variants. The large sample size available from TCGA enables us to overcome this limitation and discover *trans*-associated variants as well. Preliminary data demonstrates that we find *cis*-associations for ~10% of the analyzed genes, of which a large fraction replicates across tissue and cancer types. We also confirm a recently reported *trans*-association in the splice factor U2AF1 and detect several additional *trans*-associations with effect sizes >20%. Current work includes the follow-up analysis on these discoveries. To address the commonly found rare somatic variants in cancer, we have also designed a rare variant test for splicing QTLs in order to identify recurrently affected genes and networks leading to aberrant splicing.

3296S

Genome-wide association study of breast cancer in Japanese population. *S. Low[1], A. Takahashi[1], M. Kubo[2], T. Katagiri[3].* 1) Lab. for Statistical Analysis, Ctr. for I.M.S., RIKEN., Yokohama, Kanagawa., Japan; 2) Lab. for Genotyping Development, Ctr. for I.M.S., RIKEN, Yokohama, Kanagawa, Japan; 3) Division of Genome Medicine, Inst. for Genome Res., The Univ. of Tokushima, Tokushima, Japan.

Breast cancer is the most common malignancy among women in worldwide including Japan. Many genome-wide association studies have been performed to identify genetic variants that are associated with the risk of breast cancer. Owing to the complex linkage disequilibrium structure and various environmental exposures among different populations, it is of important to investigate associated genetic variants with breast cancer in a specific population. In this study, we conducted a genome-wide association study as well as whole-genome imputation with 2,642 cases and 2,099 unaffected female controls collected from the Biobank Japan. We further examined 13 suggestive loci ($P<1.0\times10^{-5}$) using an independent sample set of 2,885 cases and 3,395 controls and successfully validated two previously-reported loci, rs2981578 (combined *P*-value of 1.31×10^{-12}, OR=1.23; 95% CI=1.16-1.30) on chromosome 10q26 (*FGFR2*), rs3803662 (combined *P*-value of 2.79×10^{-11}, OR=1.21; 95% CI=1.15-1.28) and rs12922061 (combined *P*-value of 3.97×10^{-10}, OR=1.23; 95% CI=1.15-1.31) on chromosome 16q12 (*TOX3-LOC643714*). Weighted genetic risk score by using three significantly associated variants and two previously reported breast cancer associated loci in East Asian population revealed that individuals who carry the most risk alleles have 2.2 times higher risk of developing breast cancer in the Japanese population compare to those who carry the least risk alleles. In addition, pathway analysis suggested that variants within aryl hydrocarbon receptor signaling pathway from Ingenuity database are associated with the risk of breast cancer. Although we could not identify additional loci associated with breast cancer, our study utilized one of the largest sample sizes reported to date, and provided genetic status that represent the Japanese population. Further local and international collaborative study is essential to identify additional genetic variants that could lead to a better, accurate prediction for breast cancer.

3297M

New insights into ovarian cancer from the investigation of overall genetic sharing. *Y. Lu[1], A.B. Spurdle[2], G.W. Montgomery[3], K.T. Zondervan[4], P.D. Pharoah[5,6], G. Chenevix-Trench[7], S. Macgregor[1], Ovarian Cancer Association Consortium.* 1) Statistical Genetics, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia; 2) Molecular Cancer Epidemiology, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia; 3) Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia; 4) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 5) Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK; 6) Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK; 7) Cancer Genetics, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia.

Epithelial ovarian cancer (EOC) is not a single disease but is comprised of many tumour types, with substantial heterogeneity between the subtypes. To date, most genetic studies have focused on the common subtype, serous carcinoma. Studying rare histological subtypes of EOC (clear-cell, endometrioid, or mucinous) has been challenging because a large sample size is often required but not readily available. Epidemiological studies observed that certain subtypes of EOC are associated with other women's diseases, e.g. clear cell and endometrioid carcinoma occur more frequently in women with endometriosis. But it remains to be elucidated whether there is a genetic overlap underlying the association observed from epidemiological studies. We investigated this question in the largest ovarian cancer data set to-date from Ovarian Cancer Association Consortium, explicitly modelling the "polygenic" effect shared between EOC subtypes and other women's diseases using polygenic risk prediction approach and bivariate linear mixed models implemented in GCTA. We found a highly significant genetic overlap between ovarian and endometrial cancer, endometrial cancer and endometriosis, and a borderline significant overlap between breast and endometrial cancer. These findings are consistent with previous work on the familial aggregation of women's cancers, serving as a proof of principle of our method. We also found significant overlaps between clear cell, endometriod EOC with endometriosis, which strongly support the hypothesis that endometriosis is a probable precursor of clear cell and endometriod EOC. In addition, there were also suggestive genetic overlaps between multiple EOC subtypes and endometrial cancer. Altogether, these intriguing genetic overlaps suggest that treatment for other women's diseases may be used as preventive intervention for certain EOC subtypes.

3298T

Exploring the Role of Regulatory Variation in Linkage Disequilibrium with Cancer Risk SNPs. *D.S. Park, N. Zaitlen, J. Witte.* Bioengineering & Therapeutic Sciences, UC San Francisco, San Francisco, CA.

Genome Wide Association Studies (GWAS) have identified large numbers of single nucleotide polymorphisms (SNPs) that reproducibly influence disease risk (Manolio et al. 2009). Such SNPs are more likely to be expression quantitative trait loci (eQTL) than random SNPs (Nicolae et al. 2010), suggesting that they may exert their phenotypic effect by altering gene regulation rather than modifying the product of the implicated gene.

To investigate this phenomenon, we examined whether there is an over-representation of regulatory variants in linkage disequilibrium (LD) with cancer risk SNPs from all GWAS reported in the NHGRI catalogue, and whether these regulatory SNPs display 'variant pleiotropy' (i.e. when multiple variants affect the same gene/pathway). To test for enrichment at the SNP-level we first created random sets of SNPs by conditioning on minor allele frequency (MAF), genotyping array, function, and distance to transcription start site of cancer-associated SNPs. All proxy SNPs in LD at a threshold of $r^2 \geq 0.8$ and max-distance of 500kb were generated for the random sets, and RegulomeDB (Boyle et al. 2012) scoring categories were assigned to the proxy SNPs. The number of proxy SNPs in each scoring category was used to create an empirical distribution of functional SNPs in LD with the random sets to calculate the enrichment of the cancer SNPs for each scoring category. The cancer SNPs were pruned ($r^2 > 0.2$) while the random sets were not in order to obtain a conservative estimate of enrichment. To determine if functional variants affect similar genes/pathways we used DAVID (Huang et al. 2007) and GREAT (McLean et al. 2010).

We tested for functional variant enrichment and 'variant pleiotropy' in the HapMap populations CEU and CHB/JPT. At the SNP-level, in CEU we saw an enrichment of functional regulatory variants across all cancers as well as in bladder and prostate cancer. On the other hand, CHB/JPT was not enriched across all cancers but was in hepatocellular carcinoma, lung, and prostate cancer. Genes that were regulated by SNPs in the cancers identified were enriched for cancer related gene families and pathways such as the MAPK pathway, TGF-Beta pathway, Basic-Helix-Loop-Helix genes, and Zinc Finger genes. Our findings support the hypothesis that regulatory variants play a role in disrupting the normal function of cancer associated genes and pathways.

3299S

Association of *2R3R* polymorphism of the Thymidylate synthase gene with in breast cancer advanced stage patients. *A.M. Puebla-Pérez[1], A. Ramos[2 y 3], R. Ramírez[2 y 4], IA. Gutíerrez[2 y 4], OM. Soto[2 y 3], DI. Carrillo[2 y 4], LE. Figuera[5], MP. Gallegos[2].* 1) Laboartorio de Farmacología. Departamento de Farmacología, Universidad de Guadalajara, Guadalajara, Jalisco. México; 2) Lab. de Genética Molecular. Div de Med Mol, CIBO, IMSS; 3) Doctorado en farmacología, CUCS, U de G; 4) Doctorado en Genética Humana, CUCS, U de G; 5) División de Genética, CIBO, IMSS.

Background: The 2R3R polymorphisms in the TYMS gene have been studied in different cancers however exist contradicted results to respect at their association with breast cancer. Our aim was to evaluate the association of 2R3R polymorphism with advanced stage in breast cancer patients. Methods: DNA genomic samples from 525 patients (UMAE gynecology and obstetrician Hospital, CMNO, IMSS), were included in the study. Protocol was support by FIS/IMSS/PROT/G13/1231. The 2R2R polymorphism was determinate by polyacrylamide gels electrophoresis, previously PCR amplification analysis. The association was determinate by odds ratio. Results: The genotype 3R3R was associated with presence of metastatic nodules [1.9(IC95% 1.03-3.7), p=0.040], as well as the union of genotypes 3R3R-2R3R were associated advanced stages (III-IV) in breast cancer patients [2.2(IC95% 1.1-4.1), p=0.015]. Conclusion: The polymorphism 2R3R could be a good marker in advanced clinical stage in breast cancer patients, in the analyzed sample.

3300M

Differences of nitric oxide level in Mexican breast cancer patients. *R. Ramírez-Patiño[1,2], Jl. Delgado[3], A. Ramos[1,4], AM. Puebla[3], LE. Figuera[5], MP. Gallegos[1].* 1) Lab de Genética Molécular. Div Med Mol, CIBO, IMSS, Guadalajara, Jal,. Mex, Mexico; 2) Doctorado en Genética Humana, CUCS, u de G; 3) Laboratorio de Inmunofarmacología, Departamento de Farmacología, CUCEI, U de G; 4) Doctorado en Farmacología, CUCS, U de G; 5) División de Genética, CIBO, IMSS.

Nitric Oxide (NO) level plays an important role in several biological functions. Recent evidence suggests that NO may also increase metastatic ability in human cancers. We examined the Nitrite/nitrate levels in 234 breast cancer and 84 healthy Mexican women. The NO levels in plasma observed were 648.67±167.19 nM, with significant differences. The NO level is associated with breast cancer in the analyzed samples from the Mexican population.

3301T

Admixture scan of breast cancer in U.S. black women: the AMBER consortium. *E.A. Ruiz-Narvaez[1], L.E. Sucheston-Campbell[2], J.T. Bensen[3], S. Yao[2], S.A. Haddad[1], A.F. Olshan[3], C.B. Ambrosone[2], J.R. Palmer[1], K.L. Lunetta[4].* 1) Slone Epidemiology Center, Boston University, Boston, MA; 2) Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY; 3) Department of Epidemiology, Gillings School of Global Pubic Health, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

Breast cancer incidence varies by racial groups in the United States. Estrogen receptor (ER) positive breast cancer is more frequent in European and Asian ancestry populations, and ER negative breast cancer is more common in women of African ancestry relative to other populations. We conducted a genome-wide admixture scan in order to identify loci that may contribute to the racial differences in incidence of breast cancer and the specific ER subtypes. We analyzed 2,624 autosomal ancestral informative markers (AIMs) in 3,130 breast cancer cases (including 1,622 ER-positive, and 952 ER-negative) and 3,698 controls from the African American Breast Cancer Epidemiology and Risk (AMBER) consortium, a collaborative study from four of the largest studies of breast cancer in African American women (Carolina Breast Cancer Study, Black Women's Health Study, Women's Circle of Health Study, and Multi-Ethnic Cohort). We used ADMIXMAP software to conduct the analyses. Mean European ancestry percentage was 17.6% for controls, 17.7% for all breast cancer cases, 17.5% for ER-positive cases, and 16.7% for ER-negative cases. In case-only analyses, lower European ancestry was observed over a region on chromosome 12q24 for all cases (Z-score = -3.97, $p = 7.2\times10^{-5}$), and for ER-positive cases (Z-score = -4.04, $p = 5.3\times10^{-5}$). In case-only analyses of ER-negative women, the largest excess of European ancestry was observed over the 8q21 region (Z-score = 3.59, $p = 3.4\times10^{-4}$). In summary, we found that African American women with ER-negative tumors tended to have lower overall European ancestry relative to women with ER-positive breast cancer. These results add to the findings from the GWAS in Caucasian populations and suggest novel regions containing multiple variants with weak to modest effects that may jointly contribute to the racial differences in breast cancer risk.

3302S

Estimation of de novo mutation rates in the offspring of Lynch syndrome families. *S. Shankaracharya, C.D. Huff, F.A. San Lucas, C. Wei, J. Rother, P.A. Scheet, P.M. Lynch, M.L. Frazier.* Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX.

Lynch syndrome [LS] is a hereditary disorder characterized by an increased risk of developing colorectal cancer and other specific cancers. LS follows an autosomal dominant pattern of inheritance and is caused by genetic variants that decrease DNA mismatch repair capacity. Individuals with LS have a 50-80% lifetime risk of developing colorectal cancer, with an average age-of-onset of 45 years. In LS families, offspring develop cancer approximately 10 years earlier than their parents, but the cause of anticipation in LS is not yet understood. Here, our goal is to determine whether anticipation in LS results from an excess of single nucleotide and small indel de novo mutations due to decreased DNA mismatch repair capacity. To test this hypothesis, we sequenced 39 individuals from 3 control and 7 LS families to greater than 30X coverage on two separate whole genome sequencing (WGS) platforms, Illumina and Complete Genomics. To identify single nucleotide and small indel de novo mutation candidates, we analyzed the Illumina WGS data with four pedigree-aware variant callers (GATK PhaseByTransmission, Denovogear, Famseq and Polymutt) as well as two general-purpose variant callers (GATK HaplotypeCaller and GATK UnifiedGenotyper). We report the estimated single nucleotide and small indel de novo mutation rates in each family for candidates identified from the Illumina data and validated by the Complete Genomics data. With this approach, the expected number of false positive mutations is less than 1 per family. We also report the accuracy of the six variant callers as measured by the specificity, sensitivity and receiver operating curves (ROC) of each method.

3303M

Excess Prevalence of Gastric Cancer Family History Among Hispanic Breast Cancer Patients. *I. Solomon, A. Sunga, K. Yang, J. Weitzel.* City of Hope National Medical Center, Duarte, CA.

Purpose: As the 2nd most common cause of cancer death and 4th most common incident cancer, gastric cancer (GC) accounts for approximately 10% of all cancer deaths. Environmental and genetic risk factors have been established to play a role in the etiology of GC. However, outside of the few known inherited syndromes that cause GC, there is a lack of data exploring other contributing genetic factors, specifically family-based studies. In addition, there is a dearth of information regarding familial GC in the background of breast cancer aside from Hereditary Diffuse Gastric Cancer and Peutz-Jeghers syndromes. This report aims to characterize a population with a family history of GC from the City of Hope Clinical Cancer Genetics Community Research Network (CCGCRN) and explore the presence of familial GC in individuals with a personal history of breast cancer. We aim to contribute to the knowledge-base of familial GC and illuminate potential associations. **Methods:** The CCGCRN's Institutional Review Board-approved registry was queried (Progeny® v9.1) for the keywords “family history of GC” obtained from patients’ individual data spreadsheets. Family information was entered into a database for analysis. Individual pedigree evaluations were completed for families with more than one family member afflicted with GC. Descriptive analysis using SPSS® v19 was conducted to create frequency tables. **Results:** There were 1116 (9.4%) families with a history of GC. Notably, 33% (365) reported Hispanic ancestry as compared to 26% Hispanic representation in the registry overall. Most families had one member with GC (919), 143 had 2 family members with GC and 54 had ≥ 3. A majority (657) of these families were based on a breast cancer index case. Only 153 (14%) of these families carried a known mutation, with the majority (99) carrying an HBOC gene mutation. The most common mutation was the Ashkenazi Jewish Founder mutation BRCA1 185delAG (11), then 3492insT (7) and 6174delTT (5) in *BRCA2* . Two of these mutations have been previously shown to be common Hispanic mutations in the U.S. Gastric cancer has been previously reported in association with HBOC. Our findings raise the question of an under appreciated association of GC in HBOC. However, a larger sample is needed to further define the phenotypic variation of GC in HBOC families. In addition, this data prompts the idea of population-dependent genetic modifiers that may play a role in Hispanic HBOC families.

3304T

Association of polymorphism in GSTM1 null with obesity in breast cancer patients triple negative. *O. Soto-Quintana[1,2], B. Sánchez[6], R. Rámirez[1,3], A. Ramos[1,2], DI. Carrillo[1,3], IA. Gutíerrez[1,3], AR. Rincón[4], AM. Puebla[5], MP. Gallegos[1].* 1) División de Genética Molecular, CIBO, IMSS, Guadalajara, Jal,. Mex, Mexico; 2) Doctorado en Farmacología, CUCS, U de G; 3) Doctorado en Genética Humana, CUCS, U de G; 4) Coordinación del Doctorado en Farmacologia, Departamento de Fisiología, CUCS, u de G; 5) Laboratorio de Inmunofarmacología, Depertamento de framacología, CUCEI, U de G; 6) Consulta Externa de Oncología, UMAE Hospital de Gineco'Obstetricia, CMNO, IMSS.

Glutathione S-transferases M1 (GSTM1) is an important phase II metabolizing enzyme. The null genotype of GSTM1 causes total loss of GSTM1 enzyme activity and numerous studies have investigated the association between GSTM1 null genotype and breast cancer risk. In this study was analyzed 67 breast cancer patients triple negative. The GSTM1 null genotype was determined by PCR. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated to estimate the association. The null genotype was associated with obesity in breast cancer patient's stage IV triple negative [OR 4.8(95%CI 1.5-14.8), p=0.00642]. The GSTM1 null genotype is associated with an increased risk of breast cancer patient's stage IV triple negative.

3305S

Genetic Variants related to presence of Bladder Cancer in a high risk, arsenic-exposed population in Northern Chile (Antofagasta). *C. Vial[1], K. Espinoza[1], E. Chaparro[2], G. Repetto[1], M.I. Fernández[1,3].* 1) Centro de Genética y Genómica, Facultad de Medicina, Universidad del Desarrollo Clinica Alemana, Santiago, Chile; 2) Servicio de Urología, Hospital Regional de Antofagasta, Chile; 3) Departamento de Urología, Clinica Alemana.

Introduction: Bladder Cancer (BC) is a heterogeneous disease with a variable natural history. It is currently the fourth most common malignancy among men in the western world, following prostate, lung, and colon cancers. Incidence in an arsenic-exposed city in Northern Chile (Antofagasta 24.8/100,000) is 4 to 5 times higher than in the rest of the country. Concentrations in drinking-water in Antofagasta increased significantly from 90 to 800-900 μg/L when new drinking-water sources were introduced in 1958. Following this, nearly all of the population in Antofagasta was exposed to arsenic levels up to seventeen times over the WHO recommendation between 1958 and 1971. In the present study we aim to perform a Genome Wide Association Study (GWAS) on people exposed to Arsenic comparing cases with controls. Methodology: Individuals were invited to participate after signing an informed consent. A blood sample was obtained and a questionnaire with epidemiological and clinical information was applied. DNA samples were analyzed using Affymetrix Genome-Wide SNP Array 6.0. After filtering by missingness per individual, missingness per marker allele frequency and Hardy Weinberg Equilibrium we obtained 788,705 SNPs to be analyzed. Results: 42 BC patients and 36 control subjects have been enrolled so far in the study. All of them were exposed to arsenic at some point between 1955 and 1971. The epidemiological data showed that males are predominant among cases and controls (64.3% and 83.3%; respectively) and there is no significant difference concerning mean age, familial history of BC, occupational exposure or smoking status between groups. It is interesting to note that smoking prevalence is high and similar among cases and controls (59.5% and 58.3%; respectively; p=0.55). The sample for population stratification was analyzed using principal component analysis (PCA), clustering the different patients with the identity by state and found them to be a homogeneous population. An association test comparing cases and controls was performed and found two regions with a significant association: (a) rs4838646 in chromosome 10 (p=3.8E-06) and (b) rs12371702 in chromosome 12 (p=5.8E-06). Previous studies have linked polymorphisms in the former region to BC susceptibility. Conclusion: Initial results of a BC genomic case-control study in an arsenic exposed population are presented. Further analysis is warranted after completion of recruitment. Fondecyt1120987.

3306M

Heterogeneous DNA methylation contributes to tumorigenesis by inducing the loss of co-expression connectivity in colorectal cancer. *Q. Wang[1], P. Jia[1], F. Cheng[1], Z. Zhao[1,2,3,4].* 1) Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN; 2) Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN; 4) Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN.

Increasing evidence indicates the high heterogeneity of cancer cells. Recent studies have revealed distinct subtypes of DNA methylation in colorectal cancer (CRC); however, the mechanism of heterogeneous methylation underlying tumor process remains poorly understood. Gene expression is a natural, intermediate quantitative trait that bridges genotypic and phenotypic features. In this work, we studied the role of heterogeneous DNA methylation in tumorigenesis via gene expression analyses. Specifically, we integrated methylation and expression data in normal and tumor tissues, respectively, and examined the perturbations in co-expression patterns. We found that the heterogeneity of methylation leads to significant loss of co-expression connectivity in CRC. Moreover, validation in an independent cohort confirmed the reliability of our findings. Functional analyses showed that the lost co-expression partners participate in important cancer-related pathways/networks, such as ErbB and MAPK signaling pathways. To our knowledge, this is the first study interpreting methylation heterogeneity in cancer from the perspective of co-expression perturbation. Our analyses suggest that the loss of co-expression connectivity induced by methylation heterogeneity plays an important role in CRC. These findings provide new perspectives in tumor biology and may facilitate the identification of potential biomedical therapies for cancer treatment.

3307T

Pathways associated with susceptibility of nasopharyngeal carcinoma (NPC) identified by whole-exome sequencing in Taiwanese NPC families. *G. Yu[1], W. Hsu[1], M. Yeager[1], CGR. Cancer Genomics Research Lab[1], C. Wang[2], P. Lou[2], S. Diehl[3], C. Chen[4], A. Hildesheim[1], A.M. Goldstein[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) National Taiwan University, Taiwan; 3) Rutgers School of Dental Medicine, Newark, NJ; 4) Academia Sinica, Taiwan.

NPC is a cancer that is known to be caused by infection with Epstein-Barr virus (EBV), a ubiquitous DNA virus that establishes lifelong latency in infected individuals. While >90% of the world population is infected with EBV during childhood or early adult life, only a very small fraction of infected individuals develop NPC. It is believed that, In addition to EBV infection and other environmental factors, genetic predisposition plays an important role in the pathogenesis of NPC. However, except for the genes that encode human leukocyte antigen (HLA), no other specific genes or chromosome regions have been shown to be consistently associated with NPC. In this study, we sequenced 83 exomes from 32 NPC families with 2-4 cases by Nimblegen v2.0 and v3.0 exome capture array and the Illumina HiSeq2000 sequencer. Reads were aligned using Novoalign v.2.07.14 and variants were called using GATK software. We prioritized 637 rare deleterious variants (frequency <1% in dbSNP, 1000 genomes or NHLBI exome database; no segmentally duplicated genes/variants; not in the repeated region; predicted as deleterious mutation by PolyPhen-2) with at least 10 reads and fitting a dominant segregation model in families with multiple NPC patients sequenced. Forty genes with functions related to EBV or NPC or infection and immunity were prioritized for technical validation in other members of the sampled families. Pathway enrichment analysis for the 637 prioritized deleterious variants in Kobas 2.0 (http://kobas.cbi.pku.edu.cn/home.do) suggested that three pathways including Notch signaling, ECM (extracellular matrix)-receptor interaction and Focal adhesion were significantly enriched in NPC cases (P<0.00087). All these pathways are critical for cell survival, growth and metastasis. Another set of NPC families (218 exomes, 110 families with 2-3 cases) will be exome sequenced to confirm our results and expand the search for NPC susceptibility genes.

3308S

Female Reproductive Traits and Lifespan in *BRCA* Families. *W. Zhuang[1], C. Snyder[1], M. Casey[2,3], M. Stacey[1], H. Lynch[1].* 1) Department of Preventive Medicine and Public Health, Creighton University, Omaha, NE; 2) Department of Obstetrics and Gynecology, Creighton University, Omaha, NE; 3) Department of Pathology, Creighton University, Omaha, NE.

Objective: A longer female reproductive lifespan increases exposure to endogenous estrogen and number of ovulations, which can be risk factors for breast and ovarian cancers. We investigated the effects of *BRCA1* and *BRCA2* on female reproductive traits and lifespan to understand the differences in risks for breast and ovarian cancers in *BRCA1* compared with *BRCA2* mutation carriers. Further, we explored the potential of other genes that influence reproductive traits/lifespan and are functionally independent of *BRCA1* and *BRCA2* mutations. **Methods:** We studied a total of 1058 women from the Creighton University Hereditary Cancer Registry. The women who tested positive or negative for their family's deleterious mutation in *BRCA1* and *BRCA2*, and provided age at menarche, age at natural menopause (ANM), or both were included in the study. The linear mixed-effects model was used to explore the effects of *BRCA1* and *BRCA2* on female reproductive traits and lifespan. Heritability estimates were calculated using variance-components methods in Sequential Oligogenic Linkage Analysis Routines (SOLAR) to explore the potential of other genes with respect to female reproductive traits and lifespan. **Results:** The reproductive lifespan of the women with a *BRCA1* mutation was 3.12 years shorter (p-value = 0.048, n=38) than the non-carrier relatives. The significant difference was not observed in the *BRCA2* families. Age at menarche was significantly heritable after adjusting for the presence of a *BRCA1* or *BRCA2* mutation (heritability [h^2] = 0.57 and 0.34, and n= 810 and 416, respectively; p-values <0.01). **Conclusions:** Women with a *BRCA1* mutation may have a shorter reproductive lifespan as compared to non-carrier relatives, causing a shorter period of exposure to estrogen and lower number of ovulations. The high heritability after adjusting for the presence of a *BRCA1* or *BRCA2* mutation underscores the importance of further work to identify other specific genes that contribute significantly to the variation in age at menarche in *BRCA* families.

3309M

Assessing the Cumulative Contribution of New and Established Common Genetic Risk Factors to Early-Onset Prostate Cancer. *K.A. Zuhlke[1], J.V. Ribado[2], A.M. Johnson[1], G.R. Keele[2], J. Li[2], Y. Wang[2,3], Q. Duan[2], G. Li[4], Z. Gao[4], Y. Li[2,3], J. Xu[4], S. Zheng[4], K.A. Cooney[1,5], E.M. Lange[2,3].* 1) Department of Internal Medicine, University of Michigan, Ann Arbor, MI; 2) Department of Genetics, University of North Carolina, Chapel Hill, NC; 3) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 4) Center for Genomics and Personalized Medicine Research, Wake Forest University, Winston-Salem, NC; 5) Department of Urology, University of Michigan, Ann Arbor, MI.

Genome-wide association studies (GWAS) have identified more than 60 loci associated with prostate cancer (PCa) in men of European ancestry, including 23 loci recently identified in a large study from the PRACTICAL consortium. We assessed the evidence for these 23 new PCa loci and the aggregate predictive value of these 23 SNPs plus 40 established SNPs with respect to early-onset (EO) PCa using 931 unrelated men diagnosed with PCa prior to age 56 years from the University of Michigan Prostate Cancer Genetics Project and 1126 male controls. Ten of the 23 new SNPs demonstrated evidence ($p < 0.05$) for association with EO disease. On average, EO PCa cases carried one more total risk alleles compared to controls across these 23 SNPs (21.61 vs. 20.69, $p=2.0\times10^{-12}$) and 3.5 more risk alleles across all 63 SNPs (58.02 vs. 54.49, $p=8.9\times10^{-59}$). We constructed receiver operating characteristic curves and calculated the corresponding area under the curves (AUC) for weighted aggregate risk allele counts for the 23 new PCa SNPs (AUC=0.59), 40 established PCa SNPs (AUC=0.69) and the set of 63 total PCa SNPs (AUC=0.71). The defined risk alleles and their associated weights were based on previous reports. We found that the aggregate burden of common risk alleles across previously associated PCa SNPs is highly predictive of EO PCa for men with values of total risk burden in either tail of the total risk allele distribution. The odds of a man having EO PCa given a burden of risk alleles in the upper, or lower, 5% of the combined case-control distribution of risk alleles across all 63 SNPs was estimated to be 11.11 [95% CI (5.91,20.90)] or 0.24 [95% CI (0.14,0.40)], respectively. While attenuated, odds remained strong for men in the upper or lower 25% of the total risk allele distribution (OR=3.81 [3.08,4.72] or OR= 0.27 [0.21,0.33], respectively). Inclusion of results from the 23 new variants only modestly improved disease prediction over the 40 established stronger effect SNPs. Our results provide the first formal replication for many of the 23 new variants recently reported to be associated with PCa and firmly establishes the importance of common variants, identified to be associated with PCa in GWAS, in men with early-onset disease.

3310T

Determination of cancer susceptibility in early-onset colorectal cancer (CRC) patients. *K.A. Schrader[1,6], J. Vijai[1,6], M. Artomov[2,6], T. Thomas[1], A. Kiezun[2], X. Wei[3], M. Corines[1], L. Jacobs[1], A. Lincoln[1], D. Villano[1], R. Rau-Murthy[1], L. Margolin[2], N. Gupta[2], M. Robson[1], R.J. Klein[4], L. Ellisen[5], M. Daly[2], S. Lipkin[3], G. Getz[2], S. Gabriel[2], D. Altshuler[2], K. Offit[1], Z. Stadler[1].* 1) Memorial Sloan Kettering Cancer Center, New York, NY; 2) Broad Institute of Harvard and MIT, Cambridge, MA; 3) Department of Medicine, Weill Cornell College of Medicine, New York, NY; 4) Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai New York, NY; 5) Massachusetts General Hospital, Boston, MA; 6) equal contributions.

Germline mutations in high-penetrant cancer susceptibility genes have been implicated in CRC risk. As part of a larger project elucidating the genetic basis of CRC, we surveyed whole-exome sequencing (WES) data from early-onset CRC patients, for germline variants in known cancer predisposition genes. CRC patients, diagnosed at ≤45 years of age, with unrevealing clinical genetics evaluations, were ascertained under IRB-approved research protocols allowing for WES at MSKCC's Clinical Genetics Service. Cases were enriched for DNA mismatch repair (MMR) proficient CRC or MMR-deficient CRC due to somatic *MLH1* hypermethylation. Germline DNA was subjected to exome capture, paired-end sequencing, and variant calling by GATK Unified Genotyper. Variant curation was by Ingenuity Variant Analysis(c) with 150 cancer susceptibility genes analyzed. Variants were excluded, with minor allele frequencies (MAF) >5% in public databases, at base positions with >10% missing calls and if novel and at >20% MAF within the cohort. Preliminary analysis involved visualization of BAM files and inclusion of only unique protein truncating variants. Analysis of 63 early-onset CRCs revealed 15 patients with presumed deleterious heterozygous protein truncating variants in 11 genes; *MSH2*, *STK11*, *MUTYH*, *BAP1*, *BUB1B*, *BLM*, *FANCL*, *COL7A*, *ATR*, with 2 variants in *PMS2* and 4 in *ATM*. Family history in the *MSH2* variant carrier was consistent with Lynch Syndrome. In one *PMS2* variant carrier, the tumor showed isolated PMS2 protein loss. No polyps were detected in carriers of *MUTYH* or *BUB1B* variants, but were present in both family histories. The *STK11* variant carrier also had bronchioloalveolar carcinoma (BAC) at age 39, and no classic Peutz-Jeghers stigmata. In 2 *ATM* variant carriers, family histories included pancreatic, ovarian and breast cancer. The *BAP1* variant carrier had no relevant history. Half of the cancer susceptibility genes harboring unique protein truncating variants found by WES of early-onset CRC are not implicated in CRC susceptibility and in most cases, clinical history was not suggestive of the identified genetic variant. However, among these findings, an identified *STK11* variant carrier without classical features of Peutz-Jeghers had BAC, that has been reported in association with *STK11* germline and somatic mutations. Multiplex panels will potentially broaden the phenotypic spectrum of well-defined syndromes; clinical correlation and proof of causation will remain a challenge.

3311S

Hispanic MMR Mutations: A Multi-Institutional Report from Southwestern United States and Puerto Rico. *A. Sunga[1], C. Ricker[2], C. Espenschied[3], J. Herzog[1], S. Bannon[4], M. Cruz Correa[5], P. Lynch[4], S. Gruber[2], J. Weitzel[1].* 1) Clinical Cancer Genetics, City of Hope, Duarte, CA; 2) Clinical Cancer Genetics, USC Norris Comprejhensive Cancer Center, USC Los Angeles, CA; 3) Ambry Genetics, Aliso Viejo, CA; 4) Clinical Cancer Genetics, MD Anderson Cancer Center, Houston, TX; 5) Clinical Cancer Genetics, University of Puerto Rico Comprehensive Cancer Center, Rio Piedras, Puerto Rico.

Background Knowledge of founder mutations enable efficient and cost-effective strategies for genetic testing, a potential benefit for populations with limited access to services. We have shown several BRCA mutations, most of Spanish origin, to be founder mutations in Hispanic populations. We hypothesized that the same population factors may be operative in Lynch syndrome. There is limited literature on the spectrum of mutations in mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) in Hispanic populations. Our goal was to identify recurrent MMR mutations in Hispanic patients and explore potential ancestral origins of identified mutations. **Methods** Subjects: Hispanics patients seen for genetic risk assessment at City of Hope Clinical Cancer Genetics Community Research Network collaborating institutions, MD Anderson Cancer Center, USC Norris Cancer Center and the University Of Puerto Rico Cancer Center. A total of 397 patients underwent evaluation for Lynch syndrome by genotyping of one or more of the MMR and/or tumor analysis by IHC. A comprehensive literature review and of MMR variant databases was conducted for all mutations. **Results** Seventy one (18%) unrelated Hispanic families were found to have at least one MMR disease predisposing gene mutation. Most (90%) of the mutations were noted in the *MLH1* and *MSH2* genes and more than half (55%) were frameshift and nonsense type. Nine mutations were observed two or more times: *MLH1* 350C>T, 1790del2ins9, 2041G>A, 1852del3, 1024del16, and IVS7+5G>A; and *MSH2* 1216C>T, 425C>G, and 1705delGA. Most (6/9) recurrent mutations were detected in multiple institutions. *MSH2* 1216C>T and *MLH1* 1852del3 were seen 5 and 4 times respectively and have been reported multiple times in European populations. *MLH1* 350C>T, 2041G>A, 332C>T, and 676C>T and *MSH2* 1216C>T, exon 4-8 del, and exon 8 del have all been reported previously in Spain. The *MSH2* exon 4-8 del was seen in one Mexican family and has been reported as a Spanish founder mutation. **Conclusion** Our finding that 3/9 recurrent mutations and the *MSH2* exon 8 del identified in our cohort were previously reported in Spain supports the hypothesis of the likely influence of Spanish ancestral heritage on MMR gene mutations in Hispanic populations. While this is the largest reported cohort of Hispanic patients with MMR mutations in N America, a larger sample and haplotype analyses are needed to better define the spectrum and origin of MMR mutations in Hispanic populations.

3312M

Pleitropy between Hodgkin lymphoma and other immunological diseases. *W. Cozen[1], P. Khankhanian[2], Y. Kong[1], D. Himmelstein[2], R. Jarrett[3], J. McKay[4], A. van den Berg[5], F. Gilliland[1], K. Onel[6], S. Baranzini[2], W.J. Gauderman[1], J. Oksenberg[2], H. Hjalgrim[7], EVE Consortium.* 1) Preventive Medicine and Pathology, USC Keck School of Medicine, Norris Comprehensive Cancer Center, Los Angeles, CA; 2) Department of Neurology, University of California San Francisco 675 Nelson Rising Lane Suite 215 San Francisco, CA; 3) MRC - University of Glasgow Centre for Virology Research Institute of Infection, Immunity and Inflammation University of Glasgow Garscube Estate Glasgow G61 1QH; 4) Genetic Cancer Susceptibility group, International Agency for Research on Cancer, 150 Cours Albert Thomas, 69372 Lyon CEDEX 08, France; 5) Department of Pathology & Medical Biology University Medical Center Groningen University of Groningen Hanzeplein 1 P.O. Box 30.001 9700 RB Groningen The Netherlands; 6) Section of Hematology/Oncology The University of Chicago 900 East 57th Street, Rm 5140 MC 4060 Chicago, IL 60637; 7) Consultant Department of Epidemiology Research Statens Serum Institut 5 Artillerivej, DK-2300 Copenhagen S Denmark.

Hodgkin lymphoma (HL) is a highly heritable B-cell lymphoma that primarily affects young adults. Its epidemiology and pathology are unique and more closely aligned with immunological diseases than with other solid tumors. To test this hypothesis, we conducted studies examining the genetic overlap between HL and multiple sclerosis (MS) and asthma. We performed a meta-analyses of HL and MS consisting of 1,816 HL patients, 9,772 MS patients, and 25,255 controls using 464,424 single nucleotide polymorphisms. Genetic overlap was observed in the HLA region and genome-wide, including modest genome-wide significant associations at two loci that did not reach genome-wide significance in either disease alone: IL12B (rs2546890, $OR_{MS\text{-}HL\text{-}meta}$=1.1, p=3.37 x 10^{-8}), a gene encoding a subunit of a cytokine involved in T-helper cell Type 1 (Th1) immunity, and NCOA5 (rs2425752, $OR_{HL\text{-}MS\ meta}$=1.11, p=2.96 x 10^{-8}), a gene included in an estrogen-dependent oncogenic pathway. We conducted a separate analysis of HL and asthma combining GWAS meta-analysis of 1,816 HL cases and 7,877 HL controls with a GWAS meta-analysis of 2,088 asthma cases and 2,743 asthma controls examining 904,634 SNPs in common. We found associations with two linked SNPs in the T-helper cell Type 2 (Th2) transcription factor gene GATA3 (rs422628, $OR_{HL\text{-}asthma\ meta}$ =1.247, p=3.36 x 10^{-9} and rs444929, $OR_{HL\text{-}asthma\ meta}$ = 1.264, p=2.09 x 10^{-8}), both of which were significant at the p <0.05 level in the disease-specific GWAS, but reached genome-wide significance in the HL-asthma meta-analysis. The effect was stronger in the subset of nodular sclerosis HL (rs422628, $OR_{HL\text{-}asthma\ meta}$ = 1.32, p=2.33 x 10^{-9} and rs444929, $OR_{HL\text{-}asthma\ meta}$ = 1.32, p=4.28 x 10^{-8}). Another genome-wide significant SNP near KIAA1279 reached genome-wide significance only in the nodular sclerosis subtype subset (rs6864, $OR_{HL\text{-}asthma\ meta}$ = 0.81, p=4.65 x 10^{-8}). In a genetic diseasome analysis, HL was more closely aligned to both MS and asthma and other immunological diseases than to solid cancers. Recognition of overlap in genetic predisposition to HL and immune diseases sheds light on the complex etiology of HL and may enable novel diagnostic and therapeutic approaches.

3313T

Prospectively Identified Incident Testicular Cancer Risk in a Familial Testicular Cancer Cohort. *A. Pathak[1], C. Adams[1], J. Loud[1], K. Nichols[2], M. Greene[1], D. Stewart[1].* 1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; 2) Westat, Inc., Rockville, MD, USA.

Testicular germ cell tumors (TGCT) are the most common form of cancer in men aged 15-35 years. Approximately 8,820 new cases and 380 TGCT deaths are projected for 2014. Human testicular germ cell tumors (TGCT) have a strong genetic component and a high familial relative risk. TGCT has an estimated heritability that ranks as the 3rd highest for all cancers. Compared with most cancers - which have familial relative risks between 1.5-to-2.5 fold - sons and brothers of affected men with TGCT have a 4- to 6-fold and 8- to 10-fold increased risk of TGCT versus the general population, respectively. These risks increase to 37-fold and 76.5-fold in dizygotic and monozygotic twins. However, linkage analysis has failed to identify any rare, highly-penetrant familial TGCT (FTGCT) susceptibility loci. Currently, multiple low-penetrance genes are hypothesized to underlie the familial multiple-case phenotype. Despite the apparent heritability of TGCT, families with more than two affected members are unusual, unlike other hereditary cancer syndromes in which a single multigenerational pedigree can harbor many affected individuals. Importantly, the prospective TGCT risk in this context is unknown. Thus, we performed a prospective quantitative analysis of TGCT incidence in a cohort of multiple affected person and sporadic bilateral case families; 1,260 men from 140 families (10,207 person-years of follow-up) met our inclusion criteria. Age-, gender-, and calendar time-specific standardized incidence ratios (SIR) for TGCT relative to the general population were calculated using SEER*Stat. Eight incident TGCTs occurred during prospective FTGCT cohort follow-up (versus 0.67 expected; SIR= 11.9; 95% confidence interval [CI]=5.1-23.4). We demonstrate that the incidence rate of TGCT is much greater among bloodline male relatives in multiple-case testicular cancer families than that expected in the general population, a pattern characteristic of adult-onset Mendelian cancer susceptibility disorders. Remarkably, two incident TGCTs occurred in healthy relatives of sporadic bilateral cases (0.15 expected; SIR=13.4; 95% CI=1.6-48.6). Our data are the first to indicate that, despite relatively low numbers of affected individuals per family, members of both multiple affected person FTGCT families and sporadic bilateral TGCT families comprise high-risk groups for incident testicular cancer, which might benefit from tailored risk stratification and surveillance strategies.

3314S

The significance of N-acetyltransferase 2 (NAT2) genotypes in combination with phenotypes to risk of bladder cancer in a Chinese population. *K. Chattopadhyay[1], Y-B. Xiang[2], H.H. Nelson[3], R. Wang[1], Y-T. Gao[2], J-M. Yuan[1].* 1) University of Pittsburgh Cancer Institute, and Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; 2) Department of Epidemiology, Shanghai Cancer Institute, Shanghai Jiatong University, Shanghai, China; 3) Masonic Cancer Center, and Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA.

1) Current Affiliation: F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Department of Gastroenterology, Cedars-Sinai Medical Center, Los Angeles, California, USA. Background: The incidence rates of bladder cancer in Chinese men and women are only one-third those in U.S. whites although smoking rate, a major risk factor for bladder cancer, is higher in Chinese males than U.S. white males. N-acetyltransferase 2 (NAT2) plays a major role in catalyzing the detoxification of aromatic amines in cigarette smoke and other environmental sources. Some of aromatic amines are carcinogenic to the urinary bladder. The difference in genetically determined acetylation status may contribute to the differential risk of bladder cancer between Chinese and whites. Methods: A population based case-control study of bladder cancer (499 cases, 513 controls) in Shanghai, China was conducted. Six common single nucleotide polymorphisms (SNPs) of NAT2 gene (803A>G, 282C>T, 481C>T, 590G>A, 857G>A, 341T>C) were determined using TaqMan assay. The NAT2 acetylation phenotype was determined by the ratio of urinary caffeine metabolites. Unconditional logistic regression methods were used to calculate odds ratio (OR) and 95% confidence interval (CI) of bladder cancer associated with NAT2 phenotypes, genotypes and haplotypes with adjustment for age, sex, alcohol intake and cigarette smoking. Results: Mutant alleles of any six NAT2 SNPs were associated with reduced acetylation status in a dose-dependent manner among healthy individuals. The NAT2 genotypes referring to intermediate acetylation status was associated with significantly reduced risk (OR = 0.70, 95% CI = 0.53-0.92, p = 0.012) of bladder cancer compared to the genotypes referring to rapid acetylation status. The genotypes referring to slow acetylation status did not show any significant association with the disease. Conclusion: Our results show a novel association of NAT2 genotypes with bladder cancer in a Chinese population. The results must be interpreted cautiously and further work with a larger population is warranted.

3315M

DEPTH: A Novel Algorithm for Feature Ranking with Application to Genome-Wide Association Studies Identifies that Variation in the ESR1 Gene Region is Associated with Risk of Estrogen Receptor Negative Breast Cancer from a Small Study. *E. Makalic, D.F. Schmidt, M. Kapuscinski, J.L. Hopper.* School of Population and Global Health, The University of Melbourne, Victoria, Australia.

A conventional Genome-Wide Association Study (GWAS) involves selecting single nucleotide polymorphisms (SNPs), from among a very large set of correlated SNPs, that discriminate between two groups. The ultimate aim is to identify susceptibility genes, or gene regions; the dominant paradigm to date has been to identify SNPs associated with risk of disease. We propose a new algorithm, called DEPendency of association on the number of Top Hits (DEPTH), that achieves the aim more efficiently by using bootstrap statistics and stability selection and considering contiguous SNPs as a group. DEPTH is applicable to ordinary, logistic, and Cox regression and can be run on a commodity computer but is implemented to run much faster on the IBM BlueGene/Q supercomputer. In the context of a GWAS, the algorithm passes a sliding window across the whole genome, and can be applied to subsets of SNPs (e.g. in a region or pathway, or of a particular 'type'). We have found, using simulated data, that the algorithm shows good statistical performance when compared to several established procedures.

We applied DEPTH to data from the hypothesis-generating Phase I of a GWAS that included Australian women with breast cancer diagnosed before age 40 years (72 estrogen receptor (ER) -ve; 88 ER +ve) and 287 women 40 years and older without breast cancer, genotyped using the Illumina 610KQUAD platform with standard QC. DEPTH identified that germline variation in the region of the *ESR1* gene was implicated in risk of ER-ve disease, and more so than ER+ve disease, a finding not evident from applying the conventional 'genome-wide significant' threshold to these data. Subsequent analysis of iCOGS data for ~9,500 ER-ve and ~32,000 ER+ve cases and ~51,000 controls found evidence consistent with these hypotheses. Our insights from a small GWAS validates the utility of our approach in terms of the novel design - using cases (controls) enriched (depleted) for putative genetic risk - and analysis using DEPTH.

3316T

DEPTH: A Novel Algorithm for Feature Ranking with Application to Genome-Wide Association Studies Identifies that Variation in the CHEK2 Gene Region is Associated with Risk of Breast and Colorectal Cancer. *D.F. Schmidt[1], E. Makalic[1], M. Kapuscinski[1], W. Chen[2], P.M. Lynch[3], C.I. Amos[4], J.L. Hopper[1], M. Pande[3].* 1) School of Population and Global Health, The University of Melbourne, Victoria, Australia; 2) Department of Clinical Applications and Support, MD Anderson Cancer Center, TX, USA; 3) Department of Gastroenterology - Research, MD Anderson Cancer Center, TX, USA; 4) Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA.

A conventional Genome-Wide Association Study (GWAS) involves selecting single nucleotide polymorphisms (SNPs), from among a very large set of correlated SNPs, that discriminate between two groups. The ultimate aim of a GWAS is to identify susceptibility genes, or gene regions; the dominant paradigm to date has been to do this by identifying individual SNPs associated with risk of disease. We propose a new algorithm, called DEPendency of association on the number of Top Hits (DEPTH), that achieves the above aim more efficiently by using bootstrap statistics and stability selection and considering contiguous SNPs as a group. DEPTH: (i) exploits information from the correlation structure of the predictors without specifying an underlying model; (ii) applies to ordinary, logistic, and Cox regression, so can be used to study continuous, binary and survival data, and (iii) can be run on a commodity computer but has also been implemented to run much faster in a parallel computing environment on the IBM BlueGene/Q supercomputer. In the context of a GWAS, the algorithm can consider SNPs across the whole genome, or subsets of SNPs (e.g. in a region or pathway, or of a particular 'type'). We have found, using simulated data, that the algorithm shows good statistical performance when compared to several established procedures. We applied DEPTH to a subset of GWAS data from the Breast and Colon Cancer Family Registries to try to elucidate genes/regions associated with susceptibility to breast and/or colorectal cancer. Cases (985) were women with breast cancer and a family history of colorectal cancer, or persons with colorectal cancer with a family history of breast cancer. Controls (1769) were frequency matched to cases for age and sex and had neither breast nor colorectal cancer. DEPTH identified that germline variation in the region of the checkpoint kinase 2 (CHEK2) gene is associated with risk of breast and/or colorectal cancer, a finding not evident from applying the conventional 'genome-wide significant' threshold. Lending support to our finding, several germline mutations in CHEK2 (e.g., 1100delC, IVS2+1G>A and I157T) are known to be associated with the risk of breast, colorectal, ovarian, lung and other cancers. Our finding from studying a modestly sized GWAS validates the utility of our approach both in terms of the novel design (using family history across diseases) and analysis (DEPTH).

3317S

Integrated pathway and gene-gene interaction analysis reveals novel candidate genes for melanoma. *M. Brossard[1,2], S. Fang[3], A. Vaysse[1,2], Q. Wei[4], H. Mohamdi[1,2], W.V. Chen[1,2], N. Lavielle[1,2], E. Maubec[1,2], M.-F. Avril[5], M. Lathrop[6], J.E. Lee[3], C.I. Amos[7], F. Demenais[1,2], MELARISK and MDACC Melanoma Study Groups.* 1) INSERM UMR-946, Paris, France; 2) Université Paris Diderot, Paris, France; 3) MD Anderson Cancer Center, Houston, Texas, USA; 4) Department of Medicine, Duke University School of Medicine, Durham, USA; 5) Hôpital Cochin, Université Paris Descartes, Paris, France; 6) Genome Quebec Innovation Centre, McGill University, Montreal, Canada; 7) Geisel College of Medicine, Dartmouth College, New Hampshire, USA.

Genome-wide association studies (GWAS) have identified 17 loci associated with melanoma risk. However, these loci account for a small part of melanoma susceptibility. These GWAS, which have focused on the analysis of individual SNPs, are underpowered to detect SNPs with small marginal effect. To increase the yield from GWAS, alternative approaches that use biological knowledge and allow testing for association on the basis of functional units (pathways) have been proposed. To identify new candidate genes for melanoma, we proposed an analysis strategy that combines pathway analysis and tests of gene-gene interactions (GxG) within melanoma-associated pathways. Pathway analysis was conducted using the gene-set enrichment analysis (GSEA) approach, which searches for gene sets, defined by gene ontology (GO) classes, enriched in genes associated with melanoma. GSEA was applied to single-SNP statistics obtained from GWAS of the French MELARISK study (3,976 subjects) and the North-American MDACC study (2,827 subjects). To identify enriched GO categories, we computed the false discovery rate (FDR), using 100,000 permutations of SNP statistics. We tested all cross-gene SNP-SNP interactions within each identified GO using INTERSNP. One million Hapmap3-imputed SNPs were assigned to 22,000 genes, that were assigned to 316 Level4-GO classes. Five GO categories were found to be significantly enriched in genes associated with melanoma (FDR≤5% in both studies): response to light stimulus, regulation of mitotic cell cycle, induction of programmed cell death, cytokine activity and oxidative phosphorylation. A total of 101 genes were driving the enrichment signals in these pathways. Five of these genes (*TP63*, *IL6*, *IL15*, *MAPK1*, *NDUFA2*) were found to occur frequently with melanoma-related terms through text mining of PubMed abstracts and represent novel candidates for melanoma risk. Analysis of SNP-SNP interactions within each of the five identified GO showed evidence for interaction for 4 SNP pairs ($P \leq 10^{-4}$ in MELARISK and replication at 5% in MDACC). One of these gene pairs, *TERF1-AFAP1L2* (combined P-value over the two samples=2×10^{-6}), is particularly biologically relevant. TERF1 and AFAP1L2 proteins have been recently reported to interact for telomere signaling, a process which is emerging as playing a key role in melanoma. Thus, an integrated pathway and gene-gene interaction analysis is a powerful approach to uncover new disease genes.

3318M

Combined pathway and gene-gene interaction analysis pinpoints biologically relevant genes for a major melanoma prognosis factor. *A. Vaysse[1,2], S. Fang[3], M. Brossard[1,2], H. Mohamdi[1,2], W.V. Chen[3], N. Lavielle[1,2], E. Maubec[1,2], Q. Wei[4], M. Lathrop[5], M.-F. Avril[6], C.I. Amos[7], J.E. Lee[3], F. Demenais[1,2], MELARISK and MDACC Melanoma Study Groups.* 1) INSERM U946, Paris, France; 2) Université Paris Diderot, Paris, France; 3) MD Anderson Cancer Center (MDACC), Houston, Texas, USA; 4) Department of Medicine, Duke University School of Medicine, Durham, USA; 5) Genome Quebec Innovation Centre, McGill University, Montreal, Canada; 6) Hôpital Cochin, Université Paris Descartes, Paris, France; 7) Geisel College of Medicine, Dartmouth College, New Hampshire, USA.

Cutaneous melanoma is one of the deadliest cancers if diagnosed at a late stage. Breslow thickness (BT), a measure of invasion of melanoma in the skin, is a major predictor of melanoma survival. To date, the genetic factors underlying BT are largely unknown and have not yet been investigated at the genome-wide level. We first conducted a GWAS of BT in the French MELARISK study (966 cases) and the North-American MDACC study (1546 cases) using Hapmap3-imputed SNPs. Single SNP analysis in MELARISK showed evidence for five loci that reached $P<10^{-5}$ but none of these associations was replicated in MDACC, suggesting the existence of many genetic variants with small effect. Since the analysis of individual SNPs may be underpowered to detect SNPs with modest effect, we used a multi-marker analysis strategy that integrates pathway analysis and tests of cross-gene SNPxSNP interactions within BT-associated pathways. Pathway-analysis was based on the gene-set enrichment analysis (GSEA) approach, which searches for gene sets, defined by Gene Ontology (GO) classes, enriched in genes associated with BT. For each GO, we computed the false discovery rate (FDR), using 100,000 permutations of SNP statistics. Cross-gene SNPxSNP interactions within each identified GO were investigated through linear regression. One million Hapmap3-imputed SNPs were assigned to 21,917 genes using NCBI Build 37.1. These genes were assigned to 317 level 4-GO categories. Three GO categories were found to be significantly enriched in genes associated with BT (FDR≤5% in both studies): Hormone activity, Cytokine activity and Myeloid cell differentiation. A total of 61 genes were driving these pathways in both studies. Interestingly, expression of four of these genes (*CXCL12, TNFSF10, VEGFA, CDC42*) was reported to be associated with melanoma progression in tumors. Analysis of cross-gene SNPxSNP interactions within each identified pathway showed evidence for interaction for three SNP pairs ($P \leq 10^{-4}$ in MELARISK and replication at $P \leq 0.05$ in MDACC). One of these gene pairs (*SCINxCDC42*, combined P over the 2 samples=2×10^{-6}) has biological relevance. SCIN and CDC42 proteins are both involved in actin cytoskeleton dynamics, that plays a major role in cell migration, and have opposite roles: CDC42 stimulates actin assembly while SCIN severs the actin filaments, thus preventing their growth. This study outlines the importance of integrating various approaches for gene identification. Funding: INCa_5982.

3319T

A Population-based survey of excess cancers observed in NF1 cases and in their first- and second-degree relatives. *D. Abbott[1], D. Viskochil[2], D.A. Stevenson[2], L.A. Cannon-Albright[1,3].* 1) Division of Genetic Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT; 2) Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT; 3) George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT.

Background: Neurofibromatosis type I (NF1) is an autosomal dominant disorder that affects approximately 1 in 3000 individuals and is caused by a mutation in NF1, a tumor suppressor gene that encodes the protein neurofibromin. While NF1 individuals often have benign peripheral nerve sheath tumors (neurofibromas) and intracranial low-grade astrocytomas (gliomas), they are also at increased risk for malignant peripheral nerve sheath tumors (MPNST) and high-grade astrocytomas. Current research purports that NF1 individuals have increased risk for other cancers. We analyzed a unique population-based resource to characterize the increased risk of various cancers in NF1 individuals and their relatives in an unbiased fashion. Methods: The Utah Population Database (UPDB), a population-based Utah genealogy linked to both the Utah Cancer Registry and to hospital data from the University of Utah Health Sciences Center, was used to calculate relative risks (RR) of cancer in NF1 cases and in their first- and second-degree relatives. RR was calculated using age-, birth-state-, and birth-year-specific cancer rates. Results: We identified 237 individuals with NF1 who had ancestors in the UPDB. These cases had 1055 first-degree and 2770 second-degree relatives in the UPDB. The RRs for 6 cancers were significantly higher in NF1 cases: Malignant Peripheral Nerve Sheath Tumor-MPNST (RR=646.9; 95%CI: 260.07, 1332.83), brain (43.80: 20.03, 83.15), central nervous system (511.08: 165.94, 1192.66), lymphoma (6.04: 1.65, 15.47), skin melanoma (4.97: 1.02, 14.52), and breast cancer (3.6;1.01, 9.23). The RR of seven cancers were significantly higher in first-degree relatives: MPNST (RR=118.03:47.45, 243.18), brain (5.48: 2.20, 11.29); mature T-cell lymphoma (13.75: 1.67, 49.68), acute lymphocytic leukemia (6.26: 1.29, 18.29), lentigo maligna melanoma (6.24: 1.29, 18.24), breast cancer (1.74: 1.01, 2.82), and small intestine (7.07: 1.01, 25.54). Lastly, in second-degree relatives there was a significantly increased risk of two cancers: MPNST (16.29; 3.36, 47.60), and prostate (1.44; 1.11, 1.84). Conclusion: These results support an increased risk of distinct cancers in individuals with NF1 and also in their first- and second-degree relatives. Relative risks for various cancers are much higher for NF1 cases themselves, but first- and second-degree relatives still display a significant excess of some cancers.

3320S

Identification of somatic structural events associated with L1 element activity in 208 colorectal cancers using whole genome sequence analysis. *T. Cajuso[1,2], E. Pitkänen[1,2], R. Katainen[1,2], E. Kaasinen[1,2], J. Taipale[1,3], LA. Aaltonen[1,2], O. Kilpivaara[1,2].* 1) University of Helsinki, Helsinki, Finland; 2) Department of Medical Genetics, University of Helsinki, Helsinki, Finland; 3) Science for Life Center, Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden.

L1 insertions have been shown to be frequent events in colorectal cancer (CRC). The role of L1 insertions to CRC tumorigenesis remains, however, unclear. L1 elements are retrotransposons that spread by copying themselves into an RNA intermediate and integrating at distant locations in the genome. Whole genome sequencing (WGS) has emerged as a novel tool to systematically search for structural variation. Utilizing WGS analysis, we recently detected insertions stemming from an active L1 element, located in the first intron of *Tetratricopeptide repeat domain 28* (*TTC28*). We have extended our WGS data to 208 CRC and respective germline DNA samples. We identified somatic structural variation including deletions, tandem duplications, inversions and translocations in the WGS data with DELLY. To reduce false positive calls, data was filtered against structural variation in the germline genomes and repeat regions in the genome. Structural events with breakpoints close to L1 elements as described in retrotransposon databases were investigated in detail. We first identified a total of 661818 structural somatic events. Subsequently, we focused on the 2619 structural somatic events located in close proximity to L1 elements (< 1000bp distance). From those, we have selected the 178 events that were present in more than one CRC, for visual inspection of mapped sequencing data and validation by Sanger sequencing. Careful analysis of WGS data is a prerequisite for accurate identification of structural variation and it should include consideration of retrotransposition events in cancer.

3321M

Duplication of 7p21.3-p14.3: metastasis risk, cancer susceptibilities and ethical implications. *M.R.S. Carvalho[1], M.P. Almeida[1], M. Miranda[1], L.T. Grillo[1], T.B. Teixeira[3], A.C.A. Prado[3], A.E. Paiva[1], P.A. Fonseca[1], L.L. Leão[3], M.J.B. Aguiar[3], V.G. Haase[2].* 1) Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 2) Departamento de Psicologia, Faculdade de Filosofia e Ciências Humanas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 3) Serviço Especial de Genética Médica, Hospital de Clínicas, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil.

The advent of powerful tools to identify copy number variations (CNVs) has opened new practical and ethical questions related to the functional implication of some chromosomal variants. Specifically, these questions include incidental findings. Recently, we evaluated a 17-year-old female with a history of psychomotor developmental delay, borderline intelligence, dysmorphic features, triphalangeal thumbs, hypotonia and severe hyperextensibility. A large 7p chromosome insertion was detected in the karyotype. Using CytoSNP-850K BeadChip (Illumina Technologies), we were able to identify a previously undescribed 16.4 Mb duplication (arr[hg19]7p21.3-p14.3 (13.265.700- 29.668.638)x3). Genes related to cancer and epithelial-mesenchymal transition (and therefore, metastasis) map to this chromosomal region, including the *HOXA* cluster, *TWIST1* (MIM 601622) and *HOTAIRM1.* The *HOXA4* (MIM 142953), *HOXA5* (MIM 142952), *HOXA7* (MIM 142950), *HOXA9* (MIM 142956) and *HOXA13* (MIM 142959) have been related to leukemia. Overexpression of these five *HOXA* genes, due to duplication of chromosome 7 in a gastric cell line, has been associated with tumorigenicity. Most notably, *HOXA9* overexpression has been described in the accelerated phase of chronic myeloid leukemia (CML[MIM 608232]). Overexpression of *TWIST1* has been related to metastatic cancer in general, advanced oral squamous cell carcinoma (OSCC), and with poor survival rates. Since large duplications have been only rarely described, it is impossible properly to establish the risks and management protocols for cancer surveillance for this patient. Moreover, the high metastasis risk unrelated to one specific type of cancer represents a complex situation in genetic counseling. In conclusion, the incidental findings associated with this specific chromosomal duplication, particularly the high metastasis risk, puts forward ethical dilemmas concerning the right not to know, as well as a strategy to care the patient.

Support: FAPEMIG, PPSUS/FAPEMIG, CNPq, CNPq 307975/2010-0, CAPES, CAPES-DAAD (PROBRAL).

3322T

A copy number variation genome-wide association study identifies two new cervical cancer susceptibility loci at NEDD4L and CTDSPL. *D. Chen[1], Z. Li[2], T. Cui[1], D. Zhang[2], U. Gyllensten[1].* 1) Department of Immunology, Genetics and Pathology, Uppsala University, Sweden, Uppsala, Sweden; 2) Department of Pathology, School of Medicine, Zhejiang University, Hangzhou, People's Republic of China.

Cervical cancer is the third most common cancer in women worldwide. Genome-wide association studies (GWASs) focusing on evaluating common single nucleotide polymorphisms (SNPs) have identified several genetic susceptibility loci for cervical cancer. However, the vast majority of risk variants identified to date have small effect sizes and only explain a small fraction of the heritability. Very few studies have evaluated copy number variation (CNV), another important source of human genetic variation, in relation to cervical cancer risk. We have conducted the first CNV GWAS in 1034 cervical cancer patients and 3948 control subjects in a Swedish population using data from Illumina HumanOmniExpress BeadChip. We used two CNV detection software suites-PennCNV and QuantiSNP for detecting both rare and common CNVs. Only the samples and CNVs that passed the quality control steps of both programs were included in the final analysis. Burden analysis showed that the average number of common CNVs (frequency ≥1%) per person and that of rare CNVs (frequency <1%) per person were significantly higher in cervical cancer patients than in controls (6.56 compared to 5.55 for common deletion, $P<10^{-6}$; 1.12 compared to 0.87 for common duplication, $P<10^{-6}$; 2.35 compared to 2.0 for rare deletion, $P<10^{-6}$; and 0.94 compared to 0.74 for rare duplication, $P<10^{-6}$; respectively). We also found that one deletion in *NEDD4L* (odds ratio [OR]= 4.09, 95% confidence interval [CI]=2.08-8.04, $P=7.0\times10^{-5}$) and one deletion in *CTDSPL* (OR=2.98, 95% CI=2.14-4.15, $P=1.0\times10^{-6}$) were significantly associated with increased risk of developing cervical cancer, respectively. These findings were further validated using real-time quantitative polymerase chain reaction (qPCR) in the same subjects. *NEDD4L* encodes a member of the Nedd4 family of HECT domain E3 ubiquitin ligases. The encoded protein mediates the ubiquitination of multiple target substrates and plays a critical role in epithelial sodium transport by regulating the cell surface expression of the epithelial sodium channel. *CTDSPL*, also known as *RBSP3*, is a tumor suppressor gene implicated in major epithelial malignancies and is frequently altered in premalignant cervical lesions. Our study provides evidence for two novel cervical cancer loci and new insights into the genetic etiology of cervical cancer. These two CNVs are the strongest genetic risk variants identified so far for cervical cancer.

3323S

Comprehensive genetic analysis of cell-free circulating nucleic acids through next-generation sequencing. *J. Fan[1], X. Cai[1], S. Munchel[1], R. Pantoja[1], N. Salathia[1], H. Xu[1], Y. Zhao[1], B. Bibikova[1], F. Kaper[1], A. Aravanis[1], J. Stone[1], R. Shen[1], R. Rava[1], F. Janku[2], T. Hinoue[3], D. Weisenberger[3], P. Laird[3], R. Klausner[1].* 1) Oncology, Illumina, Inc, San Diego, CA; 2) UT M. D. Anderson Cancer Center, Houston, TX; 3) USC Epigenome Center, Keck School of Medicine of University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA.

Cell-free circulating nucleic acids (cfDNA/cfRNA), so-called "liquid biopsies", hold great promise for future molecular diagnostics development. Analysis of circulating tumor DNA (ctDNA) for non-invasive detection and monitoring of cancers has been an active area of clinical research. We have used multiple approaches to access genetic information from circulating nucleic acids. We first applied a whole-genome random shotgun sequencing methodology to detect cancer dysploidy patterns in cfDNA. Second, we developed a targeted cfDNA-based drug response assay for melanoma. This assay is a somatic mutation panel developed to assay specific genomic regions with a variety of molecular aberrations such as SNVs, CNVs, and translocations from 1-10 ng cfDNA inputs. Next we focused on DNA methylation profiling which offers additional advantages over somatic mutation analysis for the detection of cancer: (1) higher clinical sensitivity and dynamic range; (2) a multitude of affected methylation targets in disease; (3) multiple affected CpG sites within each targeted genomic region, enabling robust sequencing calls. We developed a targeted methylation panel corresponding to >1,000 differentially methylated genomic regions to measure minimal residual disease (MRD) and recurrence in colorectal cancer. Lastly, we have developed highly sensitive total RNA (including non-coding RNA) sequencing protocols that can be used to analyze RNA/miRNA expression and cDNA/mutation in exosomal RNA or cfRNA isolated from serum/plasma.

3324M

Chromosomal Mosaicism in Patients with Familial Chronic Lymphocytic Leukemia. *L.R. Goldin[1], W. Zhou[2], M. Yeager[2], M. Dean[3], M.J. Machiela[4], M.L. McMaster[1], J. Boland[2], M. Cullen[2], N.E. Caporaso[1], S.J. Chanock[2].* 1) Genetic Epidemiology Branch, DCEG, NCI, Bethesda, MD; 2) Cancer Genomics Research Lab, DCEG, NCI, Bethesda, MD; 3) Laboratory of Experimental Immunology, Center for Cancer Research, NCI; 4) Laboratory of Translational Genomics, DCEG, NCI, Bethesda, MD.

Chronic lymphocytic leukemia (CLL) is a malignancy characterized by the accumulation of small, mature-appearing lymphocytes in the bone marrow, blood, and lymphoid tissues accounting for about one third of the leukemias in US adults. About one third of patients have an indolent disease course. Studies have shown significant familial aggregation of CLL and its precursor trait, monoclonal B-cell lymphocytosis (MBL). CLL tumors are characterized by several cytogenetic abnormalities, the most common being deletions of 13q14, 11q, 17p, 6q, and trisomy 12. We have used Illumina Omni Express arrays to detect these and other abnormalities in DNA derived from blood. Our analytic method can identify abnormalities present in 7% of cells or higher, allowing us to detect abnormalities in DNA from mixtures of tumor and normal cells. We genotyped 130 CLL patients from high-risk families (including some patients with longitudinal samples), as well as 24 first degree relatives with MBL. We use B-allele frequency and log2 intensity ratio to score copy number gains and losses, as well as copy neutral loss of heterozygosity. Mosaic abnormalities were found in 55% of CLL patients, but in none of the relatives with MBL. These abnormalities included those commonly seen in CLL patients as well as more recently reported trisomies of chromosomes 18 and 19. We also found evidence of chromothripsis in some patients. The rate of mosaic abnormalities in familial CLL patients is consistent with that seen in sporadic CLL patients. Among 15 patients studied on two or more time points, changes in the pattern of abnormalities could be seen over time. We are correlating the pattern of abnormalities with disease progression and treatment in this study with more patients, relatives, and time points. Future analyses will be conducted on purified tumor cells compared to normal cells and whole genome sequencing will be performed in selected patients to better define abnormalities. Together, these studies will allow us to identify the earliest genetic changes present in CLL, highlight those associated with progression, and aid in identifying genomic alterations in relatives at increased risk for developing CLL .

3325T

Comprehensive discovery of structural variation in Multiple Myeloma via single molecules. *A. Gupta[1,2], M. Place[2], S. Goldstein[2], K. Potamousis[2], S. Zhou[2], C. Flanagan[3], Y. Li[4], E.H. Bresnick[3], N.S. Callander[3], P. Hematti[3], J. Ma[4], F. Asimakopoulos[3], D.C. Schwartz[1,2].* 1) Biophysics Graduate Program, University of Wisconsin-Madison, Madison, WI, USA; 2) Laboratory for Molecular and Computational Genomics (LMCG), Department of Chemistry, Laboratory of Genetics and Biotechnology Center, University of Wisconsin-Madison, Madison, WI, USA; 3) School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, USA; 4) Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, USA.

Cancer genomes are complex and harbor somatic variation that ranges from single nucleotide changes to numerical chromosomal abnormalities. Recent cancer genome sequencing studies have greatly improved our understanding of cancer genome variation. However, this progress is limited to the ends of the spectrum of variation: Single nucleotide changes on the short length scale and gross copy number changes/aneuploidy on the long length scale. Accordingly, current platforms and analyses do not comprehensively discover mid-sized somatic variation (structural variation; 1kb- megabases), which includes balanced and unbalanced events like insertions, deletions, translocations, inversions and other complex genomic rearrangements. In this work, we use an integrative approach that employs Optical Mapping - a single molecule whole genome analysis system, and DNA sequencing to identify widespread structural variation in a Multiple Myeloma (MM) genome. Furthermore, by studying the longitudinal evolution of this MM genome, we found evidence for increased mutational burden, clonal tides and clonal evolution at the level of structural variation. Our work demonstrates high prevalence of structural variation in a cancer genome, and highlights the importance of routinely studying structural variation in cancer genomes.

3326S

A meta-analysis of somatic copy number alterations in Hepatocellular Carcinoma. *K. Hao[1,2], Z. Zhang[1,2], Y. Hoshida[3], D. Sia[3,4,5], M. Mahajan[1,2], V. Mazzaferro[4], R. Pinyol[5], X. Sun[3], E.E. Schadt[1,2], J.M. Llovet[3,5,6].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Mount Sinai Liver Cancer Program, Division of Liver Diseases, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Gastrointestinal Surgery and Liver Transplantation Unit, National Cancer Institute, Milan, Italy; 5) HCC Translational Research Laboratory, Barcelona-Clinic Liver Cancer Group, Institut d'Investigacions Biomediques August Pi I Sunyer (IDIBAPS), CIBERehd, Liver Unit, Hospital Clinic, University of Barcelona, Catalonia, Spain; 6) Institucio Catalana de Recerca i Estudis Avancats (ICREA), Catalonia, Spain.

Introduction: Hepatocellular carcinoma (HCC) is the second leading cause of cancer mortality worldwide. With the development of high-throughput technology, several studies have surveyed genome-wide somatic copy number alterations (SCNAs) and unveiled a few altered genes as potential drivers. However, more potential driver alterations are yet to be identified. Here we conducted a meta-analysis of SCNAs in 594 resected HCCs with different etiologies. **Methods**: We collected whole-genome SNP array data of 594 HCCs and adjacent normal tissues in three published studies (Lamb et al (2011), n=162; Wang et al (2013), n=286; European Commission Framework Program 7 (HEPTROMIC), n=146). The individual SCNA profiles were quantified using a segmentation-based approach. To compare the results from different studies, the SCNA signals were properly rescaled with the germline copy number variants (CNVs) information detected in adjacent normal tissues. We applied GISTIC2 to the combined dataset as well as to etiology specific datasets. Genes within these regions were compared with existing HCC driver genes and established oncogenes and tumor suppressor genes (TSGs) in other tumors. **Results**: The three cohorts analyzed consisted mostly of East Asian and European patients with HBV (n=388, 65%) or HCV infection (n=89, 15%). Broad gains and losses of chromosome arms detected in our analysis were consistent with previous studies. Among the focal CNVs, we identified focal amplifications at 11q13.3 (FGF19, CCND1), 6p21.1 (VEGFA), 8q24 (MYC), 5p15.33 (TERT), and 7q31.2 (MET), and focal deletions at 9p21.3 (CDKN2A/B), 1p36 (ARID1A), 13q14.2 (RB1), and 10q23.31 (PTEN), which were equally distributed among HBV and HCV-infected patients. Interestingly, focal deletion at 16p13.3 (AXIN1) occurred more frequently in HBV than HCV-infected individuals (111/388, 29% vs 9/89, 10%; p< 0.01). Similarly, focal deletion of the potential TSG PTPRS (19p13.3) were found more frequent in HBV-related patients (79/388, 20% vs 8/89, 9%; p=0.03). Previously unknown focal CNVs were also unveiled, such as focal amplifications at 19p13.11 (JAK3, 5%), 19p13.12 (NOTCH3, 4%) and 9p24.1 (JAK2, 2%), and focal deletion at 8p23.2 (CSMD1, 36%). **Conclusions**: By performing a large meta-analysis, we can estimate more reliably the prevalence of recurrent SCNAs in HCC with different etiologies as a basis to develop therapeutic strategies and identify novel candidate drivers with increased statistical power.

3327M

A t(1;19) translocation involving *TCF3/PBX1* fusion within a context of a hyperdiploid karyotype in adult B-ALL. *B.J. Lasky[1], L. Yeh[2], M. Kallen[2], S. Pullarkat[2], L. Yang[2], A. Hasan[2], Y. Kim[2], D. Shabsovich[2], N. Rao[2], C.A. Tirado[2].* 1) Pathology & Laboratory Medicine, UCLA, San Pedro, CA. 90731; 2) Pathololgy & Laboratory Medicine, Cytogenetics UCLA 1010 Veteran Ave, 2212F, Los Angeles, CA 90024.

Hyperdiploidy is found in 5% of adults with B-Lymphoblastic Leukemia (B-ALL) and portends a favorable prognosis. The t(1;19)(q23;p13) is a common translocation in adult and pediatric B-ALL, often present as a lone abnormality, and is associated with an intermediate prognosis. Herein we present a 44-year-old female with a history of relapsed precursor B-ALL. First diagnosed in March 2013 with leukemic cells showing a normal karyotype, she underwent four cycles of UK ALL 14 protocol consolidation therapy, and although initially found to be in remission, she relapsed in December 2013. A bone marrow biopsy showed 85% blasts with a pre-B immunophenotype and a hyperdiploid, complex, poor-risk karyotype. In January 2014, she underwent FLAG-IDA therapy and the subsequent bone marrow biopsy showed a hypocellular marrow without a significant residual blast population. The patient presented with relapse to our institution in April 2014 for enrollment in a clinical trial with blinatumomab. A bone marrow biopsy showed extensive tumor necrosis with B-lymphoblasts constituting 5% of the total surface area. Touch preparations showed clusters of B-lymphoblasts representing 90% of viable cells. Immunophenotyping was positive for CD10, CD19, PAX-5, CD79a and TdT (weak, rare) and negative for CD34 and CD20. Chromosome analysis demonstrated a female karyotype with numerous numerical and structural abnormalities including extra copies of chromosomes 1, 8, 11, 20, 22, a (1;19) translocation, an unbalanced rearrangement of the long arm of chromosome 13 leading to 13q- and a marker chromosome of unknown origin. The (1;19) translocation is among the most common translocations in B-ALL, resulting in a fusion of *TCF3* at 19p13 with *PBX1* at 1q23, which can occur in a balanced or unbalanced form. The pattern seen here, with a complex karyotype (>3 abnormalities), extra copies of chromosomes 8, 11 and 22 (confirmed by FISH), a (1;19) translocation and the presence of marker chromosomes of unknown origin portends an unfavorable prognosis in B-ALL. This case is one of few that shows both hyperdiploidy and a TCF3/ PBX1 fusion, which underscores the importance of FISH for proper classification of these cases and to distinguish them from those with hyperdiploidy and *TCF3/PBX1* negativity.

3328T

PROMOTER-SPECIFIC ALTERATIONS OF APC ARE A RARE CAUSE FOR MUTATION-NEGATIVE FAMILIAL ADENOMATOUS POLYPOSIS. *T.T. Nieminen[1], W. Pavicic[1], A. Gylling[1], J-P. Pursiheimo[2], A. Laiho[2], A. Gyenesei[2], H.J. Jarvinen[3], P. Peltomaki[1].* 1) Medical Genetics, Haartman Institute, Helsinki, University of Helsinki, Finland; 2) University of Turku, Finland; 3) Helsinki University Central Hospital, Finland.

Familial adenomatous polyposis (FAP) syndrome predisposes to colorectal cancer (CRC). FAP individuals suffer from even hundreds or thousands of colon adenomas and if untreated, CRC develops. The average age at onset of CRC is 39 years. FAP is caused by a germline mutation in the APC gene and it is inherited by an autosomal-dominant manner. In FAP, 20% of classical and 70% of attenuated/atypical (AFAP) cases remain mutation-negative after routine testing; yet, allelic expression imbalance may suggest an APC alteration. Our aim was to determine the share of families attributable to genetic or epigenetic changes in the APC promoter region. We studied 51 unrelated families/cases (26 with classical FAP and 25 with AFAP) with no point mutations in the exons and exon/intron borders and no rearrangements by multiplex ligation-dependent probe amplification (MLPA, P043-B1). Promoter-specific events of APC were addressed by targeted re-sequencing, MLPA (P043-C1), methylation-specific MLPA, and Sanger sequencing of promoter regions. A novel 132-kb deletion encompassing the APC promoter 1B and upstream sequence occurred in a classical FAP family with allele-specific APC expression. No promoter-specific point mutations or hypermethylation were present in any family. In conclusion, promoter-specific alterations are a rare cause for mutation-negative FAP (1/51, 2%). The frequency and clinical correlations of promoter 1B deletions are poorly defined. This investigation provides frequencies of 1/26 (4%) for classical FAP, 0/25 (0%) for AFAP, and 1/7 (14%) for families with allele-specific expression of APC. We will continue investigations with the remaining 6 families with allele-spesific expression of APC by RNA-sequencing. Clinically, promoter 1B deletions may associate with classical FAP without extracolonic manifestations.

3329S

Comparison of CNV detection from whole-exome sequencing and microarray platforms using matched tumor-normal samples from TCGA. *A. O'Hara, L. Culot, Z. Che, R. Keshavan, S. Shams.* BioDiscovery Inc., Hawthorne, CA.

Copy number variants (CNVs) have been implicated as drivers in many different tumor types. The platform of choice to detect genome-wide CNVs has traditionally been microarray (including SNP arrays that can also detect copy neutral loss of heterozygosity regions); however, since samples have often undergone sequencing to discover pathogenic sequence variants, it is desirable to exploit these data to also detect CNVs. Various methods have been proposed to detect CNVs from NGS data, but little information is available which compare the success of these approaches to orthogonal methods. Using a data set of matched tumor-normal samples from the TCGA that have been subjected to both whole-exome sequencing (WES) and to a genome-wide SNP microarray, we compare the ability to detect CNVs between these platforms. CNV detection results from the WES data using ngCGH are compared with an established HMM-based method for detection of CNVs from the microarray platform. We will further discuss benefits, and potential shortfalls of each method.

3330M

A New Method for High Fidelity Copy Number Analysis in Solid Tumor Samples and its implementation in the OncoScan™ FFPE Assay Kit. *J. Schmidt[1], B. Liu[1], M. Ghent[1], B. Bolstad[1], F. Siddiqui[1], D. Abdueva[2], M. Marjanovic[1], R. Saplosky[1], A. Shukla[1], S. Venkatapathy[1], C. Chen[1], C. Bruckner[1], V. Huynh[1], L. Liu[1], K. Suyenaga[1], P. Weaver[1], L. Greenfield[2], E. Fung[1].* 1) R&D, Affymetrix, Santa Clara, CA; 2) Formerly R&D Affymetrix, Santa Clara, CA.

Copy number analysis in tumors is rapidly gaining importance in cancer therapy as a tool for differential diagnosis with impact on potential treatment, [1]. Although formalin fixed paraffin embedded (FFPE) blocks are common sources of material for both research and clinical diagnostics, DNA degradation in FFPE presents a major challenge for accurate measurement of copy number. We address this challenge using molecular inversion probes, which capture the alleles of over 220,000 SNPs at carefully selected genomic locations, evenly distributed across the genome and with increased density within ~900 cancer-related genes. Additionally, we created a reference that establishes the baseline response for a normal copy number state at each locus. This reference was generated by assessing the probe-by-probe response in 400 normal FFPE samples from over 20 sources covering a broad range of geographic locations, block ages, gender and tissue of origin. Another challenge of FFPE samples is the presence of normal cells in most biopsy samples, which affects copy number estimates. To address this problem of variable tumor burden, we developed TuScan™, an algorithm inspired by ASCAT [2], to estimate the integer copy number in the tumor at each locus. When a major clone is responsible for the majority of copy number changes, the algorithm estimates the tumor burden in the sample and reports the integer copy number in the cancer portion only, effectively subtracting the normal component, thereby enabling a comparison between samples with different tumor burden. For highly heterogeneous samples or very low tumor burden, the algorithm reports the fractional (average) copy number of all cells within the sample. To validate the method, 28 FFPE tumors which included copy number events previously determined by FISH were titrated at various percentages (80%, 70%, 60%, 40%) with their matched normal and were analyzed for copy number changes using the TuScan™ algorithm. Given that the algorithm computes copy number states in the tumor portion, the same copy number state should be reported at each locus for each of the titration levels. Over 90% of aberrant markers had concordant copy number states across titration points for 26/28 samples. For two tumor samples the algorithms were not able to determine a tumor burden for the lower titration levels.

[1] G. Ciriello et al, Nat. Gen. 45, 1127-1133, 2013. [2] Peter Van Loo et al, PNAS 107 (39) 16910-16915, 2010.

3331T

The dilution dilemma; a method to accurately estimate tumor fractions in complex tumor/normal DNA dilutions. *Z.M. Weber[1], S. Vattathil[2], Y. Hu[1], P. Scheet[2], G.E. Davies[1], E.A. Ehli[1].* 1) Avera Institute for Human Genetics, Sioux Falls, SD., USA; 2) The University of Texas M.D. Anderson Cancer Center, Houston, TX., USA.

The objective of this project was to create a series of highly accurate dilutions of matched tumor/normal DNA to test computational methods for detecting low-frequency chromosomal aberrations, such as allelic imbalance and copy number changes, in a sample of DNA obtained from heterogeneous cells. Such methods (e.g. hapLOH, GAP, BAFsegmentation, ASCAT) offer the potential to determine tumor cell fractions from surgically removed tumor sections that contain normal tissue margins and to identify tumor sub-populations in samples that possess heterogeneous genetic profiles. Qubit is a fluorescence-based method used to quantify double stranded DNA. Using this method of quantitation, we created 32 tumor/normal mixes (16 dilutions in 2 cancer cell lines) with tumor fractions ranging from 0.98 to 0.005. Dilutions containing ultra-low fractions of tumor DNA required a method of highly accurate quantitation. Breast cancer (CRL-2363) and lung cancer (HTB-172) tumor cells were grown up in culture with their respective matched normal cell line lymphocytes, CRL-2343 and CRL-5958, respectively. DNA was extracted from the four cell lines using the QIAamp DNA Blood Midi protocol. The DNA from each cell line pair was normalized and quantitated in five replicates using Qubit High Sensitivity assays. The standard deviation (SD) and coefficient of variation (CV) between the five replicates was calculated for each DNA sample. Next, a series of tumor DNA fractions was prepared for both the breast and lung cancer DNA sets. Two sets of 16 dilutions were run on Infinium HumanOmni2.5-8 Beadchips. The SNP genotype data was used to generate Log R ratios and B-allele frequency plots used for downstream analyses with hapLOH (for example). Low SD values were successfully achieved between replicate sets (0.69 - 1.45 ng/ul) along with low CV values (0.013 - 0.028). Using a mathematical algorithm which incorporated the volume of DNA used in the dilution and the mean concentrations, we were able to estimate the amount of tumor DNA in the complex mixture. The mean GenCall score across all samples was p10 GC = 0.41 and p50 GC = 0.78 following the genotyping experiment. Call rates were all > 99% and error rates between the three tumor fractions at 0.03 were all < 0.00001. The primary scientific contribution of this experiment was creating a data set on which to test methods to accurately identify mutations in extreme stromal contamination and estimate tumor/normal DNA fractions.

3332S

Age-related mosaic loss of chromosome Y is associated with cancer in cohort studies. *W. Zhou[1], M. Machiela[2], M. Dean[2], J. Sampson[2], N. Rothman[2], N. Freedman[2], S. Wacholder[2], K. Jacobs[1,5], M. Tucker[2], F. Real[3,6], L. Perez-Jurado[3,4], M. Yeager[1], S. Chanock[2].* 1) Cancer Genomics Research Laboratory, National Cancer Institute, Division of Cancer Epidemiology and Genetics, Leidos Biomedical Research Inc., Frederick, MD; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD; 3) Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain; 4) Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona, Spain; 5) Bioinformed, LLC, Gaithersburg, MD; 6) Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre (CNIO), Madrid 28029, Spain.

Prior studies indicate that loss of the Y chromosome occurs during the aging process and is associated positively with hematologic cancer, including acute myelogenous leukemia (AML), myelodysplastic syndrome (MDS), and pre-leukemia. To investigate the relationship between mosaic loss of chromosome Y and non-hematologic cancer, we utilized existing genome-wide association study intensity data from 14,099 males (8,545 non-hematologic cancer cases and 5,554 cancer-free controls) drawn from three prospective cohort studies: Alpha-Tocopherol and Beta-Carotene Cancer Prevention Study (ATBC), Prostate, Lung, Colorectal, Ovarian Cancer Screening Trial (PLCO), and Cancer Prevention Study-II (CPSII). Subjects were examined for deviations from expected log2 intensity ratio for evidence of loss of the male specific region of chromosome Y (MSY) between 6,671,498-22,919,969 (hg18/build36). DNA derived from blood or buccal samples was genotyped on one of four Illumina SNP BeadChips with sufficient coverage of MSY (Human610, Human1M, HumanOmni1M, and HumanOmni2.5M). A total of 1,000 (7.09%) individuals have detectable mosaicism of the MSY with the proportion of affected cells ranging from 9.9% to 61.7%; this is substantially higher than what has been observed for autosomal events of size >=2MB, approximately 1% overall. Mosaic chromosomal Y loss is associated with older age at DNA collection, with frequencies of 1.09% in individuals less than 60 and 18.63% in those 80 years or older (OR per year of age = 1.11, 95% CI = 1.10-1.13; P = 5.34E-69) and ever-smoking (OR = 1.43, 95% CI = 1.21-1.71; P = 8.16E-05). In addition, mosaic chromosomal Y loss is more common in individuals developing cancer (8.05%) than in cancer-free controls (5.37%) (OR = 1.57, 95% CI = 1.36-1.81; P = 1.02E-10). Mosaic Y loss was associated with bladder cancer (OR = 1.47, 95% CI = 1.19-1.84; P = 7.15E-04) and prostate cancer (OR = 1.29, 95% CI = 1.08-1.56; P = 7.52E-03) but not lung cancer. Further studies are required to determine the role of mosaic Y loss in cancer.

3333M

The actual impact of Fluorescence in Situ Hybridization (FISH) in the diagnosis and follow up of Acute Lymphoblastic Leukemia (ALL). *H. Akin, E. Karaca, B. Durmaz, A. Durmaz, A. Aykut, O. Cogulu.* Ege University Medical Faculty, The Department of Medical Genetics, IZMIR, Turkey.

Acute lymphoblastic leukemia (ALL) is a disease characterized by abnormal clonal proliferation of lymphoid precursors which lose their ability to differentiate. It is the most common malignancy in childhood, accounting for 80% of all leukemia. Although it can occur at any age, its incidence is highest among children from 2 to 5 years at rate of about 70%, decreasing among adolescents and young adults. Cytogenetics and FISH have been the gold standard for defining genetic abnormalities and facilitating therapeutic stratification of pediatric ALL cases. In this study, it is aimed to determine the cytogenetic and molecular cytogenetic profile of submitting patients with ALL during the period from January 1st, 2013 to December 31st, 2013 to Ege University Medical Faculty, Medical Genetics Department. In ALL, the most important molecular markers are BCR/ABL, all MLL translocations with their respective partner genes, and c-MYC aberrations that have to be detected at diagnosis and for minimal residual disease (MRD) studies . Furthermore, for more than 90% of ALL patients specific markers can be detected by IGH mutations and serve as the most important MRD marker. We used a standardized a FISH panel for ALL including, BCR/ABL, cMYC, TCF/PBX, IGH, MLL, p16 and TEL/AML. We performed 150 FISH analyses for 77 patients .The most frequent aberration rate was 10.67% for IGH. The second most frequent aberration was detected for delp16 as 5.33%. They were all commented as diagnosis and follow up markers for ALL. The vast majority of this study group are of pediatric age group (85.33%). The repetitive analyses by the aim of follow up and cytogenetic results were also evaluated. It was noted a substantially lack of an algorithm at diagnosis and follow up request for ALL patients. A more reliable approach would be to standardize wider FISH panels in order to detect more anomalies and using an algorithm to reach effective diagnose and follow up in ALL patients.

3334T

System for high throughput Identification of breast and ovarian cancer associated chromosomal abnormality. *Y.W. Chang[1], J. Chen[5], N. Limthong[4], Y.Y. Chang[1], D. Chen[3, 4], Y. Elshimali[2].* 1) La Sierra University, Riverside, CA; 2) Charles R. Drew University/UCLA, Los Angeles, CA; 3) Dept. of Path & Lab Med, David Geffen School of Med at UCLA, Los Angeles, CA; 4) Dept. of Path & Lab Med, UCI Med Center, Orange, CA; 5) Rosemead Clinic, Rosemead, CA.

Investigation of biochemical pathways of breast and ovarian cancer has greatly increased our understanding of the pathogenesis of breast and ovarian cancer. 48 breast cancer susceptible genes have been identified in the literature. These genes are potential candidates to expand the genetic testing panel for both the familial and somatic forms of breast and ovarian cancers. One challenge for current genetic testing technology is the relative inefficiency in identifying complex chromosomal structure abnormalities. Non-allelic homologous recombination (NAHR) and non-homologous end joining (NHEJ) are major mechanisms involved in such chromosomal changes. Since the locations for potential chromosomal changes are non-random and limited by the genomic location of homologous sequence pairs, a targeted DNA sequencing approach for chromosomal abnormalities in these candidate genes can significantly facilitate the depth of reads and thereby increase sensitivity at detecting chromosomal changes. Specifically, we conducted a computational analysis of the genomic sequences around all 48 potential breast and ovarian cancer susceptible genes and globally identified the potential target pairs for NAHR and NHEJ events. Over 700 potential genomic sequence pairs were identified with greater than 200bps length and 85% sequence homology. We designed a solution phase genomic DNA capture system for capturing genomic DNA of tumor specimens. We then use NGS technology to sequence the genomic fragments. Using normal genomic DNA for optimization, we have conducted deep sequencing with the goal of 50x depth. Overall, 94% of genomic areas are tagged as “readable sequence”. In addition, 86% of which achieve sequencing depth of greater than 50X. In conclusion, the system has the potential to serve as screening system for breast and ovarian specific chromosomal changes.

3335S

Unique Recurrent Cytogenetic Aberrations Distinguish Between Molecular Subtypes of DLBCL and Burkitt Lymphoma: An Analysis of Unsupervised Clusters and Logistic Regression Based Models. *R. Garcia, N. Guruju, P. Koduru.* Pathology, UT Southwestern, Dallas, TX.

Morphologic classification of non-Hodgkin lymphoma is reliable in majority of cases. Nonetheless, overlapping features between diffuse large B-cell lymphoma (DLBCL) and Burkitt lymphoma (BL) with MYC rearrangements are persistent and problematic to distinguish them. The aim of this study was to build and test predictor models based on recurrent cytogenetic aberrations (RCAs) that may help distinguish between DLBCL and BL. This study involved the analysis of available information from the Mitelman Database, other publications and institutional cases. Two separate cytogenetic datasets were used to build and test hierarchical clusters (HCs) and logistic regression (LR) predictor models. The first dataset comprised of 254 DLBCL and 84 BL tumors that were used to identify RCAs. The test dataset comprised of 117 cases of DLBCL and 60 BL obtained from the published literature and institutional sources; this was used to test predictability of the RCAs identified in the first dataset. An independent group t-test, performed with the Statistical Analysis Software (SAS), was used to assess the differences in the frequency of aberrations between comparative groups. A Fisher exact test was then used to determine correlations between the different types of RCAs and comparative groups. A p-value less than .05 was considered significant. Specificity of RCAs was also determined. The R package was used to build HC models and SAS was used for LR analysis. We identified 21 RCAs significantly associated with DLBCL than with BL. These included +12, +3, and +18 previously reported as associated with DLBCL than BL. HC and LR analysis of the second dataset was performed to determine the predictability of the remaining RCAs. This analysis identified RCAs +16, +2, -8, -2, 14qL, +X, 16qL, 1p36L, 15qL, 17pL, -4, 19pL and 22qL as predictors of DLBCL by their closeness in Euclidian distance to BCL2, +12, +3 and +18 and by their distance away from MYC . Further evaluation of the data revealed two major clusters one associated with GCB and the other with the ABC DLBCL molecular subtypes. Predictive value of the models was 95% for the HC heat map, 96% for the PAM algorithm, 85% for the LR and 98% for the HC with p-values. Our findings indicate HC with p-values based on Euclidean distance is a better predictive model. Future studies correlating these RCAs with molecular signatures such as gene expression profiles are warranted. Key words: DLBCL, BL, RCAs.

3336M

A rare transformation case report: from Chronic Lymphocytic Leukemia to Multiple Myeloma. *C. Hangul[1], C. Aydin[1], O.K. Yucel[2], O. Salim[2], S. Berker Karauzum[1].* 1) Department of Medical Biology and Genetics, School of Medicine, Akdeniz University, Antalya, Turkey; 2) Department of Hematology, School of Medicine, Akdeniz University, Antalya, Turkey.

Chronic Lymphocytic Leukemia (CLL) and Multiple Myeloma(MM) are closely related B-cell malignancies. The genetic characterization of CLL has made significant progress in the recent years. Genetic aberrations are detected in over 80% CLL cases with FISH methods. In the other hand, 30-50% cases of multiple myeloma shows random genetic rearrangements. Transformation from CLL to MM is seen in very rare cases. In this report, we present 62 years old patient who was first diagnosed with Chronic Lymphoid Leukemia in 2009. Marked lymphoid population (CD5+, CD19+, CD23+) was detected by flow cytometric immunophenotyping in peripheral blood and bone marrow. In the first diagnosis in 2009, his karyotype revealed from bone marrow sample was designed as 45, X, -Y[4]/46,XY[10] and in FISH analysis, an 11q22.3 deletion was found of 55% in this case as a sole anomaly. At the second biopsy, cytoplasmic kappa light chain positivity was seen. Lambda and CD20 staining were found negative in 2012. Bone marrow immunophenotyping revealed 30% of clonal plasma cells (CD38+, CD56+) and 0.9% of abnormal B Lymphocytes (CD5+,CD23+,CD19+) with these findings the patient was diagnosed as Multiple Myeloma. In the second bone marrow sample, his karyotype was found to be: 46,XY,der (6)t(1;6)(q11;q23),t(11;14)(q13;q32), dup(17)(q23q25)[17]/46,XY[2]. Additionally two hundred nuclei were analysed for each probe from MM FISH panel and t(11;14) (q13.1;q32) translocation and deletion of 6q23 was found of 48% in interphase nuclei. All these results supported the diagnosis of MM and by the same time to transformation from CLL to MM. Clinical and labaratory findings will be discussed in detail with the company of literature.

3337T

Specific Gene Expression Profiles of Diffuse Large B-cell lymphoma and Burkitt Lymphoma Identified a Set of Enriched Genes that Positively Correlates with Cytogenetics Data. *P. Koduru, R. Garcia, N. Guruju.* Dept Pathology, Univ Texas Southwestern Med Ctr, Dallas, TX.

In 2008, the World Health Organization created a provisional category for Unclassifiable B-cell lymphoma with features intermediate of diffuse large B-cell lymphoma (DLBCL) and Burkitt lymphoma (BL) mainly due to overlapping morphological features. The purpose of this study was to identify differences between DLBCL and BL at the molecular level and to correlate these differences with the recurrent cytogenetic abnormalities (RCAs) between them identified by us previously. The data for this analysis consisted of 44 lymphoma tumors characterized by HGU 133A Affimatrix gene chip, and extracted from the National Center of Biotechnology Information (NCBI) Gene Expression Omnibus database (GSE4475 dataset). These 44 tumors included 21 DLBCL and 23 BL. The statistical package 'R-package' was used to analyze the expression data. Samples were normalized and differentially expressed genes were determined by a p-value <.01 with a false discovery rate adjustment applied to the p-value in order to decrease false positives. To further characterize the gene expression in both tumor types, a gene set enrichment analysis was performed using the GSEA software package. The identified differentially expressed genes were uploaded onto Ingenuity program to gain insight into the molecular interactions and biological processes in these lymphomas. In addition, a list of RCAs and copy number alterations previously identified by us was correlated with gene expression data. This study revealed differentially expressed genes in DLBCL and BL previously reported in the literature, mainly MYC, ID3, SOX11 and CSE1L. Moreover, our analysis revealed 180 and 3 upregulated genes in DLBCL and BL respectively, and 137 and 29 down regulated genes for DLBCL and BL respectively. This data closely correlated with recurring cytogenetic abnormalities (+X, +2, -2, +12, 14qL, +16, 16qL, +3, +18, 1p36L, 15qL, 17pL, 19pL, -4 and 22qL) we identified as markers to distinguish these tumors. In addition we identified a number of genes not previously reported in the literature in association with DLBCL. These included: CDK5R1, SAMSN1, FNBP1 and ELK3. Expression profiling data from additional tumors are needed to validate these findings. In brief, our analysis identified molecular expression markers that may be used to distinguish DLBCL and BL, and these molecular markers correlated with RCAs identified by us in a previous study. Key words: Gene Expression, DLBCL, BL, RCAs.

3338S

The Clinical Utility of CpG-Oligodeoxynucleotide Stimulation in Chromosomal Analysis for patients with Chronic Lymphocytic Leukemia (CLL) and a possible secondary Myeloid Neoplasm. *C.A. Marcou, P.T. Greipp, T.A. Goble, D.L. Van Dyke.* Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN., USA.

CLL is the most common leukemia affecting adults in the western world. Detection of cytogenetic abnormalities in CLL is useful in the diagnosis, prognosis, and therapeutic guidance for this disease. However, the detection of these abnormalities proves difficult due to the low mitotic rate of mature B cells. Interphase FISH analysis has been the standard test to detect 11q, 13q, and 17p deletions, trisomy 12, and *IGH* rearrangements, but has the limitation that cytogenetic complexity and other recurrent CLL-associated changes are not detected. To overcome these limitations, CpG-oligodeoxynucleotide (CpG) stimulation has been widely adopted for chromosome studies of CLL. Our cytogenetics laboratory receives blood & bone marrow samples referred from Mayo Clinic-affiliated sites and other clients via Mayo Medical Laboratories. In 2013, we began using CpG-stimulation in addition to unstimulated cell cultures to identify either a myeloid or lymphoid neoplasm in samples referred for chromosome analysis with a "reason for referral" that included lymphocytosis, CLL, small lymphocytic lymphoma, chronic lymphoproliferative disease, or a history of a CLL FISH study. Many and perhaps most of these samples were from patients with CLL or other lymphoproliferative disease who presented with a new cytopenia. Presumably, the principle concern was to identify a cytogenetically abnormal myeloid neoplasm if present; however, in most cases we did not obtain a final pathologic diagnosis. We performed a retrospective analysis identifying 568 cases that fit the criteria above. Of these, 253 (44.5%) had a normal karyotype in the CpG-stimulated and unstimulated cultures. The remaining 315 cases had an abnormal karyotype; of these, 40 (13%) had no analyzable metaphase cells in the unstimulated cell cultures. An abnormal karyotype interpreted as representing a myeloid neoplasm was present in 14/568 cases (2.5%). Among the 301/568 cases (53%) representing a lymphoid neoplasm, 97 exhibited trisomy 12, 87 had a 13q-, 78 had an 11q-, 63 had a 17p-involving TP53 loss, and 42 had a 14q deletion or rearrangement. In addition, 23 cases exhibited translocations between *IGH* and *MYC*, *BCL2*, *BCL3*, or *CCND1*. It is clear that CpG-stimulation is useful to identify the cytogenetically abnormal lymphoproliferative clone, and along with unstimulated cell cultures provides dual surveillance of both lymphoid and emerging secondary myeloid neoplasms in patients with lymphoproliferative disease.

3339M

Chromosomal Aberrations associated with methylation in ependymomas in Mexican pediatric patients. *M. Pérez Ramírez[1,5], A. Hernández[1,5], A.J. Hernández[3], A. Guerrero[3], A. García[3], F. Chico[4], E. Benadón[4], I. Del Angel[3], G. Ballesteros[3], M. Pérezpeña[4], A.G. Siordia[2], A. Sánchez[4], D.J. Arenas[1], F.A. Salamanca[1], N. García Hernández[1].* 1) Unidad de Investigación Médica en Genética HuUnidad de Investigación Médica en Genética Humanamana, Centro Médico Nacional "Siglo XXI", IMSS, Mexico DF, Mexico; 2) Servicio de Patología, Hospital de Pediatría "Dr. Silvestre Frenk Freud", Centro Médico Nacional "Siglo XXI", IMSS; 3) Servicio de Neurocirugía Pediátrica, Hospital General "Dr. Gaudencio González Garza", Centro Médico Nacional "La Raza", IMSS; 4) Hospital Infantil de México "Federico Gómez; 5) Universidad Nacional Autónoma de México.

Ependymoma (EP) is a slow-growing tumor, originates from the ependymal cell lines of the ventricular system, plexus choroid and central canal, located in regions within the Central Nervous System, principally developed at the fourth ventricle and less frequently in the brain parenchyma as a result of migration of ependymal cells during embryogenesis. It is known that the EP can proceed of cells derived from the transformation of neural stem cells and their progenitors to contribute to the initiation and the development of brain tumors. In recent studies in brain tumors, it has been found that the methylation affects genes responsible for DNA repair, regulation, migration and cellular apoptosis. Methylation is associated with the development of cellular malignancy by oncogene activation, promotion of genomic instability and loss of genomic imprinting. Additionally, the chromosomal aberrations are important for the development of EP, as amplifications in 7p, 9q, 15q and losses in 14q, 6q; it is known that changes in gene expression in oncogenic processes are responsible for large cell changes; so it is interesting to analyze the methylation pattern to relate with the chromosomal aberration in ependymomas. Ependymal tumors were collected surgically in pediatric patients, genomic DNA was extracted by organic method of phenol-chloroform-isoamyl, subsequently were performed microarrays methylation and aCGH with the platform of Agilent microarrays following the manufacturer's recommended protocol. In this work was found amplification in 14q32.3 (100%), 2p11.2 (83%), 8p22 (83%), and losses in 11q11 (83%), 15q11.2 (66%), 16q12.2 (66%), with respect to the methylation pattern, found hypomethylated regions in 2p11.2, 8p22, whereas the region 14q32.33 showed hypo and hypermethylated of genes related to growth and neuronal development, apoptosis, transcription factors and system immune. It is suggested to the region 14q32.33 as alteration of great importance for the Mexican population, because of the genes located in this chromosomal region.

3340T

Toward rapid identification of coding fusions and structural rearrangements in cancer genomes: Multiple Myeloma First. *M. Rossi[1,2], H. Dai[4], H. Cao[4], M. Saghibini[4], X. Yang[4], A. Pang[4], A. Hastie[4], T. Dickinson[4], J. Hauenstein[2], C. Curnow[2], A. Nooka[3], J. Kaufman[3], S. Matulis[3], L. Zhang[2], D. Saxe[2], K. Mann[2], D. Jaye[2], L. Boise[3], S. Lonial[3].* 1) Radiation Oncology, Winship Cancer Institute of Emory University, Atlanta, GA; 2) Pathology and Laboratory Medicine, Emory University, Atlanta, GA; 3) Hematology and Medical Oncology, Emory University, Atlanta, GA; 4) Bionano Genomics, San Diego, CA.

The use of G-banded karyotype and FISH have been standard diagnostic tools in monitoring response to treatment and disease progression in hematological disorders, including multiple myeloma, for over 40 years. However, with the availability of array and NGS technologies in most clinical diagnostic laboratory settings, it is time to consider evaluated the use of more modern methods in diagnosing plasma cell dyscrasias. Although we routinely run RNA-Seq, SNP-CN arrays and a deep sequencing cancer panel as a cost effective workflow for clinical trials, we find the validation of translocations and structural rearrangement by PCR and/or FISH cumbersome. Moreover, because of the limited applicability of using whole genome sequencing as a clinical diagnostic tool for multiple myeloma, we remain confined to the use of conventional cytogenetics for primary risk stratification of all patients. More recently, we have identified the BioNano Genomics Irys System which utilizes nanochannel technology and high resolution imaging for whole genome mapping of translocations and copy number abnormalities. As proof of concept, we have used the Irys system to analyze the multiple myeloma cell line, KMS11, and will demonstrate the ultility, speed and accuracy in detecting structural rearrangements in this line compared to whole genome, RNA-seq and SNP-CN datasets. We will also demonstrate the use of the system with CD138+ enriched bone marrow samples from consented patients currently on clinical trials compared to FISH, G-band karyotype and SNP-CN arrays. We are confident that these data demonstrate that the Irys ststem is a disruptive innovation with broad applicability to genome research and refinement of normal variation and disease.

3341S

Conventional Cytogenetic and Molecular Genetic Analysis of Pancreatic Cancer. *D. Shabsovich, N. Rao, J. Ji, K. Park, L. Yang, C.A. Tirado.* Pathology and Laboratory Medicine, UCLA, Los Angeles, CA.

Pancreatic cancer is the fourth leading cause of cancer-related death in the United States, with an average 5-year survival rate of less than 5 percent. Cytogenetic analysis has allowed for the elucidation of clinically significant chromosomal abnormalities in both solid tumor and hematological malignancies, but pancreatic cancer still does not have any FDA-approved cytogenetic assays despite its known intratumor cytogenetic heterogeneity. In the present study, conventional cytogenetic and molecular genetic analysis was performed on the following 9 pancreatic cancer cell lines: AsPC-1, MIA PaCa-2, BxPC-3, Capan-2, CFPAC-1, HPAF-II, Panc-3.27, Panc-10.05, and Panc-2.13. Giemsa-banding (G-banding), targeted fluorescence in situ hybridization (FISH), and chromosomal microarray (CMA) analysis was completed for each cell line, and results were pooled to elucidate commonly aberrant regions identified by all three methods. Recurrent numerical and structural abnormalities were observed on chromosomes 7, 8, 9, 18, and 20, and these findings were further corroborated by interphase and multicolor FISH studies. Abnormalities included deletion of *SMAD4* (18q21.2), which has been associated with increased metastasis and consequent poor prognosis in pancreatic cancer, deletion of *DCC* (18q21.1), and deletion of *p16* (9p21). Ultimately, phenotypic correlation of cytogenetic abnormalities using a larger population of patients can provide a framework for the development of diagnostic and prognostic cytogenetic assays as well as targeted therapies to combat pancreatic malignancies.

3342M

The *ZNF384* gene in pediatric acute lymphoblastic leukemia - multiple partner genes, immature (CD10 negative) immunophenotype, and potential good outcome. *M. Shago[1,3], O. Abla[2,4], J. Hitzler[2,4], S. Weitzman[2,4], M. Abdelhaleem[1,3].* 1) Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Paediatrics, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; 4) Department of Pediatrics, University of Toronto, Toronto, Ontario, Canada.

The *ZNF384* (*CIZ*) gene, located distal to the *ETV6* (*TEL*) gene at 12p13.31, is a putative zinc finger transcription factor which has been reported to be rearranged in acute leukemia. Rearrangement of the *ZNF384* gene results in attachment of various 5' partner gene sequences to form *ZNF384* fusion genes. The *ZNF384* gene has three known partners: *TAF15* at 17q12 (19 cases), *EWSR1* at 22q12 (4 cases), and *TCF3* (E2A) at 19p13 (3 cases). The proteins encoded by the fusion genes consist of the transactivation domains of the 5' partner genes fused to the entire coding region of ZNF384. We present detailed karyotypic, genomic copy number, and gene expression analyses of seven new pediatric pre-B ALL patients with *ZNF384* gene rearrangement. Our patients had lymphoblasts with a CD10-negative immunophenotype, similar to the immunophenotypic profile seen in *MLL* gene-rearranged ALLs. Follow up on the patients ranges from 5 to 7 years, and none of the patients have relapsed. Identification of the rearrangements was facilitated using dual-colour breakapart probes for the *ZNF384*, *TCF3*, and *EWSR1* loci. Four of the patients had *TCF3-ZNF384* gene rearrangement and one had *EWSR1- ZNF384* gene rearrangement. Two patients had novel *ZNF384* gene translocations involving regions on chromosomes 6 and 22, identified as the *ARID1B* and *EP300* genes respectively by RNA sequencing analysis. Similar to the previously identified *ZNF384* partners, these novel partner genes, which function in chromatin remodeling, contribute 5' gene sequences to the *ZNF384* gene fusion. Based on the total number of ALL patients seen at our centre during the study period, our data suggests that, collectively, *ZNF384* gene rearrangement may have an incidence of ~3% in pediatric pre-B ALL, with an incidence of at least 18% in CD10-negative pre-B ALL. *ZNF384* gene rearrangement may be associated with a more favorable prognosis than *MLL* gene rearrangement. Both good and poor outcomes have been proposed in the literature, however, the majority of the data are based on *TAF15-ZNF384* rearrangement in adult patients. The *ZNF384* gene rearrangements described above are cryptic and not easily identified by G-banding. Inclusion of *ZNF384* FISH analysis in patients with CD10-negative ALL, after *MLL* gene rearrangement has been excluded, will assist with the determination of the prognostic significance of *ZNF384* gene rearrangement in pediatric ALL.

3343T

A complex karyotype with a cryptic t(11;14)(q13;q32) in a Blastoid Crisis of Mantle Cell Lymphoma. *C.A. Tirado, Y. Kim, L. Yang, G. Abusaba, B.J. Lasky, N. Rao.* Pathology & Laboratory Medicine, UCLA, 669 Levering Ave, apart 15 Los Angeles, CA 90024.

Mantle cell lymphoma (MCL) comprises 3-10% of all NHL. The t(11;14)(q13;q32) is present in virtually all cases of MCL, independent of their morphologic or clinical presentation. Herein we describe an 80-year-old female patient with splenomegaly, thrombocytopenia, elevated absolute lymphocyte count, and anemia. A bone marrow biopsy demonstrated a hypercellular marrow for age showing 50% cellularity with 30-40% of the total cellularity involved by a mature B-cell neoplasm with immunohistochemistry showing co-expression of CD20 and BCL-1 in the aggregates of small B-cells. Flow cytometry of the bone marrow aspirate revealed a monotypic, kappa-restricted B-cell population representing 37% of total cells, positive for CD5, CD19, CD20, CD22, FMC-7, and CD38; and negative for CD23 and CD10. These results were consistent with a diagnosis of mantle cell lymphoma. Conventional cytogenetics performed on the bone marrow showed a complex abnormal female karyotype with numerical and structural abnormalities but no evidence of a (11;14) translocation in 6/20 metaphase cells examined. These results were described as: 44-45,XX,del(6)(q25),add(8)(p21),add(10)(p15),add(11)(q13);add(11)(q13), -14, tas(12;21)(pter;pter),+mar[cp6]/46,XX[14]. However, FISH on interphase nuclei using the CCND1-IGH (CCND1-IGH@) dual-color, dual-fusion from Abbott showed evidence of *CCND1-IGH@* fusion signals in 23% of the interphase cells studied. These findings were described as: nuc ish (CCND1,IGH@)x3(CCND1 con IGH@x2)[46/200]. FISH on previously G-banded metaphases showed a *CCND1-IGH@* fusion signal on the derivative chromosome 10 which was cryptic and not identifiable by conventional cytogenetics and another fusion signal on one of the abnormal copies of chromosome 11. Therefore the karyotype was reinterpreted as: 44-45,XX,del(6)(q25),add(8)(p21),der(10)(11q25->11q13::14q32->14pter::10p15->10qter),der(11)t(11;14) (q13;q32),add(11)(q13),14,tas(12;21)(pter;pter), +mar[cp6].ish der(10)(p15)(CCDN1+,IGH+),der(11)t(11;14)(IGH+,CCND1+)/46,XX[14]. This particular pattern with a complex karyotype with three or more numerical and structural changes along with 6q-, t(11;14) and a marker chromosome are common findings in mantle cell lymphoma of the blastoid subtype and are usually associated with an unfavorable prognosis. This case also highlights the importance of FISH on previously G-banded metaphases to elucidate cryptic rearrangements.

3344S

Identification of Semi-Cryptic and Variant Translocation Partners of RUNX1 gene in Acute Myeloid Leukemia (AML). *A. Yenamandra[1], F.C. Wheeler[1], A. Hollis[1], D. Zalepa[1], M. Kapp[1], N.C. Richardson[2], S.C. Borinstein[2].* 1) Department of Pathology, Immunology and Microbiology, Vanderbilt Univ. Medical Center, Nashville, TN; 2) Dept. Pediatrics, Division of Pediatric Hematology/Oncology, Vanderbilt University School of Medicine, Nashville, TN.

Rearrangements of RUNX1 locus at 21q22 include translocation of t(8;21), t(12;21), t(1;21), t(3;21), t(5;21), t(17;21) and t(16;21) in myeloid and lymphoid malignancy. We report here three cases of AML with variant translocation partners of RUNX1; a pediatric AML with t(7;21), and two adult secondary AML cases with t(10;21) and t(16;21) respectively. Sequential metaphase FISH and Cytogenetics were performed in all cases to identify the partner chromosome in the RUNX1 rearrangements. Case #1. A 9 year old male was referred for fever, fatigue, and pancytopenia. Bone marrow analysis revealed hypercellularity with a mixture of myoblasts and immature monocytic elements in 90% of cells as determined by both morphology and flow cytometry. Molecular analysis for mutations in FLT3, NMP1 and C-KIT were negative, although FISH revealed rearrangement of RUNX1 locus in the chromosome 7p22 region. Karyotype analysis confirmed this finding, revealing a unique translocation: 46,XY,t(7;21)(p22;q22)[20]. Three adults and one child have been reported in the literature with AML or high grade MDS with t(7;21) involving RUNX1-ubiquitin-specific protease gene (USP42-)fusion, however, to date no consistent prognostic information has emerged from these cases. Our patient had reinduction chemotherapy and is currently in remission, awaiting stem cell transplant. Case #2. A 58 year old male was diagnosed with treatment related AML following urothelial cell carcinoma. In this case, cytogenetics was normal (46,XY) at diagnosis, but follow up bone marrow analysis four and eleven months later revealed progressive abnormalities (46,XY,del(9)(q13q22)[2]/46,XY[18] and t(10;21)(p13;q22)[12]/46,XY[9], respectively). The latter BMA was FLT3+ also demonstrated RUNX1 rearrangement. The gene at the 10p13 region is unknown. Case #3. A 73 year old female was diagnosed with a myeloproliferative disorder. Karyotype revealed 46,XX,t(16;21)(q24;q22)[16]/46,XX[4] with FLT3 positivity. FISH analysis revealed rearrangement of RUNX1 with MTG16 gene locus at 16q24. The majority of the cases with t(16;21) (MTG16; RUNX1) translocation have been reported in treatment occur in the context of MDS/ANLL and the prognosis is poor. We present here three cases with semi-cryptic RUNX1 translocation in pediatric primary and adult secondary AML with unique RUNX1 rearrangement indicating the importance of the regulatory role of RUNX1 gene in hematopoiesis.

3345M

Somatic gene fusions in human cancer revealed by whole exome sequencing. *L. Yang, P. Park.* CBMI, Harvard Medical School, Boston, MA.

A small fraction of somatic genomic rearrangement breakpoints in human cancer locate in exons or near exon-intron boundaries, and thus can be identified by paired end reads from the whole exome sequencing (WES) data. The main advantage of using WES data to detect somatic rearrangements is the larger sample size (at least 10 times more for The Cancer Genome Atlas [TCGA] samples) compared to whole genome sequencing (WGS) data. We first used 120 TCGA patients that have both WES and WGS to test sensitivity and specificity and to design additional filters in order to remove WES specific artifacts. Using Meerkat, an algorithm we previously developed to detect genomic rearrangements, we found about 1% of the somatic rearrangements detected by WGS can be recovered in WES data and about 50% of the somatic rearrangements detected by WES present in WGS. Some WES specific somatic rearrangements, which were of low allele fraction and thus probably sub-clonal events, weren't detected by WGS due to higher sequencing coverage of WES data. Over 9,000 high confidence somatic rearrangements were identified from TCGA WES data in a total of 4,600 human cancer patients across 15 tumor types. The somatic rearrangement breakpoints occur preferentially in highly express genes. Hundreds of novel gene fusions were identified in addition to known cancer driver fusions such as RET fusions in thyroid cancer, FGFR3-TACC3 fusion in GBM and TMPRSS2-ERG fusion in prostate cancer, etc. RNA-seq data were used to identify activating fusions with following criteria: the 3' fusion partners are up-regulated and the expression changes in at least one of the fusion breakpoints in exon specific expression profile. We identified a total of 150 activating gene fusions. The Gene Ontology (GO) term enrichment analysis for activating fusions revealed that the 5' fusion partners are often house-keeping genes while the 3' fusions partners are enriched in tyrosine kinase and chromatin regulators. We also identified 200 patients (4%) with massive rearranged chromosomes (chromosomes significantly over-represented with somatic rearrangement breakpoints) using the binomial model with Bonferroni correction. The massive rearrangements are often associated with up-regulation of oncogenes. For example, chr17 is often massively rearranged in breast cancer which is associated with up-regulation of ERBB2. Such association can only be revealed by analyzing thousands of patients.

3346T

A Personalized Genomic Signature of Lung to Brain Metastasis. *J.M. Furgason[1], E.M Bahassi[1,2].* 1) Hematology & Oncology, University of Cincinnati, Cincinnati, OH; 2) UC Brain Tumor Center, Cincinnati, OH.

Brain metastases(BMs) are commonly diagnosed intracranial neoplasms that occur during the progression of many systemic malignancies. Approximately 40-50% of these lesions come from lung cancer. Previous efforts to characterize patients that will develop BMs have yielded relatively little results. The purpose of this study is to leverage the power of next-generation sequencing in order to circumvent prior issues in characterizing the metastatic potential of lung tumors. Whole genome sequencing data was obtained from 3 primary non-small cell lung tumors and matched BMs as well as and normal lung tissue. Data was then processed using the GATK quality control pipeline and somatic variants were called using Breakdancer and Somatic Sniper. Copy number variation was determined using Control-FREEC. Calls were validated using Integrative Genomics Viewer. Data was visualized using Circos Diagrams. In all three cases, we were able to identify specific mutations leading to the activation of pathways that have been implicated in metastasis, including those involved in cellular adhesion, motility, and proliferation. Several genes that have been identified are now candidates for future studies. This evidence not only highlights the merits of this analysis in classifying primary tumors based on metastatic potential and stratifying patients based on the risk of developing BMs, but also provides potential signaling pathways that can be therapeutically targeted to prevent brain metastases.

3347S

RAS driver mutations are present in 36% of acute lymphoblastic leukaemia cases in children with Down syndrome and are mutually exclusive with JAK2 mutations. *D. Nizetic[1,2,3], S. Nikolaev[4], M. Garieri[4], F. Santoni[4], E. Falconnet[4], P. Ribaux[4], M. Guipponi[5], A. Murray[6], J. Groet[1], E. Giarin[6], G. Basso[6], S.E. Antonarakis[4,7].* 1) The Blizard Institute, Barts and The London School of Medicine, Queen Mary Unive, London, Switzerland; 2) LonDownS Consortium, The Wellcome Trust, United Kingdom; 3) Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Unit 04-11, Proteos Building, 61 Biopolis Drive, 138673 Singapore; 4) Department of Genetic Medicine and Development. University of Geneva Medical School. 1 rue Michel Servet, 1211 Geneva 4, Switzerland; 5) Geneva University Hospitals - HUG. Service of Genetic Medicine. 4 Rue Gabrielle-Perret-Gentil, 1211 Geneva 4, Switzerland; 6) Dipartimento di Salute della Donna e del Bambino, University of Padua, IRP - Istituto di Ricerca Pediatrica - Fondazione Città della Speranza, Padua, Italy; 7) IGE3 institute of Genetics and Genomics of Geneva. 1 rue Michel Servet, 1211 Geneva 4, Switzerland.

Children with Down syndrome (DS) and acute lymphoblastic leukaemia (ALL) have poorer survival and more relapses, than non-DS children with ALL, highlighting an urgent need for deeper mechanistic understanding of DS-ALL. Using full exome or cancer genes-targeted sequencing of 42 ALL samples from 39 DS patients, we uncover driver mutations in RAS (KRAS and NRAS) recurring to a similar extent (15/42) as JAK2 (12/42) mutations or P2RY8-CRLF2 fusions (14/42). RAS mutations were almost completely mutually exclusive with JAK2 mutations (p=0.016), driving a combined total of two thirds of analysed cases. Clonal architecture analysis revealed that both RAS and JAK2 drove sub-clonal expansions primarily initiated by CRLF2 rearrangements, and/or mutations in chromatin remodellers and lymphocyte differentiation factors. Remarkably, in 2/3 relapsed cases there was a switch from a primary JAK2 or PTPN11 mutated sub-clone to a RAS-mutated sub-clone in relapse. These results provide important new insights informing the patient stratification strategies for targeted therapeutic approaches for DS-ALL.

3348M

Integrated analysis of transcriptome and exome in cancer samples improves interpretation and reveals additional therapeutic insights. *S.M. Boyle[1], M.J. Clark[1], E. Helman[1], D.M. Church[1], S. Lou[1], S. Kirk[1], P. Sripakdeevong[1], M. Karbelashvili[1], M. Pratt[1], M. Snyder[2], R. Altman[2], R. Chen[1], J. West[1].* 1) Personalis, Menlo Park, CA; 2) Department of Genetics, Stanford University, Stanford, Ca.

Next-generation sequencing is increasingly applied as a method for cancer analysis in both research and clinical settings. Many research studies used DNA or transcriptome sequencing to assess the genetic basis of tumor progression and, on some occasions, guide therapeutic decision-making. While gene panel, exome and whole genome sequencing of tumors are widely used for clinical guidance and translational research, transcriptome sequencing has yet to be widely adopted in the clinical environment. However, analyzing the transcriptome allows for improved detection of expression changes, gene fusions, and alternative splicing events. To assess these unique features of RNA, we performed whole transcriptome sequencing along with ACE exome sequencing of cancer samples. Our sample set included commonly used cell lines, cell lines modified to contain specific mutational ratios, and real-world patient samples. We assessed important genetic alterations and limits of detection through analysis of gene expression levels, gene fusion events, and RNA-derived variant calls. These calls were then filtered and annotated with our cancer gene database. There are a growing number of therapeutically relevant gene fusions, which are critically important for accurate tumor mutational analysis. We identified important gene fusion events at the expression level and cross-referenced these with exome findings, where present. Many fusion events would have been missed by exome or gene panel analysis alone. Additionally, transcriptome analysis improves upon assessment of the impact of CNVs on expression. We performed gene expression analyses to accurately quantify differential regulation, something that can only be assumed when analyzing CNVs from DNA alone. By combining expression and DNA-based CNV analyses, we were able to generate more accurate CNV interpretations. A third area transcriptome gives us additional information on is the expression of mutant alleles. To that end, we called variants directly from the RNA, and identified variants that are truly expressed in the tumor. We observed variants in important cancer driver genes which were selectively identified in either genomic or transcriptomic samples through allele specific expression, poorly covered regions, or lack of expression. Taken together, these joint approaches demonstrate how dual interpretation substantially bolsters accuracy and leads to changes in both research results and clinical decisions.

3349T

Deep targeted sequencing for accurate identification of low frequency somatic variation in cancer. *D. Burgess, D. Green, K. Jefferson, M. Brockman, R. Bannen, M. D'Ascenzo, K. Heilman, L. Feng, J. Patel, J. Voichick, T. Richmond.* Research & Development, Roche NimbleGen, Inc., 500 South Rosa Road, Madison, WI.

Somatic genome variation is a key feature underlying the pronounced cellular and clinical heterogeneity observed across a wide range of cancers. The ability to detect and accurately quantify this variation in tumors, particularly as it evolves over time, may be useful for predicting the effectiveness of therapy and monitoring subsequent tumor burden, the emergence of pharmacoresistance, the occurrence of metastasis and, ultimately, projecting a patient's prognosis and survival. If functionally relevant variants can be detected at a low molecular frequency during monitoring, the time and options available for clinical intervention would be greater. For a tumor genome analysis method to be effective for this application, but also practical for widespread implementation, it must be sensitive, accurate, reproducible, fast, low-cost, uncomplicated, amenable to automation and have robust data analysis and reporting outputs. For primary tumor analysis, and for utility in noninvasive screening or monitoring applications, efficacy with low sample input is also an essential element. Targeted next-generation sequencing (tNGS) approaches are a good fit for many of these requirements. We developed a tNGS method based on optimized, multiplexed, molecular inversion probes (MIPs) for the identification of low frequency somatic variation in cancer. A panel of ~700 MIPs was synthesized to interrogate mutational hot spots within a set of 55 cancer genes. The probes targeted both DNA strands and were designed to facilitate bioinformatic error correction. They also incorporated unique molecule identifiers (UIDs) to facilitate tagging of PCR duplicates and the accurate assessment of molecular complexity free of amplification bias. A performance evaluation was executed using a series of normal and cancer mutation reference samples. Analysis of experimental results using ≤100ng of input sample DNA and <40Mb of raw sequencing (paired-end Illumina MiSeq) indicated that >98% of the target region was covered to at least 50x coverage depth, with a duplicate read rate less than 20%, and with >96% of the probes exhibiting ≥20% of the mean panel coverage. Initial evaluation of sensitivity for the panel demonstrated that known single nucleotide variants in the samples could be reliably detected when present at frequencies down to below 1%, with low false-positive rates. We conclude that optimized MIP panels are a promising solution for important applications in cancer genomics.

3350S

Exome sequencing identifies novel cancer-predisposing genes in familial thyroid cancer. *A. Chaudhuri[1], R. Gupta[1], T. Boben[2], P. Kumar[1], V. Ram Prasad[1], S. Santhosh[1].* 1) MedGenome Inc., San Francisco, CA; 2) Kerala Institute of Medical Sciences, Kerala, India.

Thyroid cancer is the most common endocrine malignancy accounting for ~1.7% of all cancer incidences in the world. In India, incidence of thyroid cancer ranges from 2-3% of all cancer incidences. However, certain regions in India - in particular, the state of Kerala, have a significantly higher incidence (3-4-fold higher) of thyroid cancer than the rest of the country. There is no clear mechanistic explanation for this observation. We have applied next generation sequencing (NGS) to identify novel cancer predisposing genes in this population. To this end, we have sequenced exomes of four affected members (3 females + 1 male) of a family diagnosed with thyroid cancer, along with a normal female member from the same family. Three of the four affected were diagnosed with papillary thyroid carcinoma (PTC) and one with follicular thyroid carcinoma (FTC) leading to the identification of variants specific to each cancer type. Our analysis revealed 209,135 variants in genes, which included 36,084 coding region variants. By selecting variants with minor allele frequency of ≤ 5% from 1000genome, and those that resulted in non-synonymous alterations in the coding sequence, we identified multiple variants in genes involved in tissue inflammation, cell survival and extracellular matrix organization in all samples of thyroid cancer. However, further analysis of variants present only in the FTC sample identified truncating mutations in two negative regulators of NFκB pathway, along with mutations in the catalytic subunit of phosphoinositol-3 kinase catalytic subunit 2 alpha (PIK3C2A) suggesting activation of pro-survival pathways in this cancer type. By contrast, the PTC subtype was characterized by mutations in genes associated with growth, cell polarity and cell adhesion functions. In conclusion, our analysis revealed the underlying mechanism of familial thyroid cancer to be inflammatory in origin with distinct pathways participating in the origin of FTC and PTC.

3351M

Dysregulation of TGFB pathway in the formation of chordoma. *W. Chen[1], Q. Hu[1], L. Wang[2], J. Zhang[2], C. Zeng[1].* 1) Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China; 2) Beijing Tian Tan Hospital, Capital Medical University, Beijing, China.

As a rare malignant tumor chordomas arises from notochordal remnants and frequently occurs in skull base of patients. miRNA dysregulation has been observed in chordomas when comparing miRNA array data in chordoma tissues with skeletal muscle, notochord, or nucleus pulposus. To give a more comprehensive profiling of miRNA in chordomas, we sequenced both miRNA and mRNA in two chordoma tissues and two fetal notochords on HiSeq 2000. In both notochords and chordomas, there are over 600 miRNA expressed. The top five most abundantly expressed miRNAs accounted for more than 50% of total miRNAs in the whole expression library of chordomas or notochords, and among these top miRNAs, two tissues had 3 in common. A total of 45 significantly differentially expressed miRNAs were identified in chordomas, including 16 up-regulated and 29 down-regulated with changed transcriptional activities of their targeted mRNAs accordingly. The functions of these genes were enriched in retinoic acid receptor activation, DNA methylation, transforming growth factor beta (TGFB) pathway, etc. The chordoma up-regulated miR-29b-3p and its down-regulated target gene TGFB3 were further validated in additional 8 chordoma tissues and 8 notochords. Compared to notochords, miR29b-3p was over expressed in all chordoma tissues together with TGFB3 down expressed as revealed by qPCR. As shown by analysis of TGFB pathway in KEGG, we further found the genes involved in apoptosis were dysregulated in chordorma in comparison with notochords, suggesting the missing regulation of TGFB induced cell death had important contribution to chordoma. To identify some variations in DNA level that cause the dysregulation of miR-29b-3p and TGFB3, we further compared the copy number of their corresponding genes between chordoma tissues and their paired blood samples. Among the 8 chordoma patients, 5 had somatic copy number gain at miR-29b1 and miR-29b2 loci, while 4 had copy number loss in TGFB3 locus, suggesting suggests the dysregulation of TGFB pathway may play a critical role in the formation of chordoma.

3352T

Whole Exome Sequencing study of HPV-Positive and HPV-Negative Oropharyngeal Squamous Cell Carcinoma: Mutational Profile and Predisposition Gene Identification. *JS. CHOI[1], YS. JUNG[3], JI. KIM[1,2].* 1) Dept. of Biomedical Science, Seoul National Univ, College of Medicine, Seoul, South Korea; 2) Genomic Medicine Institute, Seoul National Univ, College of Medicine, Seoul, South Korea; 3) Head and Neck Oncology Clinic, National Cancer Center, Goyang-si, Gyeonggi-do, South Korea.

Head and neck squamous cell carcinoma (HNSCC) is the sixth leading cancer around the globe, arising in the oral cavity, oropharynx, larynx, and hypopharynx. There are 0.5 million new cases a year worldwide and the five-year survival rate of patients with HNSCC is about 40-50%. Head and neck cancer is strongly associated with certain environmental and lifestyle risk factors, including tobacco smoking, alcohol consumption, UV light, particular chemicals, and certain types of viruses, such as human papillomavirus (HPV). Oropharyngeal squamous cell carcinomas (OSCC), a subtype of head and neck squamous cell carcinoma, is a disease where malignant cells form in the tissue of the oropharynx and which is strongly associated with HPV infection. Oropharyngeal cancers can be divided into two groups, HPV positive, which are associated to HPV infection, and HPV negative cancers, which are usually related to heavy alcohol or tobacco consumption. Particularly, a rapidly rising HPV related oropharyngeal cancers in the developed countries has been well shown in some epidemiological studies. But there are not enough therapeutic molecular targets for HNSCC. We performed whole exome sequencing (WES) on three sets (blood, normal mucosa issue and oropharyngeal tumor tissue) of each 6 HPV positive and 3 HPV negative patients to reveal what genetic factors affect tendency to predispose to OSCC and to compare genetic profile on the basis of HPV infection. We found two of three HPV negative cases have sporadic mutation on TP53 and some others previously known genes to be involved in HNSCC (for example, CDKN2A, CSMD1), while HPV negative cases don't have alterations of TP53. We also found some OSCC predisposition genes by sequential comparisons with normal mucosa tissues. Furthermore, we found that HPV negative samples show more aberrant genetic alteration compared to HPV positive cases and these aberrations do not appear at the level of normal mucosa. This work reveals distinctive mutational genetic profiles of HPV positive and negative OSCC and shows genome wide mutational states of OSCC predisposition giving a clue how normal mucosa changes into malignant tumor cells. These findings open new approach for characterization of biological mechanisms and finding therapeutic targets of OSCC.

3353S

Evaluation of a multiplex PCR targeted enrichment approach for the detection of actionable mutations in FFPE samples via Next-Generation Sequencing. *F. de Abreu[1], J. Peterson[1], R. Samara[2], J. Collins[2], G. Tsongalis[1].* 1) Department of Pathology , Dartmouth Hitchcock Medical Center , Lebanon, NH; 2) QIAGEN Sciences, LLC, Germantown, MD.

Targeted Next Generation Sequencing (NGS) is rapidly becoming an essential tool in personalized cancer treatment. Patient tumor samples can be comprehensively genotyped to identify mutations with a wide variety of both therapeutic and prognostic implications. The often poor quantity and quality of patient samples, however, remains a significant challenge in routine clinical sequencing. The GeneRead DNAseq Panel for Tumor Actionable Mutations is capable of identifying actionable mutations in 8 cancer-associated genes from relatively low-concentration formalin-fixed paraffin embedded (FFPE) samples. This study will show the validation of this panel for the detection of tumor actionable mutations in FFPE samples. The GeneRead DNAseq Panel workflow can be completed in 3-4 days, and can be divided into 2 steps: library preparation, and sequencing/data analysis. 20ng of FFPE gDNA is required for library preparation. Ten samples and two controls were amplified using the GeneRead DNAseq Panel Primer Mix Pool. Sequence adapters were incorporated and purified using the GeneRead Size Selection Kit, and size selection was done using AMPure XP beads. Purified libraries were PCR amplified, and quantified using the GeneRead DNAseq Library Quant Array. Finally, libraries were normalized, pooled, and sequenced on Illumina's MiSeq® system. FASTq files were uploaded to the GeneRead DNAseq Variant Calling portal, where sequence alignment, annotation, and variant classification were performed. The GeneRead panel was able to detect and annotate previously-described mutations with very similar allelic frequencies. Additionally, the panel was able to detect previously-described insertions and deletions; however their annotation was incomplete. Using a commercially-available and thoroughly-annotated control, all five point mutations were correctly identified with similar allelic frequencies to manufacturer specifications. The Qiagen GeneReader Panel for Tumor Actionable Mutations is capable of obtaining both high read-depth and uniformity of coverage using samples with low concentration, producing high-confidence clinically significant variant calls.

3354M

Whole Genome Sequencing of Aggressive, Treatment-Naïve Prostate Tumors. *B. Decker[1,2], D.M. Karyadi[1], E. Karlins[1], B.W. Davis[1], L.S. Tillmans[3], J.S. Stanford[4], S.N. Thibodeau[3], E.A. Ostrander[1].* 1) Cancer Genetics Branch, NHGRI, National Institutes of Health, Bethesda, MD; 2) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, UK; 3) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 4) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109.

Prostate cancer (PCa) is the most common and second most lethal malignancy in American men, with approximately 239,000 diagnosed cases and 30,000 deaths in 2013. Early identification of men with aggressive versus indolent disease is a critical challenge in the clinical management of PCa. Recent sequencing-based studies have illuminated important aspects of PCa tumor development and molecular biology. In spite of recent progress, the coordinated role of SNVs, indels, and structural variants has not been well described, especially in the most aggressive tumors. To access the genomic landscape of aggressive prostate tumors, we performed deep whole genome sequencing of nine fresh frozen Gleason 8-10 prostatectomy samples and one nodal metastasis from a tenth patient. Subjects had a mean age at diagnosis of 63.6 years and all were of Caucasian ancestry. Tumor DNA was sequenced to an average depth of 95.3X, with an average germline coverage of 48.3X. Somatic SNVs, indels, and SVs were identified, and coordinated evaluation of somatic mutations is underway. Tumors harbor 2,387 - 15,500 somatic single nucleotide variants, and 96 - 491 such variants fall in coding regions. We also found 274 - 2,578 high-confidence somatic indels, and 112 - 927 somatic structural variants per tumor. All variants that are predicted to damage the encoded protein are being evaluated with gene set enrichment analysis, in order to uncover pathways and functional gene groups that are enriched for somatic mutations in aggressive PCa. We are also investigating the order of somatic genetic events in the evolution of aggressive PCa tumors using allelic imbalance, which informs variant clonality and enables analysis of earlier versus later mutations. Future work will focus on recurrently mutated pathways and comparison of our highly aggressive tumors with PCa tumors from other PCa Tumors. Ultimately, we aim to 1) Define the spectrum of somatic variation in ten high-grade, treatment-naïve PCa tumors; 2) Characterize genes and pathways that are recurrently disrupted by somatic mutations; and 3) Describe somatic variant chronology and clonal evolution in aggressive PCa tumors. As more tumors across the aggressiveness spectrum are sequenced, this approach may uncover valuable markers for PCa diagnosis, prognosis and treatment selection.

3355T

iCAGES: integrated CAncer GEnome Score for understanding personal cancer genomes. *C. Dong[1,2], H. Yang[1], X. Liu[4,5], K. Wang[1,2,3].* 1) Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA; 2) Biostatistics Division, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) Department of Psychiatry, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA; 4) Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, Texas, 77030, USA; 5) Division of Epidemiology, The University of Texas Health Science Center at Houston, Houston, Texas, 77030, USA.

Cancer is a disease caused by accumulation of somatic mutations. Integration of personal genomic information with prior biological knowledge allows better understanding of cancer on a personal and systematic level. Here, we propose iCAGES (integrated CAncer GEnome Score), a statistical model that integrates multiple sources of information, from personal genomic information, to prior knowledge of its genes and networks and finally to drug-gene interactions. For each patient, iCAGES takes as input the genomic somatic mutations and calculates a prioritized list of cancer driver genes and candidate drugs, using a three-layer estimation procedure: The first layer summarizes personal genomic information using radial SVM (Support Vector Machine) algorithm based on the observation of a novel non-linear relationship between evolutionary conservation and function alterations of potential cancer driver mutations retrieved from COSMIC database, modeled with nine different deleteriousness prediction scores (PolyPhen-2, SIFT, MutationTaster, Mutation Assessor, FATHMM, LRT, GERP++, SiPhy and PhyloP, retrieved from ANNOVAR website and dbNSFP database) as its features. The second layer is based on previous knowledge on the association of each gene with cancer, modeled using Phenolyzer. Natural language processing schema was used for summarizing related biological research publications compiled from various databases, such as OMIM, ClinVar and GWAS catalog. Overall conditional probability for each gene on cancer was calculated and then integrated with protein-protein interaction and other network information from sources, such as HPRD protein interaction, NCBI's Biosystem and HGNC gene family, to adjust for known protein-protein interactions. The third layer is based on gene-drug interaction, currently modeled using publicly available algorithm from DGIdb (Drug-Gene Interaction database), but novel algorithms are also under development. We evaluated the performance of iCAGES using somatic nsSNV data retrieved from a recently published case on sorafenib-sensitive lung adenocarcinoma. Within 10 minutes, iCAGES prioritized ARAF gene and sorafenib as the candidate gene and drug, using a list of nsSNP as input, consistent with the original publication. In summary, we developed a statistical model that leverages personal genomic information with prior biological knowledge, shedding light into cancer driver genes identification, personalized drug discovery and cancer treatment.

3356S

The Activation of LINE-1 Retrotransposition in Barrett's Esophagus and Esophageal Carcinoma. *T.T. Doucet[1,2], N. Rodić[3], T. Jungbin Choi[1], J. Young Ahn[1], Y. Cheng[4], K.H. Burns[1,3], S.J. Meltzer[4], H.H. Kazazian[1].* 1) Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Pre-doctoral training program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA; 3) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America; 4) Division of Gastroenterology and Hepatology, Department of Medicine and Department of Oncology, Johns Hopkins University, Baltimore, Maryland.

One cause of genomic instability is the insertion of human retrotransposons that copy and paste themselves into new sites. We investigated how retrotransposition is involved in genomic instability in esophageal carcinogenesis. Using the L1-seq technique, we enriched for young, active LINE-1 (L1) sequences and utilized next generation sequencing coupled with our analysis pipeline to isolate L1 insertions unique to the tumors of twenty-one esophageal adenocarcinoma (EAC) cases and squamous cell carcinoma (SCC) patients, divided into three cohorts. Our first cohort contained 1 EAC case and 4 SCC cases; the second cohort had eleven EAC cases. The third cohort contained five cases of EAC with matched DNA from the precursor lesion, Barrett's esophagus (BE). Between the three cohorts, we confirmed 86 somatic insertions many of which occurred in introns of genes known to be involved in carcinogenesis. To investigate whether these confirmed insertions were clonal, we validated a subset of them in BE as well as tumor DNA. After validation, eight insertions were present only in BE, while two occurred in both BE and tumor from the same individual. In esophageal carcinoma cases for which we had tissue, at least 20 of the 42 high-stringency insertions were present in multiple regions of each tumor, suggesting that these insertions were clonal and occurred early in tumorigenesis. By immunohistochemistry, we also observed high levels of ORF1p, one of two L1 proteins, in 6 of 9 patient samples evaluated. In the individuals in whom we observed ORF1p expression, protein concentration increased linearly with progression from dysplasia to frank cancer. We are also investigating germline insertions within introns of genes involved in carcinogenesis to determine whether or not certain insertions predispose individuals to develop esophageal carcinoma. 174 germline insertions, including the genes PADI6, AKT3, and THADA (not polymorphic in the general population), were recurrent in multiple cohorts of patients. Furthermore, there was one germline insertion predicted to reside in an internal exon of a CDH17 gene in two cohorts. We are currently elucidating effects of these insertions on gene expression. In light of these data, we conclude that the process of retrotransposition is a potential cancer driver in at least a subset of EAC patients. Using multiple approaches, we now seek to understand how retrotransposition contributes to genomic instability in tumorigenesis.

3357M

A big family of adenomatous polyposis with different extraintestinal manifestations. *M. Duz[1], A. Gezdirici[1], H. Ulucan[1], G. Guven[1], M. Seven[1], A. Yuksel[1,2], M. Ozen[1,3].* 1) Department of Medical Genetics, Cerrahpasa, Istanbul, Turkey; 2) Biruni University, Istanbul, Turkey; 3) Department of Pathology & Immunology Baylor College of Medicine, Houston, TX, 77030, USA.

Familial adenomatous polyposis (MIM #175100) (FAP) is an autosomal dominant disorder characterized by predisposition to colorectal cancer. Affected individuals usually develop hundreds to thousands of adenomatous polyps of the colon and rectum. These findings can progress to colorectal carcinoma in a small proportion of these patients. Moreover, extraintestinal manifestations can appear as osteomas, dental anomalies, congenital hypertrophy of the retinal pigment epithelium, soft tissue tumors, and desmoid tumors. Molecular genetic testing of APC, which is frequently used in the early diagnosis of at-risk family members, in addition to verification of the diagnosis of FAP and attenuated FAP in individuals with suspected findings, detects disease-causing mutations in about 90% of individuals with typical FAP. Here, we report a big family of five generations with thirty five individuals whose referral reasons were ranging from gastrointestinal problems to total colectomy. The proband, a 23-year-old female, had altered bowel habits with 20 adenomatous polyps in her colonoscopy. Her father has also history of colorectal cancer, accompanied with hundreds to thousands of adenomatous polyps of the colon and rectum at age 28 years, followed by total colectomy. Molecular testing revealed deleterious nonsense germline mutation c.C1370G, p.S457X in 11th exon of APC in thirteen of twenty three members of the family studied. This clinical picture and mutation within the family are compatible with FAP. The mutation carriers in the family show different symptoms and extraintestinal manifestations. Our results suggest that a single APC germline mutation might come up with clinical variability even in closely related individuals due to modifier genes of APC and epigenetic factors.

3358T

Exomic and transcriptomic patterns of colitis-associated carcinoma. *D. Esser, M. Falk-Paulsen, K. Aden, P. Rosenstiel.* Institute of Clinical Molecular Biology, Kiel, Schleswig-Holstein, Germany.

Inflammatory bowel diseases (IBDs) are chronic inflammatory disorders in the gastrointestinal tract, which arise as a result of the interaction of environmental and genetic factors. The risk of developing colorectal cancer (CRC) is increased for a person with IBD and correlates with the degree of colonic inflammation and disease duration. Patients with colitis-associated cancer (CAC) are younger than patients with sporadic colorectal cancer and show a poorer survival in the advanced stage. Both cancer types are multi-step processes, but differ in certain mutation types. Compared to the more frequent sporadic CRC, the genomic signatures of CAC are far less understood.

In order to investigate the genetic background of colitis-associated carcinoma we chose a well established inflammation-associated colorectal carcinogenesis model in mice (AOM/DSS) and compared the results from early and late stage cancer with inflamed tissue from DSS-treated and with healthy samples from untreated mice. Three experimental settings with different DSS concentrations and treatment durations were investigated. After tissue collection, the whole exome as well as the whole transcriptome of tumor and all proximal colonic non-tumor samples were sequenced on the Illumina HiSeq 2000, followed by bioinformatic analyses. Beside known CAC-associated single nucleotide variants (SNVs) in e.g. *Ctnnb1*, we were able to identify a multitude of novel potentially damaging mutations. Furthermore, we observed regional clustering of somatic SNVs in several tumor samples and could show that the cancer stage is associated with the mutation rate and the mutation pattern. Additionally, we found a tumor specific gene expression profile. This also included the investigation of novel transcriptionally active regions, fusion transcripts and splice variants. Based on exome and transcriptome data, relevant genes and pathways were filtered out using different methods.

We performed one of the first studies analyzing the genetic background of CAC with next-generation sequencing technologies and demonstrated that a combined application of whole exome and strand-specific whole transcriptome sequencing can reveal new insights into the development of CAC.

3359S

Analysis of metastatic diffuse gastric cancer genomes in a Mendelian family with an inherited *CDH1* mutation. *S. Greer[1], J. Bell[2], E. Hopmans[2], L. Miotke[1], S. Grimes[2], H. Ji[1].* 1) Stanford University School of Medicine, Stanford, CA; 2) Stanford Genome Technology Center, Stanford, CA.

Gastric cancer (GC [MIM 137215]) is the fourth most common malignancy and the second leading cause of cancer deaths worldwide. Based on distinctive histopathological features, GC can be classified into two subtypes: intestinal GC and diffuse GC, with evidence suggesting that different carcinogenic pathways are involved in the development of each subtype. Compared with intestinal GC, the incidence of diffuse GC is increasing more rapidly and has an overall worse prognosis as a result of its diagnosis after metastasis has occurred. For this study, we investigated the underlying genetics of diffuse gastric cancer metastasis, leveraging the highly informative Mendelian genetics of a family with Hereditary Diffuse Gastric Cancer (HDGC), an inherited cancer syndrome in which germline mutations in the *CDH1* gene confer a 70% lifetime risk of developing diffuse GC. *CDH1* [MIM 192090], which encodes the E-cadherin protein, is well known to be associated with diffuse GC. However, very few additional cancer driver genes that work in coordination with *CDH1* loss-of-function in the diffuse subtype have been identified to date. In the case of Mendelian cancer genomes, the initial cancer driver event is known, which offers a unique scientific opportunity to delineate additional cooperative cancer drivers involved in metastasis. With whole genome sequencing, we analyzed a primary gastric tumor and ovarian metastasis (Krukenberg tumor) from an individual with a germline, hereditary *CDH1* mutation, as well as a Krukenberg tumor from her sister. Both sisters were found to have a germline splice site mutation in intron 10 of *CDH1*, leading to loss of function of the critical *CDH1* tumor suppressor, representing the first hit in the development of their primary tumors. Several cooperative cancer drivers were identified in this study, including SNVs in *TP53* [MIM 191170], and amplification of the *FGFR2* [MIM 176943] gene was observed in the tumors of both sisters. This study has significant implications in the consideration of precision medicine approaches for treating metastatic cancer by shedding light on the genetics of diffuse gastric cancer, molecular mechanisms of metastasis, and oncogenic events that are therapeutic targets.

3360M

The mutational landscape of peritoneal malignant mesothelioma. *O. Harismendy[1,3,5], H. Alakus[2,3,6], S.E. Yost[3,7], C.-P Cheng[3,8], D.V. Jaquish[2], E.S. Mose[2], R.P. French[2], B. Woo[1,3], K. Jepsen[4], K.A. Frazer[1,2,4,5], A.M. Lowy[1,2].* 1) UC San Diego Moores Cancer Ctr, La Jolla, CA; 2) UCSD Department of Surgery, La Jolla CA; 3) UCSD Department of Pediatrics, La Jolla CA; 4) UCSD Institute for Genomic Medicine, La Jolla CA; 5) UCSD Clinical and Translational Science Institute, La Jolla CA; 6) Department of General, Visceral and Cancer Surgery, University of Cologne, Cologne, Germany; 7) UCSD Bioinformatics Graduate Program, la Jolla CA; 8) National Cheng Kung University, Tainan, Taiwan.

Malignant mesothelioma (MM) arises from mesothelial cells that line the pleural, peritoneal and pericardial surfaces. The majority of MM are pleural and have been associated with asbestos exposure. Previously, pleural MM have been genetically characterized by the loss of BAP1, either somatic or inherited, as well as loss of NF2 and CDKN2A. The rare peritoneal form of MM (PeMM) occurs in ~10% cases. With only ~300 cases diagnosed in the US per year, its link to asbestos exposure is not clear and its mutational landscape unknown. We analyzed the somatic mutational landscape of epithelioid PeMM using whole exome sequencing (N=7) and copy number analysis (N=9). In total, we identified 87 non-silent mutations in 83 genes with a median of 8 mutated genes per tumor, therefore resulting in a relatively low mutation rate (~1.6 10-6). BAP1 is the most recurrently mutated gene in PeMM, affecting 6/9 samples. BAP1 deletions occurred as focal events in 4/9 cases, in the context of a loss of chromosome 3p21 and are frequently accompanied by somatic mutations of the remaining allele. In one additional case, the loss of the entire chromosome 3 (encoding for BAP1) leaves a non-functional copy of BAP1 carrying a rare nonsense inherited variant, thus suggesting a potential genetic predisposition in this patient. The PeMM mutational landscape, dominated by the loss of chromosome 3p21, is reminiscent of the one observed in clear-cell renal carcinoma. In contrast to pleural MM, we do not observe deletions of NF2 and CDKN2A in PeMM, suggesting genetic differences between the two diseases. These findings support the use of molecularly guided clinical trials and therapies in PeMM.

3361T

The Landscape of Inherited and Somatic Mutations in Fallopian Tube Carcinoma. *M.I. Harrell[1], T. Walsh[2], S. Gulsuner[2], K.J. Agnew[1], B. Norquist[1], M.K. Lee[2], S. Bernards[1], K.P. Pennington[1], M.C. King[2], E.M. Swisher[1].* 1) Obstetrics and Gynecology, Division of Gynecologic Oncology, University of Washington, Seattle, WA; 2) Medical Genetics, University of Washington, Seattle, WA.

Fallopian tube (FT) carcinoma is a rare gynecological malignancy, whose incidence is most likely underestimated due to misdiagnoses as ovarian cancer. Recent advances in pathology with more thorough serial sectioning techniques have led to better identification. We undertook to define germline and somatic mutations in 98 FT carcinoma cases enrolled at diagnosis in IRB-approved studies at the University of Washington, excluding cases with carcinoma identified at the time of risk-reducing surgery. We sequenced DNA from blood and neoplastic tissue using BROCA, a targeted capture and massively parallel genomic sequencing approach that detects all classes of mutation in 64 genes. The fraction of all primary ovarian, peritoneal and FT carcinoma defined as a FT primary has increased over time: FT diagnoses accounted for 23% of cases in the past five years compared to 5% in the previous 10 years (p=0.002). Of 98 subjects with FT carcinoma, 32 (33%) had loss of function germline mutations in 9 genes. Germline mutations were identified in *BRCA1* in 18 (19%), and in *BRCA2* in 4 (4%) patients. Germline mutations were also identified in other known or candidate breast and/or ovarian cancer susceptibility genes including 2 in *BRIP1* (2%), 2 in *CHEK2* (2%), and 1 each in *BARD1*, *BLM*, *MRE11A*, and *ATM* (1%). One mutation was found in *FANCL*. Mosaic mutations were found in *TP53* (1%) and in *PPM1D* (2%). Together, germline and mosaic mutations were identified in 34 (35%) cases. Neoplastic DNA from 30 FT advanced stage carcinomas sequenced with BROCA revealed 27 somatic mutations in *TP53* (90%), 1 in *PTEN* (3%), and 1 in *CHEK2* (3%). The germline mutation rate was higher in FT carcinoma compared to our previous studies in ovarian carcinoma (35% versus 22%), and the profile of mutant genes was similar. Recently identified ovarian cancer susceptibility genes such as *BRIP1*, *BARD1*, and *PPM1D* are also associated with hereditary fallopian tube cancer. Known breast cancer genes (i.e. *CHEK2, ATM*) and other cancer-associated genes (*BLM*, *FANCL*) require further study to test their association with FT or ovarian cancer. Increased recognition by pathologists of the FT as a primary site of disease is changing the relative distribution of FT, ovarian, and peritoneal carcinoma as primary diagnoses, making the profiling of FT cancers increasingly important.

3362S

A rare somatic mutation in the TEL patch of telomere protein TPP1 acts as a driver of childhood acute lymphoblastic leukemia. *J. Healy[1], J.F. Spinella[1], P. Cassart[1], N. Garnier[1], C. Drullion[1], R. Vidal[1], V. Saillour[1], C. Richer[1], M. Ouimet[1], S. Busche[2], B. Ge[3], T. Pastinen[2,3], D. Sinnett[1,4].* 1) Division of Hematology-Oncology, Sainte-Justine UHC Research Center, Université de Montréal, Montreal, QC, Canada; 2) Department of Human Genetics, McGill University, Montreal QC, Canada; 3) McGill University and Genome Quebec Innovation Centre, Montreal, QC, Canada; 4) Department of Pediatrics, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada.

Childhood acute lymphoblastic leukemia is a heterogeneous disease both biologically and clinically, and is the leading cause of cancer-related deaths among children. Thorough investigation of pediatric ALL genomes, and identification of not only recurrent but also rare/private somatic driver mutations, is crucial in order to understand the underlying genomic complexity of this disease and thus better diagnose and treat it. Here we performed deep re-sequencing of a hyperdiploid pre-B childhood ALL patient and used a strict filtering strategy to identify somatic events with high driver potential. Among the candidate mutations, we identified a rare non-synonymous mutation (p.G223V) in the oligonucleotide-/oligosaccharide-binding(OB)-folds of the TEL patch of telomere protein TPP1/ACD, member of the shelterin complex that protects chromosome ends and regulates telomerase activity. Using in vitro cytotoxicity assays in ALL cells, we demonstrated a strong protective effect of this mutation against apoptosis associated with telomere maintenance. Thus, for the first time we identified a somatic mutation in TPP1 with functional implications in cellular transformation, suggesting a role for TPP1 as a novel driver of childhood ALL and highlighting the importance of rare/private somatic mutations in understanding disease etiology.

3363M

Mutational spectrum of RET Proto-oncogene in Iranian Patients with Medullary Thyroid Carcinoma. *M. Hedayati[1], M. Zarif Yeganeh[1], S. Sheikholeslami[1], F. Azizi[2].* 1) Cellular and Molecular Research Center, Research Institute for Endocrine Science, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 2) Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Introduction: Thyroid cancer is the most common endocrine malignancy. It divides into four groups, including papillary (80%), follicular (10-15%), medullary (5-10%), and anaplastic thyroid cancer. Medullary Thyroid Carcinoma (MTC) is one of the most aggressive thyroid tumors which occur in both hereditary (25%) and sporadic (75%) forms. Mutations of the RET proto-oncogene in MTC development have been well demonstrated. The aim of the study was to investigate the mutational spectrum of exons 3, 5, 8, and 10-18 of RET proto-oncogene in MTC patients. Material and Methods: This retrospective study has been started since 2001 in Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, and Tehran, Iran. 370 participants, including 224 patients (168sMTC, 39FMTC, 8MEN2A, 3MEN2B, 5pheochromocitoma), and 146 relatives were evaluated. Genomic DNA was extracted by the standard Salting Out/ProteinaseK method and mutation detection was performed through direct DNA sequencing. Sequence analysis was performed by Chromas Software version 2.3. Results: Totally, in 224 patients (124 Female, 100 Male) and 146 relatives (80 Female, 66 Male), 79 mutations were identified in RET main exons, including exons 10, 11, 13-16. Furthermore, 282 Single Nucleotide Polymorphism (SNP) were found in exons 3, 13, and 14. Interestingly, SNPs G691S and S904S were 100% in linkage disequilibrium in 125 patients and 70 relatives. The most common mutation in our population was C634Y (4%) whereas C618R, C618S, C620G, L887L mutations had rare allele frequency (0.3%). Moreover, R886Q mutation was detected in exon 15 in two members of a family affected with MTC, for the first time. Discussion: Exon 11 and after that exon 10 were the most frequently mutated exons of RET proto-oncogene in MTC patients in Iranian population. As about half of patients with the hot spot mutations had the G691S/S904S haplotype simultaneously, further analysis needs for clarifying the collective effects of multiple risk alleles in MTC development. Key words: Medullary Thyroid Cancer, RET proto-oncogene, Germline mutation.

3364T

Analysis of RNA-Sequencing Data Reveals Association of JAK-STAT Pathway with NK/T-Cell Lymphoma. *JH. Hwang[1], HY. Park[1], SB. Lee[2], JS. Choi[1], YH. Ko[3,5], WS. Kim[4,5], SJ. Kim[4,5], HY. Yoo[5], JI. Kim[1,2].* 1) Seoul National Univ, College of Medicine, Seoul, South Korea; 2) Genomic Medicine Institute, Medical Research Center, Seoul National University, Seoul, Republic of Korea; 3) Department of Pathology, Samsung Medical Center; 4) Department of Medicine, Samsung Medical Center; 5) SungKyunkwan University School of Medicine, Seoul, Republic of Korea.

NK/T-cell lymphoma is a rare and aggressive subtype of non-Hodgkin lymphoma. Although it has been identified that lymphoma is associated with JAK-STAT pathway, there are few studies of NK/T-cell lymphoma using RNA-sequencing. We sequenced the transcriptomes of three tissues, seven cell lines of NK/T-cell lymphoma, and three normal NK/T cell lines to identify a diverse transcriptional spectrum such as gene fusions, differentially expressed genes, and transcription factors that affect differentially expressed genes. In this study, we discovered 24 in-frame fusion genes in two tissues and five cell lines. Among these, we found SND1-LEP fusion in a single cell line, and theLEPgene in this sample was expressed 20 times more than normal samples. It has been reported that leptin is involved in the tumorogenesis by activation of cellular signal transduction of JAK-STAT pathway. The genes showing abnormal expression patterns were also shown to be enriched in the JAK-STAT pathway(p-value = 1.92×10-4). Based on the expression data, we tested enrichment of predicted transcription factor binding sites for genes specifically activated in each sample. The binding sites of KLF4, MZF1, and SP1, which were known to be associated with JAK-STATpathway and lymphoma, were enriched in both tissues and cell lines.Taken together, these results show the association between NK/T-cell lymphoma and JAK-STAT pathway again at the transcriptional level.

3365S

Exome sequencing reveals novel mutation hotspots in microsatellite unstable colorectal cancer. *U.A. Hänninen[1,2], S. Tuupanen[1,2], J. Kondelin[1,2], P. von Nandelstadh[2], T. Cajuso[1,2], A.E. Gylfe[1,2], R. Katainen[1,2], T. Tanskanen[1,2], H. Ristolainen[1,2], J. Böhm[3], J-P. Mecklin[4], H. Järvinen[5], L. Renkonen-Sinisalo[5], C. Lindbjerg Andersen[6], M. Taipale[2,7], J. Taipale[2,7], P. Vahteristo[1,2], K. Lehti[2], E. Pitkänen[1,2], LA. Aaltonen[1,2].* 1) Haartman Institute, Department of Medical Genetics, University of Helsinki, Finland; 2) Research Programs Unit, Genome-Scale Biology Research Program, University of Helsinki, Finland; 3) Department of Pathology, Jyväskylä Central Hospital, University of Eastern Finland, Finland; 4) Department of Surgery, Jyväskylä Central Hospital, University of Eastern Finland, Finland; 5) Department of Surgery, Helsinki University Central Hospital, Finland; 6) Department for Molecular Medicine (MOMA), Aarhus University Hospital, Denmark; 7) Science for Life Center, Department of Biosciences and Nutrition, Karolinska Institutet, Sweden.

Colorectal cancer (CRC) with microsatellite instability (MSI) defines a subgroup (15%) of CRCs with distinct clinical characteristics. The mutation pattern in MSI CRC consists of small insertions and deletions in microsatellite regions throughout the genome. This pattern results from defects in the mismatch repair system, having either a hereditary (Lynch syndrome) or sporadic origin. Improved cancer diagnostics, novel therapeutic interventions, and personalized care require more comprehensive knowledge on mutations affecting tumor formation. The most frequently mutated cancer genes are already well known; mutations occurring at intermediate or low frequencies remain largely unidentified, however. Because mutations in oncogenes emerge recurrently at the same amino acid positions, a newly suggested criterion for defining new driver genes, rather than frequency, is mutation pattern. Genes with recurrent codon-specific somatic mutations are likely drivers of tumorigenesis, and potential therapeutic targets. The exomes of 25 sporadic MSI colorectal tumors and their corresponding normal tissue samples served as the discovery set for identifying base-specific somatic mutation hotspots. These were found in 43 genes, of which three were the following known oncogenes: BRAF (V600E), CTNNB1 (T41A) and PIK3CA (H1047R). The findings were validated by Sanger sequencing. The known genes and 33 novel potential hotspot genes were confirmed and thus further screened in an additional set of 254 MSI tumors. Fourteen potential hotspot genes displayed mutations also in the validation set with a mutation frequency of 0.4-2.8%. A database search revealed many of the identified hotspot mutations in other cancer types.

3366M

Detection of mutation hotspots through mutation set enrichment analysis. *P. Jia[1,2], Q. Wang[1], Q. Chen[1,3], K.E. Hutchinson[4], W. Pao[4,5], Z. Zhao[1,2,4,6].* 1) Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Center for Quantitative Sciences, Vanderbilt University, Nashville, TN; 3) Department of Biostatistics, Vanderbilt University, Nashville, TN; 4) Department of Cancer Biology, Vanderbilt University, Nashville, TN; 5) Department of Medicine/Division of Hematology-Oncology, Vanderbilt University, Nashville, TN; 6) Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN.

Although cancer genomes are extremely heterogeneous both within and among tumors, mutation hotspots are a prevalent disruptive mechanism in many cancer genes. Many of these mutation hotspots disrupt functional domains or active sites, leading to gain- or loss-of-function of the protein product. We leveraged these observations to identify and prioritize candidate cancer genes and formulated a mutation set enrichment analysis (MSEA). MSEA assesses whether somatic mutations within a gene have a tendency to cluster within certain regions of the gene's coding sequence. We proposed two complementary approaches, MSEA-clust and MSEA-domain. MSEA-clust simulates a walk through the sequences and renders a quantitative measurement of the location and extent to which mutations cluster. MSEA-clust is hypothesis-free because the convergent regions to be discovered are independent of a priori annotations of domains or functional sites. In contrast, MSEA-domain explicitly detects whether the mutations occur in a pre-defined functional domain more frequently than remaining regions. MSEA-domain implements a regression-based test and can only assess a priori domains; thus, it is hypothesis-driven. We assessed the two MSEA approaches using somatic mutation data from a comprehensive public resource (the Catalog of Somatic Mutations in Cancer database, COSMIC) and found that many known cancer genes have mutation hotspots. We then applied these approaches to available somatic mutation data in eight cancer types from The Cancer Genomic Atlas (TCGA) project. With false discovery rate (FDR) being less than 0.2, both MSEA-clust and MSEA-domain were effective in detecting well-studied cancer genes and new candidates that have not been intensively studied in cancer. Collectively, these analyses demonstrate that MSEA methods are well suited to detect cancer genes based on mutation hotspots and provide valuable tools for future cancer gene studies.

3367T

Advanced qualification and quantification of amplifiable genomic DNA (gDNA) for PCR-based targeted enrichment prior to next-generation sequencing. *Q. Jiang, Q. Peng, J. DiCarlo, Y. Wang, R. Samara, V. Devgan, E. Lader.* Biological Research Content, QIAGEN Sciences, LLC, Frederick, MD, USA.

Formalin-fixed paraffin-embedded (FFPE) tissue archives are an invaluable source for the molecular characterization of disease using next-generation sequencing (NGS). Unfortunately, heavy damage and fragmentation from the preservation process often results in FFPE DNA samples that cannot be amplified by PCR. Commonly used methods, such as fluorometry and spectrometry, are inadequate in assessing the amount of PCR-amplifiable DNA present in a sample. To overcome these challenges, we describe the development of a novel qPCR-based method for quantifying and qualifying FFPE samples for NGS. This method utilizes two qPCR assays to target multi-copy loci mapping to regions distributed across the genome. 329 assays in 85 distinct genomic regions were screened for qPCR performance. Assays with outstanding performance and consistency across all major human populations were selected. Plotting assay CT values against amplicon sizes generated a QC score, which provided size distribution information of sample gDNA. Additionally, the inclusion of a high-quality reference gDNA allowed quantification of amplifiable gDNA molecules, without the need for serial dilutions. To address suitability of this method for NGS, DNA from more than 100 FFPE samples with varying quality characteristics was extracted and subjected to qPCR analysis. For comparison, 50 FFPE samples were also quantified by spectrometry. All samples underwent PCR-based targeted enrichment, and were sequenced on either the PGM®, MiSeq®, or HiSeq®. All (high- or low-quality) FFPE samples quantified (and guided for input amount) by this strategy achieved good library yields, high specificity (>90%), and fairly even coverage on all sequencing platforms. Qualitatively, NGS results from these samples indicated that the calculated QC score predicted false positive rates. However, spectrometry was of limited usefulness in qualifying or quantifying these samples, which frequently failed to yield satisfactory NGS results. These samples produced varied library yields, mostly of low quantity, and low-specificity NGS reads. This advanced system for sample QC prior to NGS can assess both the quality and quantity of gDNA samples. This is critical for identifying whether a particular gDNA sample is suitable for NGS analysis and determining the appropriate DNA input for successful targeted enrichment via multiplex PCR. The applications presented here are for research use only. Not for use in diagnostic procedures.

3368S

Genetic Alterations and Evolutionary Behavior in Liver Metastatic Colorectal Cancer. *B. Lim[1], J. Kim[2], S. Kim[1].* 1) Medical Genomics Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, Republic of Korea; 2) Institute of Innovative Cancer Research and Asan Institute for Life Sciences, Asan Medical Center, Seoul, Republic of Korea.

Colorectal cancer (CRC) develops through well-established pathogenesis steps accompanying with specific mutations. However, mutations associated with CRC liver metastases (LMs) and their evolutionary behaviors are not fully understood. To investigate somatic mutations found in CRC LMs, we performed exome- and RNA-sequencing using 101 tissues from 41 CRC patients, comprising 22 non-metastatic and 19 liver-metastasized patients. Global genomic patterns, including variant allele frequency (VAF), base substitution spectrum and chromosomal aberration, of liver-metastatic primary tumors (LMPTs) were distinct from those of non-metastatic primary tumors (NMPTs) as well as LMs. Particularly, *PIK3CA* mutations were more frequently detected in LMPTs than NMPTs and significantly associated with clinical characteristics of advanced cancer. We observed 4 categories of evolutionary behavior of mutations (from Class 1 to 4) based on dynamic patterns of VAFs from LMPTs to LMs. The Class 1, which probably has high fitness to metastasis, was significantly associated with clinical characteristics of advanced cancer and frequently co-occurred with chromosomal aberrations in LMs compared with the Class 2, which probably has low fitness to metastasis. We also observed discrepant VAFs between exome and transcriptome. Importantly, VAFs of mutant alleles from *TP53* and *APC* at the transcriptomic level were significantly correlated with expression level of transcriptional target genes, revealing the functional implications of selective expression of mutant alleles. Especially, *TP53* mutant alleles tended to be positively selected through chromosomal aberration and allele specific expression in LMs. The examination of mutations and evolutionary behaviors at the genomic and transcriptomic level is required to apply precision medicine for liver-metastatic CRC patients.

3369M

Whole exome sequencing reveals that DNA repair and apoptosis pathways are affected in hereditary breast cancer cases. *J.L.D. Mentoor[1], F. Joubert[1,2], E. J. van Rensburg[1].* 1) Department of Genetics, University of Pretoria, Pretoria, South Africa; 2) Bioinformatics and Computational Biology Unit, Department of Biochemistry.

Family history serves as the most significant risk factor for breast cancer (BC). Variants in the *BRCA1&2* breast cancer susceptibility genes account for ~30-60% of hereditary BC. Moderate risk genes (e.g. *RAD51C, PALB2 & ATM* etc.) account for ~3-5% of familial cases. A considerable number of familial BC cases are not due to mutations in high/moderate susceptibility genes identified thus far i.e. *BRCAx* families. Mutant alleles in novel risk genes may account for the missing heritability. The aim of our research is to discover variants in novel high-penetrant genes that may increase susceptibility for BC.

Nine cases (*BRCA1&2* negative) were selected from 6 families with a history of breast and/or ovarian cancer (≥3 cases). Paired-end exome sequencing was completed by the Beijing Genomics Institute (Illumina). The Genome Analysis Toolkit 2.4.9 was used for variant calling and annotation. Mutations were filtered and interrogated from multiple biological perspectives by the Ingenuity variant analysis tool. High-quality, rare variants (truncating & splice-site) were selected based on minor allele frequency thresholds i.e. 1000genomes (<0.01) and ESP6500 (<0.01). High-priority mutated genes were cross-referenced to known BC and/or ovarian cancer genes with the ToppGene gene enrichment analysis suite and compared to somatic variant data derived from breast cancer resources. To further refine this list Sanger sequencing was used to confirm true-positive variants in high-priority genes within selected family members of index cases. Genes were ranked according to the presence of mutant alleles in index cases and affected family members. We then acquired gene expression data of highly ranked genes from databases containing comprehensive information of tissue specific gene regulation.

Thirteen biologically significant genes were identified that play a role in the inhibition of cell growth as well as DNA homologous recombination repair. Three of the most promising genes include; a novel putative tumour suppressor (*TCHP*) that is pro-apoptotic; the XPF-endonuclease homolog, *EME2*; and a POLQ-like helicase enzyme (*HELQ*). The latter two genes code for proteins that promote cross-linked DNA repair. Currently these genes are being screened in high-risk families and cohorts of women with breast cancer in order to explore their possible roles in breast cancer susceptibility.

3370T

Selective depletion of abundant RNAs to enable transcriptome analysis of low input and highly degraded RNA from FFPE breast cancer samples. *D. Munafo[1], B. Langhorst[1], S. Russello[1], F. Stewart[1], D. Sinicropi[2], J. Morlan[2], K. Qu[2], M. Liu[2], J. Jeong[2], M. Pho[2], R. Ambannavar[2], E. Dimalanta[1], T. Davis[1].* 1) New England Biolabs, Inc. 240 County Road, Ipswich, MA 01938, USA; 2) Genomic Health, Inc. 301 Penobscot Drive, Redwood City, CA 94063, USA.

Next generation sequencing of cDNA prepared from total RNA (RNA-seq) has become the method of choice for transcript profiling, and discovery. The standard whole-transcriptome approach faces a significant challenge as the vast majority of reads map to ribosomal RNA (rRNA). One solution is to enrich the RNA sample for polyadenylated transcripts using oligo (dT)-based affinity matrices; however, this also eliminates other biologically relevant RNA species, such as noncoding RNAs, and relies on having a high quality and quantity RNA sample. Here, we present a method to eliminate abundant RNAs from total RNA based on hybridization of probes to the targeted abundant RNA, followed by subsequent enzymatic degradation. We applied this method to remove cytoplasmic and mitochondrial rRNA from different eukaryotic total RNA samples (human, mouse and rat). We optimized this method for RNA with different degradation levels, from intact RNA to highly degraded formalin-fixed paraffin-embedded (FFPE) samples. We applied this method to remove ribosomal RNA from FFPE RNA extracted from breast cancer biopsy with an archive age of 1 and 10 year-old. We evaluated the depletion efficiency and off target effect of this method using strand specific RNA high-throughput sequencing. Ribosomal RNA depletion resulted in a minimal percentage of total reads mapping to rRNA sequences (<1% total reads), regardless of the species, input amount (1μg or 100 nanograms), or degradation level (intact or FFPE RNA). Additionally, there was very good transcript expression (FPKM) correlation (R>0.93) between rRNA depleted and non-depleted libraries. This indicates that rRNA depletion by this method is not altering the transcript expression levels. This method offers a robust and simple solution for transcriptome analysis of a variety of samples, including low quality and low quantity clinical samples such as FFPE RNA. Moreover, it is amenable to high-throughput sample preparation and robotic automation. This method is sensitive, specific, and produces increased coverage of less abundant, non-targeted transcripts in RNA-Seq studies.

3371S

Development of a novel Hotspot Frequency Ladder for Next Generation Sequencing (NGS) assay workflows. *N. Nataraj, M. Shahbazian, A. Lau, K. Norman.* AcroMetrix, ThermoFisher Scientific, Benicia, CA.

Increasing adoption of next-generation sequencing (NGS) technology has shed light on the need for more standardized controls to evaluate and optimize system performance. However, samples containing mutations of interest are difficult to source and cell line pooling experiments to determine limit of detection require significant investments of time and money. To simultaneously evaluate variant calling performance in >200 unique amplicons across 50 genes targeted by NGS tests, AcroMetrix® has developed a proprietary genomic/synthetic DNA blended material. This material contains over 400 commonly sequenced COSMIC mutations. All variants were confirmed by Sanger sequencing and the genomic DNA was characterized by the Genome in a Bottle Consortium. A material was developed to contain ~400 variants at six finely tuned frequencies, with one frequency level per tube. Variants were quantified by digital PCR, enabling the materials to be manufactured consistently at a range of frequencies including near the limit of detection for NGS or PCR platforms. The frequencies constructed were 48%, 29%, 18%, 11%, 5%, and 3%, and samples were tested using the Ion AmpliSeq™ Cancer Hotspot Panel v2 (CHPv2) with the Ion Torrent Personal Genome Machine® and Illumina® TruSeq® Amplicon Cancer Panel and TruSight® Tumor Panel on the MiSeq® instrument. At each frequency, the variants observed on each platform was compared to the expected engineered mutations. The limit of detection was then determined for ~400 variants. At the 48% level, the Ion AmpliSeq™ and TruSeq® tests detected greater than 99% of all SNPs (383/385 for CHP2 and 351/351 for TruSeq®). Detection dropped to 78% (302/385 for CHP2 and 279/351 for TruSeq®) at the 3% level. Indel detection at the 48% level was 79% (23/29) for CHPv2 and 74% (25/34) for the TruSeq® test. At the 3% level, indel detection ranged from 28% (8/29) for CHPv2 and 44% (15/34) for TruSeq® test, respectively. To determine whether the material performs similarly to tumor tissue samples, tumor samples were diluted with matched normal samples to mimic a range of frequencies. Linearity and limit of detection between the material and diluted tumor tissue samples were compared. Overall, highly multiplex controls with tunable frequencies allow for much more extensive, yet streamlined, assay evaluation and facilitate implementation and impart confidence to NGS testing. The Hotspot Frequency Ladder is under development.

3372M

Shared driver genes of familial and sporadic pancreatic cancer may explain the similar age of onset. *A.L. Norris[1], N.J. Roberts[1,2], R.H. Hruban[1,3], A.P. Klein[1,3], J.R. Eshleman[1,3].* 1) Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD; 2) Ludwig Center for Cancer Genetics, Johns Hopkins School of Medicine, Baltimore, MD; 3) Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD.

Familial pancreatic cancer (FPC) accounts for approximately 10% of pancreatic ductal adenocarcinoma cancer (PDAC) cases. While an earlier age of onset is a hallmark of most familial cancer syndromes, FPC does not present at an obviously earlier age. In this study, we collated the age of onset in FPC and sporadic PDAC reported in the literature. FPC cohorts had a reported mean or median age of 52 to 69 years old at diagnosis, while the range for sporadic PDAC cohorts was 60-74 years old. In our own small cohort, the median age of onset was 66 years old for both the FPC and PDAC patients. From our analysis of the literature, we show that any difference in the age of onset in FPC and sporadic PDAC cohorts is likely a result of ascertainment bias. The molecular progression of sporadic PDAC is well documented and the high prevalence driver genes are *KRAS* (>90%), *CDKN2A/p16* (95%), *TP53* (50-75%), and *SMAD/DPC4* (55%). We determined the prevalence of alterations in these PDAC driver genes in our cohort of FPC cell lines using an integrated genomics approach of whole exome sequencing (WES), whole genome sequencing (WGS), RNA-Seq, and high density SNP microarrays. We found that our FPC cohort had high frequencies of alterations in the 4 PDAC driver genes. Activating KRAS mutations were identified in all (12/12) FPC cases, predominately at the codon 12 hotspot (11/12, 92%). Inactivation of CDKN2A/p16 was identified in all (12/12) FPC cases, predominately by homozygous deletion (9/12, 75%). TP53 mutations were identified in 92% (11/12), predominately by single base substitutions with loss of heterozygosity (LOH) (9/11, 82%). Inactivation of SMAD4/DPC4 was identified in 75% (9/12) of FPC cases, predominately by homozygous deletion (5/9, 56%). Our collation of the literature confirms that FPC has a similar age of onset as its sporadic counterpart, when ascertainment bias is accounted for. Our annotation of alterations in the dominant 4 PDAC driver genes using an integrated genomic approach, confirms that these 4 genes are also altered in the majority of FPC cases. We conclude that FPC molecular progression parallels that of sporadic PDAC, possibly accounting for the similar age of onset.

3373T

Features of variants called from whole exome sequencing versus transcriptome sequencing in lung cancer. *T. O'Brien[1], P. Jia[2], J. Xia[2], H. Vuong[2], Q. Wang[2], H. Jin[3], U. Sukhija[4], M. Aryee[5], A.J. Iafrate[5], J. Engelman[6], R.S. Heist[6], M. Mino-Kenudson[7], L.P. Le[7], W. Pao[3,8], Z. Zhao[2,9,10].* 1) Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN; 3) Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN; 4) Department of Cancer Biology, Massachusetts General Hospital, Boston, MA; 5) Department of Pathology Services, Massachusetts General Hospital, Boston, MA; 6) Department of Medicine Service, Division of Hematology and Oncology, Massachusetts General Hospital, Boston, MA; 7) Department of Pathology, Massachusetts General Hospital, Boston, MA; 8) Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN; 9) Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN; 10) Department of Cancer Research, Vanderbilt University School of Medicine, Nashville, TN.

Whole exome sequencing (WES) and RNA sequencing (RNA-Seq) are two widely applied next-generation sequencing (NGS) technologies used for detecting somatic mutations in cancer. In this study, we used 27 matched tumor-normal samples that have both WES and RNA-Seq data to systematically compare single nucleotide variants (SNVs) called from WES and RNA-Seq of the same patients. WES reads were mapped to the human reference genome (hg19) using bwa. Post-processing of the initial mapping included steps to mark duplicate reads using Picard and perform local realignment using GATK. RNA-Seq reads were mapped using TopHat 2. We used MuTect to call SNVs for both WES and RNA-Seq, VarScan 2 to determine supporting reads of SNVs, and Cufflinks to compute gene-based expression levels (FPKM) of RNA-Seq data. Interestingly, we found a low overlap rate (on average 14%) of SNVs called from WES and RNA-Seq. Among WES-unique SNVs, 38 - 73% were missed in RNA-Seq due to no coverage (<1 read) and 10 - 23% due to low coverage (<8 reads). Interestingly, 9 - 53% of WES-unique SNVs were highly covered with ≥8 reads in RNA-Seq but were still not called. We next explored WES-unique SNVs' gene expression levels in RNA-Seq. As expected, 55% of WES-unique SNVs were located in unexpressed genes (FPKM<1). However, 3 - 12% of SNVs were in highly expressed regions (FPKM >20) yet remained undetected. We further examined WES-unique SNVs with available cDNA information, and found on average 51% were located on the non-transcribed strand causing them to be undetected in RNA-Seq. We also explored exon skipping as a factor for WES-unique SNVs being missed in RNA-Seq, and preliminary work discovered at least 4 mutations which reside in potentially skipped exons. Among RNA-Seq-unique SNVs, 22 - 89% were in positions not covered by the WES kit. For SNVs covered by the WES kit, 81 - 97% had a callable coverage (≥ 8 reads) in WES. These SNVs were not missed due to technical issues. We therefore analyzed their allele frequency and determined only 3% (0 - 12%) of the alternate alleles occurred with a frequency ≥ 20% in WES. We also analyzed the mutation patterns of all RNA-Seq-unique SNVs. 54% of them displayed a T:A → C:G pattern, which is a signature of potential adenosine deaminase acting on RNA (ADAR) induced RNA-editing occurring in these samples. In summary, this work provides a systematic view of SNV calling using two main NGS platforms.

3374S

Patterns of somatic mutations in hepatitis B virus-associated hepatocellular carcinomas. *Q. Pan[1], W. Yin[1], S. Tong[1], D. Li[2], X. Li[3], D. Zhao[4], M. Li[5], H. Hu[1], H. Ren[1], K. Ding[1].* 1) Key Laboratory of Molecular Biology for Infectious Diseases, Ministry of Educati, Chongqing, China; 2) Department of Hepatobiliary Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, P. R. China; 3) Department of Medical Oncology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, P. R. China; 4) Department of Pathology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, P. R. China; 5) Department of Hepatobiliary Surgery, Suining Central Hospital, Suining, Sichuan Province, P. R. China.

Hepatocellular carcinoma (HCC) is one of the most common malignancies worldwide. Most hepatocellular carcinomas develop in the background of the advanced liver fibrosis and cirrhosis through a stepwise accumulation of various genetic alterations from cirrhotic nodule to HCC. The patterns of somatic mutations in the liver cancer and its matched cirrhotic tissues were not fully characterized. We performed whole exome sequencing on liver cancer tissues, its matched cirrhotic liver tissues and peripheral blood leukocyte (PBL) in 13 patients of hepatitis B virus-associated HCC. Diagnoses of HCC cases and the matched liver cirrhosis were histologically confirmed by two pathologists. Only HCCs with the percentage of tumor cells more than 70% were used for the analysis. Whole exome sequencing resulted in approximate 106 million raw reads with an average coverage of 110× in each sample. A bioinformatics pipeline was used to analyze sequencing reads and identify somatic mutations. The prevalence of somatic mutations differed significantly among HCCs. Among 6,972 non-silent mutations (i.e., missense, nonsense mutations, splice sites and translation start sites) identified in 13 HCC patients, we found that T:A>A:T transversion showed the highest percentage of non-silent mutations, followed by C:G>T:A transversion and C:G>A:T transversion. By comparison of somatic mutations identified in liver cancer tissues and cirrhotic tissues, we found that somatic mutations were not accumulated in genes in HBV-associated cirrhotic liver tissues. Notably, two genes (TP53 and KRTAP4-3) were shown to be significantly mutated by MutSigCV analysis, suggesting that there were evidence for 'driver' mutations of these genes contributing to the development of liver cancer. Our study may provide an overview of patterns of somatic mutation in hepatitis B-virus associated HCC and its precancerous lesions (i.e., cirrhosis).

3375M

Highly sensitive, non-invasive detection of colorectal cancer mutations using single molecule, third generation sequencing. *G. Russo[1], A. Patrignani[1], L. Poveda[1], F. Hoehn[2], B. Scholtka[3], R. Schlapbach[1], A. Garvin[2].* 1) Functional Genomics Center Zurich - ETH/UZH, Zurich, Switzerland; 2) Droplet Diagnostics SAS, Mulhouse, France; 3) Department of Nutritional Toxicology, Institute of Nutritional Science, University of Potsdam, Nuthetal, Germany.

Colorectal cancer (CRC) represents one of the most prevalent and lethal malignant neoplasms and every individual of age 50 and above should undergo regular CRC screening. Currently, the most effective procedure to detect adenomas, the precursors to CRC, is colonoscopy, which reduces CRC incidence by 80%. However, it is an invasive approach that is unpleasant for the patient, expensive, and poses some risk of complications such as colon perforation. A non-invasive screening approach with detection rates comparable to those of colonoscopy has not yet been established. The current study applies Pacific Biosciences third generation, single molecule sequencing to the inspection of CRC-driving mutations. Our approach combines the screening power and the extremely high accuracy of circular consensus (CCS) third generation sequencing with the non-invasiveness of using stool DNA to detect CRC-associated mutations present at extremely low frequencies and establishes a foundation for a non-invasive, highly sensitive assay to screen the population for CRC and early stage adenomas. We performed a series of experiments using a pool of fifteen amplicons covering the genes most frequently mutated in CRC (APC, Beta Catenin, KRAS, BRAF, and TP53), ensuring a theoretical screening coverage of over 97% for both CRC and adenomas. The assay was able to detect mutations in DNA isolated from stool samples from patients diagnosed with CRC at frequencies below 0.5 % with no false positives. The mutations were then confirmed by sequencing DNA isolated from the excised tumor samples. Our assay should be sensitive enough to allow the early identification of adenomatous polyps using stool DNA as analyte. In conclusion, we have developed an assay to detect mutations in the genes associated with CRC and adenomas using Pacific Biosciences RS Single Molecule, Real Time Circular Consensus Sequencing (SMRT-CCS). With no systematic bias and a much higher raw base-calling quality (CCS) compared to other sequencing methods, the assay was able to detect mutations in stool DNA at frequencies below 0.5 % with no false positives. This level of sensitivity should be sufficient to allow the detection of most adenomatous polyps using stool DNA as analyte, a feature that would make our approach the first non-invasive assay with a sensitivity comparable to that of colonoscopy and a strong candidate for the non-invasive preventive CRC screening of the general population.

3376T

Integrating eQTLs from a range of normal human tissues with cancer genomics to help identify germline risk alleles in cancer driver genes. *A.V. Segre[1], D.S. DeLuca[1], T. Sullivan[1], E. Gelfand[1], L. Lappalainen[2,3], M. Lawrence[1], A. Kiezun[1], D.G. MacArthur[1,4], K. Ardlie[1], G. Getz[1,4], the GTEx consortium.* 1) Broad Institute of Harvard and MIT, Cambridge, MA; 2) New York Genome Center, NY; 3) Department of Systems Biology, Columbia University, NY; 4) Massachusetts General Hospital, Boston, MA.

A comprehensive analysis of somatic point mutations in whole exome sequences from >4,700 human cancer and their matched normal tissue samples has increased the number of cancer driver genes to 260 for 21 different tumor types. The extent to which germline variants in these somatic cancer genes may predispose people to certain cancers is not yet known. Furthermore, the contribution of gene expression-altering variants to tumor development compared to changes in protein structure or function is not well understood. To address these questions and gain mechanistic insight into the potential role of germline variants in tumorigenesis, we are testing the hypothesis that increase in expression of oncogenes and/or decrease in expression of tumor suppressor genes in normal human tissues may increase risk of cancer development. To this end, we will annotate the cancer genes as potential oncogenes and/or tumor suppressor genes, based on large-scale experimental gene perturbation studies that measure the effect on cell proliferation, growth arrest or cell death. We will then examine which cancer genes are affected by cis-eQTLs (expression quantitative trait loci) or allele specific expression (ASE), using the Genotype-Tissue Expression (GTEx) project with data in 40 normal human tissues and up to 450 unrelated individuals. Based on the pilot phase of GTEx with 178 individuals, 36% of the 260 somatic cancer genes, e.g. AKT1, CASP8 and CDKN2A have significant cis-eQTLs at FDR<5% in at least one of 9 tissues tested, including skin, lung and nerve. Following characterization of cancer genes with eQTLs, we will test whether the germline genotypes of tumor samples in the pan-cancer cohort, such as melanoma, lung cancer, and blood cancers, are enriched for eQTL or ASE alleles that increase oncogene expression or decrease tumor suppressor expression compared to a control panel. If so, this would suggest that the regulatory variants are cancer risk factors. Furthermore, we will integrate the results with somatic mutations from whole exome and whole genome sequences in the pan-cancer project and loss of heterozygosity (LOH) analyses, to evaluate how eQTLs acting on cancer genes may be contributing to tumor progression. This study may help propose new cancer germline risk variants and shed light on potential causal regulatory mechanisms in cancer. In future work, such integrative studies can be extended to testing the regulatory role of somatic mutations during tumorigenesis.

3377S

Identification of mutations in oral cavity squamous cell carcinoma induced by betel quid chewing in Taiwan. *Y. Shih[1], L. Wang[4], J. Chen[5], N. Limthong[7], H. Chen[6], D. Chen[3, 7], Y. Elshimali[2].* 1) La Sierra university, Riverside, CA; 2) Charles Drew Univ Med. Los Angeles. CA; 3) Dept of Path and Lab Med David Geffen School of Medicine at UCLA. Los Angeles. CA; 4) Rite Aid Pharmacy, Fontana, CA; 5) Rosemead Clinic, Rosemead, CA; 6) Maternal & Child Center, Taipei, Taiwan; 7) Dept of Path and Lab UCI Medical Center, Orange, CA.

Squamous cell carcinoma of the head and neck (HNSCC) is the sixth most common cancer worldwide, the most common site being the oral cavity. In Taiwan, squamous cell carcinoma of the oral cavity is the fourth most common type of cancer. Environmental carcinogens, including betel quid chewing and alcohol, are the major identifiable risk factors. The development of squamous cell carcinoma of the oral cavity arising from premalignant lesions, such as leukoplakia, follows a multi-step process involving sequential activation and inactivation of oncogenes and tumor suppressor genes, resulting in a clonal population of cells. Chromosomal deletions at 3p and 9p21 are characterized as the first step of the tumorgenesis process. However, additional genomic abnormalities are well characterized. Such abnormalities can contribute to histological variations often observed in a tumor sample, such as invasiveness and, more specifically, the ability to penetrate the basement membrane. To better characterize any chromosomal structural abnormality, we employed next-generation DNA sequencing technology on a patient sample with squamous cell carcinoma of the tongue. The patient has a known history of betel quid chewing. Specifically, we isolated DNA samples from normal, premalignant (leukoplakic), in situ and invasive components of the tumor. A mate-pair library was constructed from each of the four DNA samples. Individual DNA samples were sequenced using a next-generation Illumina H2000 sequencer. By analyzing the sequences, we have identified a novel chromosomal structural abnormality, as well as known chromosomal variations, including deletions of 3p and 9p21, and a deletion at the genomic area containing the p53 gene. In addition, several mutations specific to the invasive component of the tumor were identified. Our findings confirm the previously known fact that tumorgenesis can result in the evolution of subclonal populations of tumors.

3378M

Subclonal evolution and genomic drivers of relapse in childhood acute lymphoblastic leukemia. *JF. Spinella[1], V. Saillour[1], N. Garnier[1], R. Vidal[1], J. Healy[1], D. Sinnett[1,2].* 1) Hematology Oncology, CHU Sainte-Justine (Montreal University), Montreal, Quebec, Canada; 2) Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, Qc, Canada.

Refractory childhood acute lymphoblastic leukemia (cALL) has a dismal outcome with overall long-term survival rates around 45%, making it the leading cause of death by disease among children. We now know that intratumoral genetic heterogeneity and clonal evolution contribute to treatment resistance. In order to build a catalogue of somatic-specific variations and identify genomic drivers of relapse in cALL, wWe hypothesized that relapse-specific single nucleotide variants (SNVs) or showing a significant gain of allele frequency in the relapse material compared with the matched-primary tumor, are putative drivers of ALL relapse performed whole-genome-sequencing on 10 matched primary tumor (diagnosis), relapse and normal (remission) cALL patient trios. We hypothesized that putative relapse drivers would either be relapse-specific (detected in relapse samples only) or show a significant shift in variant allele frequency in the relapse versus matched-primary tumour samples, due to clonal evolution. Combining deep re-sequencing with genome-wide genotyping (to determine local ploidy), allowed us to accurately call single nucleotide variants (SNVs) with a wide range of frequencies, including SNVs from minor subclonal populations. Changes in the fitness of somatic SNVs (rise or fall of allele frequencies) were assessed and clustering analysis was used to illustrate subclonal complexity and clonal evolution within the individual ALL tumours. Strikingly, our data suggest the presence of a fitter ancestral subclonal population, harbouring specific relapse-drivers, that is present at diagnosis and that rises during relapse. In a minority of tumours the major diagnostic clone was resistant to therapy and was maintained at relapse. We queried putative relapse drivers against public pharmacogenomics databases (DGIdb, PharmGKB) and identified known drug-gene/pathway interactions which could lead to new avenues for improved targeted treatment of refractory cALL. By dissecting the clonal evolution of individual tumors and identifying the underlying genetic events driving cALL relapse, we hope to improve prognosis for refractory cALL and ultimately provide targeted personalized care for children diagnosed with ALL.

3379T

Integrative analysis of regulatory aberrations in lung adenocarcinoma cell lines. *A. Suzuki[1], T. Kohno[2,3], K. Tsuchihara[2], Y. Suzuki[1].* 1) Department of Medical Genome Sciences, The University of Tokyo, Japan; 2) Division of TR, The Exploratory Oncology Research and Clinical Trial Center, National Cancer Center, Japan; 3) Division of Genome Biology, National Cancer Center Research Institute, Japan.

Integrative multi-omics analysis is a powerful approach to understand the biological relevance of the genomic, epigenomic and transcriptomic aberrations in cancer. In this study, we performed whole-genome sequencing, RNA-Seq, target-captured bisulfite sequencing and ChIP-Seq for eight chromatin marks using 26 lung adenocarcinoma cell lines. By this approach, we conducted a multi-omics analysis of cancer cells, which still remain technically difficult in clinical samples.

Using whole-genome sequencing data, we identified a mean of 536 non-synonymous SNVs and indels in the protein-coding regions and 13,573 mutations in the promoters and enhancers. In representative tumor-suppressor genes, we detected splice-site mutations which cause aberrations of splicing patterns in RNA sequences. For transcriptome analysis, we examined gene expression abundances using RNA-Seq data and detected an average of 352 higher and 1,967 lower differentially expressed genes such as EGFR and SMARCA4. We also detected 135 fusion transcript candidates including known ALK and RET fusion transcripts. We analyzed DNA methylation statuses using bisulfite sequencing and identified an average of 118 higher and 278 lower differential DNA methylated genes such as the IGF1R, NRAS and CDKN2A (p16) genes. We also analyzed ChIP-Seq data for seven histone modification (H3K4me1, H3K4me3, H3K9me3, H3K9/14ac, H3K27ac, H3K27me3 and H3K36me3) and RNA Polymerase II, which were further subjected to ChromHMM analysis to identify eight chromatin states. To publicly open and visualize the multi-omics data, we have also developed a database (http://dbtss.hgc.jp/).

In this study, we particularly focused on the genomic and epigenomic aberrations affecting eventual gene expression levels. We detected different patterns of aberrations depending on genes. For example, genomic alterations were characteristic to the STK11. On the other hand, expression levels in the CDKN1A were repressed by DNA methylation or repressive marks of histone modification (H3K27me3 and H3K9me3). By complementing whole-genome and exome sequencing data of clinical samples, the multi-omics analysis using cell lines should be important to understand how variable regulatory aberrations result in transcriptional alterations in cancer cells.

3380S

Molecular profiling in diagnosis and determining prognosis of "early" myelodysplastic syndrome. *M. Thangavelu[1], S. Brodie[2], C. Mixon[3], S. Agersborg[1], E. Wei[2], M. Albitar[1].* 1) Neogenomics Laboratories, Irvine, California; 2) Neogenomics Laboratories, Ft. Myers, Florida; 3) Neogenomics Laboratories, Nashville, Tennessee.

Confirming diagnosis of myelodysplastic syndrome (MDS) can be very difficult, especially in the early stages of the disease. Cytogenetic abnormalities are detected by conventional cytogenetics (and FISH) in approximately 50% of cases of MDS and are used to confirm diagnosis. Identification of mutations in RNA splicing, DNA methylation, chromatin modification, transcription regulation, DNA repair, signal transduction, and cohesin complex genes has heightened the interest of utilizing these characteristics in the diagnosis of MDS. Most published data on molecular characterization are retrospective studies. We report the use of a panel of 14 genes in a prospective fashion as an objective means for confirming the presence of MDS rather than a reactive process. 133 patients with cytopenia involving at least one lineage, no evidence of cytogenetic abnormalities and blast count less than 5% were included in the study. DNA from bone marrow or peripheral blood specimens were analyzed for mutations in the following genes: ASXL1, ETV6, EZH2, IDH1, IDH2, NRAS, CBL, RUNX1, SF3B1, SRSF2, TET2, TP53, U2AF1 and ZRSR2. Direct bidirectional Sanger sequencing was used in all patients and next generation sequencing was used in some patients. Fifty five of the 133 patients (41.4%) had mutation in 1 to 4 genes. Mutations were seen in TET2 (23 cases), ASXL1 (15 cases), SF3B1 (15 cases), SRSF2 (10 cases), U2AF1 (6 cases), ZRSR2 (5 cases), ETV6 (3 cases), and IDH2, RUNX1, NRAS and EZH2 (2 cases each), and CBL, IDH1 and TP53 (1 case each). TET2 mutations were the most common occurring in was mutated in 17.3% of cases studied. Although only 41.4% of the tested patients had a mutation confirming the diagnosis of MDS, most patients without mutations had no diagnostic morphologic evidence of MDS. All patients were suspected of having MDS based on cytopenia(s) at the time of testing only, without documented previous history of cytopenia. Longer follow up is needed to rule out MDS and confirm a reactive process in patients without a mutation. These results suggest that utilizing an appropriately designed, relatively small, molecular panel can provide valuable objective diagnostic means for the diagnosis of MDS in patients without cytogenetic abnormalities. Furthermore, such a panel provides very important prognostic information that can be used for stratifying patients and determining therapeutic approach.

3381M

The Mutation Profiles of JAK2, MPL, CALR, LNK, CBL, ASXL1 and DNMT3A genes in BCR/ABL1 and JAK2V617F Negative Myeloproliferative Neoplasms. *B. Türkgenç[1, 2], E. Pinarbasi[3], S. Ratip[4], A. Uzay[4], A. Ozer[2], C. Akyerli Boylu[5], M.C. Yakicier[1, 5].* 1) Acibadem Genetic Diagnostic Center, ISTANBUL, Turkey; 2) Department of Medical Biology and Genetics, Marmara University, Istanbul, TURKEY; 3) Department of Medical Biology, Cumhuriyet University, Sivas, TURKEY; 4) Department of Hematology, Acibadem Kadikoy Hospital, Istanbul, TURKEY; 5) Biochemistry and Molecular Biology, Department of Acibadem University, Istanbul, TURKEY.

Background: Myeloproliferative neoplasms (MPNs) are clonal myeloid cancers characterized by overproduction of mature blood cells. BCR/ABL1 and JAK2V617F are the most frequent mutations in MPNs. Other JAK2-exon12 and MPL-exon10 mutations were less common in suspected MPNs (2% and 3-7%, respectively). Several other genes such as LNK, CBL, ASXL1, DNMT3A and TET2 have been found to be responsible for small minority of MPNs. Recently, in the majority of MPN patients, CALR mutations have been identified. Purpose: The aim of our study is to establish the mutation profiles of BCR/ABL1 and JAK2V617F negative MPNs. Methods: 152 non-mutated BCR/ABL1 and JAK2 V617F patients were selected among a thousand patients who were referred from different centers to Acibadem Diagnostic Center, from May 2007 to February 2014. Genomic DNA and RNA samples were isolated from bone marrow and peripheral blood samples. RNA samples were used for detection of JAK2-exons 11-17 mutations and MPL-exon 10 mutations by RT-PCR and nested PCR followed by Sanger sequencing. DNA samples were used to detect the frequent mutations in other reported genes; CALR-exon9, CBL-exon8-9, DNMT3A-exon23, ASXL1-exon13 and LNK-whole exons by PCR and DNA sequencing. Results: We identified 11/152 JAK2-exon11-17 (7,2%) and 6/152 MPL-exon10 mutations (3,9%). Although, all MPL mutations were reported previously, 6/11 JAK2 mutations were novel. We found 17/152 CALR mutations (11,2%); while L367fs*46 (8/17) and K385fs*47 (5/17) were the most common variants, the remaining 3 out of 4 variants were novel. Also, we identified 3 novel mutations all in the coding sequence of LNK (2%). In CBL, 2 patients had the same splice site variant in intron 8 (1,3%). We identified 10/152 ASXL1 mutations (6,6%) 2 of which were novel. None of the patients had DNMT3A mutation. Conclusion: To the best of our knowledge, this is the first mutation profile study associated with BCR/ABL1 and JAK2V617F negative MPNs in Turkish population. Our study enabled to improve the screening strategy of MPNs and provided genetic diagnosis of 49 of the selected 152 BCR/ABL1 and JAK2V617F negative MPNs. We observed several novel mutations as well as previously reported ones. Mutation profiles will contribute to the development of therapeutic drugs (such as tyrosine kinase inhibitors-TKIs) and diagnostic strategies. Functional studies should be carried out in order to reveal the mechanism underlying these unique variants.

3382T

Whole genome sequencing of high-risk families to identify new mutational mechanisms of breast cancer predisposition. *T. Walsh, S. Gulsuner, S. Casadei, M.K. Lee, J. Mandell, M-C. King.* Division of Medical Genetics, University of Washington, Seattle, WA.

Multiple genes in addition to *BRCA1* and *BRCA2* are known to harbor mutations that significantly increase risk of breast cancer. The increasing use of comprehensive cancer gene panels allows women to be tested for all known breast cancer susceptibility genes in a single assay. For families severely affected with breast cancer, but without mutations in any known breast cancer gene, exome sequencing has been applied to identify new candidate genes, with a somewhat disappointing yield. In our experience, approximately 50% of families with four or more relatives with breast cancer remain unresolved after comprehensive panel and exome sequencing. We hypothesize that in many of these families, breast cancer is due to individually rare alleles of moderate-to-severe effect located in the non-coding regions of the genome. A subset of these are likely influencing their mutational effect via alterations on expression or regulation of known breast cancer genes. We selected pairs of cousins with young-onset breast cancer from each of 20 very severely affected kindreds. Whole genome sequencing was carried out on genomic DNA of the 40 individuals. We are filtering variants shared by the members of a pair with the following criteria: (i) rarity, (ii) within 5 MB of a known breast cancer gene, (iii) conservation of the site harboring the variant, and (iv) annotation of potential regulatory function. In parallel, we are performing targeted RNAseq of all known breast cancer genes from lymphoblast derived RNA of the same individuals (see abstract by Casadei S et al.). RNAseq data may reveal genes with asymmetrically expressed alleles in a family, thereby focusing regions of interest in the whole genome data from that family. Candidate variants will be tested for co-segregation with breast cancer in the host kindred. Our long-term goal is to integrate newly identified mutational mechanisms into clinical practice.

3383S

DICER1 mutations occurring in childhood anaplastic sarcoma of kidney. *M. Wu[1], H. Druker[2], P. Thorner[3], J. Traubici[4], R. Grant[3], S. Albrecht[5], E. Weber[6], A. Charles[7,8], J.R. Priest[9], M.R. Fabian[10], N. Watanabe[11], G. Vujanic[12], W.D. Foulkes[1,6,13].* 1) Medical Genetics, Lady Davis Institute, McGill University, Montreal, Quebec, Canada; 2) Cancer Genetics Program, The Hospital for Sick Children and Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 3) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; 4) Department of Diagnostic Imaging, Hospital for Sick Children and University of Toronto, Toronto, ON, Canada;; 5) Department of Pathology, Montreal Children's Hospital, McGill University Health Centre, Montreal, QC, Canada; 6) Research Institute of the McGill University Health Centre, Montréal, QC, Canada; 7) School of Paediatrics and Child Health (SPACH), University of Western Australia (M561) Crawley, Western Australia; 8) Department of Paediatric Pathology, Princess Margaret Hospital for Children, Perth, Western Australia; 9) Minneapolis, MN, USA;; 10) Departments of Oncology and Experimental Medicine, McGill University, Montréal, QC, Canada; 11) Department of Pathology, Nihon University School of Medicine, Tokyo, Japan; 12) Institute of Cancer & Genetics, Cardiff University School of Medicine, Cardiff, UK; 13) Program in Cancer Genetics, Departments of Oncology and Human Genetics, McGill University, Montréal, QC, Canada.

Anaplastic sarcoma of the kidney (ASK) is an extremely rare renal neoplasm with fewer than 30 cases reported. We report two cases of children with ASK. The first patient (Case 1) carried a germline *DICER1* mutation (c.2062C>T) which is predicted to cause a nonsense mutation (p.R688X) and her ASK bore a somatic in trans missense *DICER1* mutation (c.5425G>A) which is predicted to cause an amino acid substitution at a critical residue in the catalytic domain of RNase IIIb (p.G1809R). In addition, Case 1 had presented prior to the ASK with a pleuropulmonary blastoma (PPB) which bore a different somatic mutation in *DICER1* (c.5125G>A; predicted p.D1709N). The ASK in Case 2 also possessed a somatic *DICER1* RNase IIIb mutation (c.5125G>A; predicted protein change p.D1709N). The germline *DICER1* mutation status of Case 2 is pending. Thus, the somatic mutations in both ASKs are in a DICER1 catalytic domain (RNase IIIb) and are predicted to affect the production of the "5p" class of microRNA by DICER1 protein. A previous report identified several cases of "DICER1 renal sarcoma", two of which had germ-line and/or somatic *DICER1* mutations. Taken together, we hypothesize that these tumors and ASKs currently described are the same entity. Our findings, therefore, build upon previous observations and suggest that *DICER1* mutations may be an important cause of these rare tumors.

3384M

Single-cell mutation detection with multiplex PCR-based targeted enrichment sequencing. *Z. Wu, R. Vijaya Satya, R. Samara, Y. Wang, E. Lader.* Biological Research Content, QIAGEN Sciences, LLC, Frederick, MD, USA.

Mutations are the driving molecular causes of various biological processes such as development and cancer. Recent findings of genomic heterogeneity among ostensibly homogeneous cell populations such as cancer cells demand genomic characterization of mutations at the individual cell level to better understand the underlying biology. Additionally, single-cell technologies make genomic analysis feasible for the characterization of rare cells, such as circulating tumor cells and in-vitro fertilized embryos. Due to its high sensitivity, next-generation sequencing (NGS) represents the ideal technology to analyze a collection of mutations in single cells. The challenge, however, is the limited amounts of DNA, which need to be amplified prior to NGS. To overcome this challenge, whole genome amplification (WGA), coupled with multiplex PCR-based targeted enrichment, was tested for mutation detection in single cells isolated from two colon cancer cell lines, Lovo and HT29. A NGS panel targeting genes relevant to colon cancer was used to enrich DNA from WGA single-cell DNA (WSC), DNA from bulk cells (BC), or bulk-cell DNA which underwent WGA (WBC), followed by mutation detection with NGS. Forty (40) ng DNA was used across 4 PCR pools in a multiplex PCR set up to enrich for 38 genes known to harbor colon cancer-relevant mutations. The amplicons went through a standard library construction protocol, followed by sequencing on a MiSeq®. Previously reported mutations such as KRAS (G13D) in LoVo and BRAF (V600E) in HT29 cells were successfully detected with this enrichment method on WSC as well as on BC. These mutations were further confirmed by Pyrosequencing. A strikingly similar mutation frequency was observed between targeted NGS and Pyrosequencing in each sample. Interestingly, mutation frequencies observed between BC and WBC were similar, but varied greatly with respect to individual cells. Similar variations were observed on additional heterozygous mutations in WSC as compared to BC. Whether the variation is due to tumor cell heterogeneity, selective WGA amplification, or both needs to be further defined. These results demonstrate that multiplex PCR targeted enrichment can be successfully applied for mutation detection on WSC. However multiple single cells need to be analyzed for accurate interpretation of genomic alterations among heterogeneous cell populations. The applications presented here are for research use only. Not for use in diagnostic procedures.

3385T

Patient-oriented functional genomics analysis of p53 mutations in cancer. *O. Zill[1], T. Shamu[1], S. Fields[2,3], B.S. Taylor[1].* 1) Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA; 2) Departments of Genome Sciences and Medicine, University of Washington, Seattle, WA; 3) Howard Hughes Medical Institute, Seattle, WA.

Over the last few years, thousands of cancer genomes have been sequenced and hundreds of recurrent cancer-associated mutations have been identified. A defining challenge for cancer genetics is to determine which mutations in a patient's tumor genome are biologically or clinically relevant and how two or more driver mutations co-occurring in that tumor might interact. We are using multi-dimensional computational analysis of cancer genomics data combined with human somatic-cell genetics experiments to address these questions. We have focused initially on the most frequently mutated cancer gene, TP53, because of the wealth of experimental data available for many mutations and to provide sufficient sample numbers to power association tests across many tumor types. Using a compendium of somatic mutations from the sequenced exomes and genomes of over 10,000 human tumors, we assessed the co-occurrence of codon-specific TP53 mutations with those in 188 significantly mutated genes in cancer. Co-mutation patterns fell into several distinct categories that were partly explained by tumor-type-specific TP53 mutational patterns. Interestingly, certain missense TP53 mutations preferentially co-occurred with mutations in specific oncogenes or tumor suppressors. To determine the relative functionality of these specific-SMG-enriched TP53 mutations, we are introducing them into human cell lines and competing them against each other in FACS-based transcriptional reporter assays. For these mutant competition assays, we constructed transcriptional reporters using the promoters of p53 target genes whose pattern of expression from RNA sequencing data across 12 tumors types was associated with TP53 mutation status. An example of a mutation co-occurrence we detected was the hotspot alleles IDH1-R132H and TP53-R273C, which co-occurred predominantly in Lower Grade Glioma samples. In parallel to the mutant competition assays, We are testing whether IDH1-TP53 co-mutation contributes to specific, cancer-relevant phenotypes by introducing the mutations into human cell lines. The mutation combinations we identified are being evaluated for associations with alterations in copy number, gene expression, and methylation in tumors. This work should help delineate the functional contributions of specific combinations of driver mutations to inform personalized cancer treatments.

3386S

Expression of genomic somatic mutations at the levels of transcriptome and proteome in a patient of hepatocellular carcinoma with MSH2 haploinsufficiency. *K. Ding[1], S. Wu[2], W. Ying[2], W. Yin[1], S. Tong[1], Q. Pan[1], X. Li[3], D. Zhao[4], H. Hu[1], X. Qian[2], H. Ren[1].* 1) Key Laboratory of Molecular Biology for Infectious Diseases, Ministry of Educati, Chongqing, China; 2) State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing, P.R. China; 3) Department of Medical Oncology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, P. R. China; 4) Department of Pathology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, P. R. China.

The expression of genomic somatic mutations, i.e., at the levels of transcriptome and proteome, were not fully characterized in cancer genome sequencing studies. It is unknown how genomic mutations were transmitted from DNA to messenger RNA, and finally affected its function of protein. A patient of hepatitis-B associated hepatocellular carcinoma that showed an extremely elevated prevalence of somatic mutations with a heterozygous nonsense mutation in the mutS homolog 2 (MSH2) gene (leading to MSH2 haploinsufficiency) was present in Peking Union Medical College Hospital. Using this patient as a model, we performed whole exome sequencing, transcriptome sequencing, and proteome profiling on liver cancer tissues and its matched liver cirrhotic tissues (i.e., adjacent 'normal' tissues). By whole-exome sequencing, we identified 20,382 genomic somatic mutations (single nucleotide variants, SNVs), and 4,998 were classified as non-silent mutations. For non-silent somatic mutations, we found that, 1,823 somatic mutations (36%) in genes were not expressed or covered, 1,892 (37.8%) where only the wild-type allele were expressed, 1199 (24.0%) where both alleles were expressed, and 84 (1.7%) where only the mutant allele were expressed by transcriptome sequencing. At the levels of proteome, we identified 118 mutations whose mutated peptides are identified in the liver cancer tissues, and their un-mutated peptides were identified in the liver cirrhotic tissues. These results suggested that only a small fraction of genomic somatic mutations were expressed, and leveraging 'trans-omics' strategies may provide important implications for conducting cancer-genome sequencing studies.

3387M

***APC* promoter 1B deletion in seven American families with familial adenomatous polyposis.** *A.K. Snow[1], T.M.F. Tuohy[1], N.R. Sargent[1], L.J. Smith[1], R.W. Burt[1,2,3], D.W. Neklason[1,2,4].* 1) High Risk Clinical Research, Huntsman Cancer Institute, Salt Lake City, UT 84112; 2) Department of Medicine, University of Utah, Salt Lake City, UT 84112; 3) Division of Gastroenterology, University of Utah, Salt Lake City, UT 84112; 4) Division of Genetic Epidemiology, University of Utah, Salt Lake City, UT 84112.

Familial adenomatous polyposis (FAP [MIM 175100]) is a colorectal cancer predisposition syndrome caused by mutations in the adenomatous polyposis coli (*APC* [MIM 611731]) gene. University of Utah has a large Hereditary Gastrointestinal Cancer Registry with over 600 FAP patients from 244 kindreds. A clear deleterious mutation is known in 122 of these kindreds, but in the remaining kindreds, *APC* genetic testing has either failed to detect a mutation (16%) or has not been pursued (40%). We have applied a simple and cost-effective approach to screen for families who may harbor identical pathogenic mutations due to a common founder. Members of our registry with FAP have been genotyped using a set of 4 short tandem repeat markers (STR) with high heterozygosity across the chr.5 *APC* locus. Following the inclusion of probes specific for *APC* promoter 1B, multiplex ligation-dependent probe amplification (MLPA) identified a proband from our registry with a deletion of promoter 1B. Six additional families shared the *APC* haplotype and the deletion, which is distinct from previously reported promoter 1B deletions. The clinical phenotype of 17 mutation carriers is classic colonic polyposis with colectomy at an average age of 24. The majority report having a large number of duodenal and gastric polyps, and there were single reports of a desmoid and a hepatoblastoma. Measurements of allele-specific expression of *APC* mRNA using TaqMan assay specific to a common polymorphism, rs459552, on cDNA from 3 tissue sources confirmed that the relative expression was reduced in the allele containing the promoter 1B deletion. The relative reduction in expression of the mutant allele was more pronounced in mRNA extracted from duodenal tissue (88%) and blood (98%) than in mRNA extracted from EBV transformed lymphoblast cell lines (42%). Previous work has shown down regulation of the alternative *APC* promoter 1A through methylation. This methylation, along with deletion of promoter 1B, is thought to be the underlying mechanism for the allelic imbalance in primary tissues. This study confirms the importance of *APC* promoter deletions as a cause of FAP. Alternations in promoters and other regulatory elements should be considered critical targets of future genetic research as a cause of syndromes. This study also highlights the utility of using STR analysis to identify patients lacking a genetic diagnosis by linking them to a genetic haplotype of a known deleterious mutation.

3388T

Therapy-related Acute Myeloid Leukemia Transformed from Juvenile Myelomonocytic Leukemia with Loss of a PTPN11 Somatic Mutation. *Y. Kim[1], J. Kim[2], K-A. Lee[3].* 1) Department of Laboratory Medicine, Samkwang medical laboratories, Seoul, Seoul, South Korea; 2) Department of Laboratory Medicine, Yonsei University Wonju College of Medicine, Wonju, Korea; 3) Department of Laboratory Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea.

Therapy-related myelodysplastic syndrome and acute myeloid leukemia (t-MDS/t-AML) is a well-recognized clinical syndrome. However, therapy-related juvenile myelomonocytic leukemia (t-JMML) induced by prior chemotherapy for treatment of primary neoplasm has not previously been reported. Here, we report the first known case of a t-JMML with PTPN 11 somatic mutation and monosomy 7 developing during chemotherapy for retinoblastoma with RB gene mutation, which, during treatment for JMML, progressed into t-AML harboring monosomy 7 but not the PTPN11 mutation initially seen at the time of diagnosis of JMML. The patient was born at 39 weeks and was a healthy female at birth. At 5 months, she was diagnosed with sporadic bilateral retinoblastoma. After diagnosis, she received a total of 26 courses of chemotherapy from June 2010 to August 2012. During the follow-up, complete blood count (CBC) showed a hemoglobin (Hb) level of 6.1 g/dL, a platelet count of 28,000/L, and a WBC count of 8,150/L with immature granulocytes, January 2013. The peripheral blood monocyte count had significantly increased, to 1.98 x 106/L (45%), and the findings of touch imprint of bone marrow biopsy were most consistent with JMML, with less than 20% blasts in the bone marrow aspirate. PTPN11 mutation (c.215C>A, p.Ala72Asp) was detected, but no mutation was found in codons 12, 13, and 61 of KRAS and NRAS. Monosomy 7 was observed in all 20 of 20 metaphases. The patient was diagnosed with JMML in February 2013, and treated with 6-mercaptopurine (50 mg/m2) and cis-retinoic acid (100 mg/m2). After 4 months, bone marrow examination revealed development of t-AML with 25.0% myeloblasts among all nucleated cells. At the time of t-AML diagnosis, the pathogenic mutation (p.Ala72Asp) of PTPN11 initially observed at the time of t-JMML diagnosis was not detected. Monosomy 7 was still observed in all 20 of 20 metaphases. We evaluated the known secondary mutations (SETBP1 and JAK3) which are detected with relatively high frequency in JMML. Based on the hypothesis that new mutations might develop with leukemic transformation and loss of PTPN11 mutation, mutations in other genes have been sought, but no mutations in the SETBP1 or JAK3 genes were detected in our case. We propose that PTPN11 mutation may be associated with t-JMML, but in the present case, other factors including monosomy 7, and cytotoxic chemotherapy might contribute to the leukemic transformation.

3389S

High-resolution characterization of a leiomyoma on Mayer-Rokitansky-Kuster-Hauser syndrome. *Y. Wu[1], M. Kan[2], Y. Wu[3], Z. Zhang[4], D. Zhou[4], Y. Liu[5], H. Huang[1].* 1) The International Peace Maternity & Child Health Hospital of China welfare Institute, Shanghai Jiao Tong University, Shanghai, 200030, China; 2) Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Graduate School of the Chinese Academy of Sciences, Shanghai, 200031, P. R. China; 3) School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China; 4) Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai, 200030, P.R. China; 5) Institutes of Biomedical Sciences, Fudan University, Shanghai, 200032 P. R. China.

Mayer-Rokitansky-Kuster-Hauser (MRKH [MIM 27700]) syndrome is characterized by Mullerian duct aplasia (uterus plus upper two-thirds of the vagina) with female phenotype and a 46, XX karyotype. MRKH is infrequent thus uterine leiomyoma of the rudimentary uterus in MRKH is an extremely rare case. We performed whole-genome sequencing (WGS) for uterine leiomyoma, rudimentary uterus and peripheral blood from a patient with MRKH syndrome, and conducted analyses for chromosomal structure variations (SVs), copy number variations (CNVs) as well as point mutations. We further performed array comparative genomic hybridization (CGH) for rudimentary uterus and peripheral blood to validate CNVs from the WGS findings. Our results confirmed 8q23.1 germline deletion in MRKH patient and we also observed several somatic CNVs in leiomyoma including deletions in 1q24.2, 4q34.3 and 19q13.33. Interestingly, the number of CNVs in rudimentary uterus showed more abundant than that in leiomyoma, suggesting that leiomyoma might be initiated from undifferentiated mesenchymal cells or immature smooth muscle cells instead of mature smooth muscle cells in rudimentary uterus. We also detected two translocation events t(16:21) and t(11:x) in rudimentary uterus and three translocation events t(1:7), t(2:10) and t(10:16) in uterine leiomyomas. Additionally, many somatic point mutations were newly identified in leiomyoma with probably damaging function involved in other type of cancers or tumors. Our study initially characterized rudimentary uterus and uterine leiomyoma from whole genome level, which may provide insight into development of rudimentary uterus in MRKH syndrome and pathogenesis of leiomyoma in further studies.

3390M

Comparison of NGS solutions for rapid and cost-effective analysis of degraded FFPE and cancer biobanked specimens with limited quantity. *A. Brooks[1], M. Moreau[1], I. Liang[1], A. Sahota[1], K. Hirschfield[2], L. Rodriguez[2], H. Zhong[2], J. Tischfield[1], R. DiPaola[2], S. Ganesan[2].* 1) RUCDR Infinite Biologics, Rutgers University, Piscataway, NJ; 2) Rutgers Cancer Institute of New Jersey, Rutgers University, Piscataway, NJ.

In order to unlock genetic sequence information hidden in millions of tissue biopsies and cancer biobanks around the world, scientists require a robust Next-Generation Sequencing (NGS) platform that offers consistent, cost-effective, high quality and high coverage results. We examined the utility of two of the leading NGS content enrichment methods with key evaluation criteria including input material needs, sequencing coverage specificity and sensitivity, workflow ease and robustness, scalability, and all-in total sample cost. We evaluated the combination of the RainDance ThunderBolts™ Cancer Panel and Illumina MiSeq System with the Ion Torrent AmpliSeq™ Cancer Hotspots v2 panel and Proton™. We used the RainDance ThunderBolts™ Cancer Panel to enrich 10ng of both FFPE and fresh frozen (FF) samples of colorectal cancer tumor and adjacent normal tissue. The amplicon-based ThunderBolts Cancer Panel uses single molecule PCR technology to target 50 important cancer genes including tumor suppressors, hotspots and drug resistance markers and features a DirectSeq™ workflow method that integrates sequencing adapters and attaches sample indexers directly, without additional library preparation steps. The enriched samples were sequenced on an Illumina MiSeq System. The results were compared to those using 20ng of the samples with the Ion Torrent AmpliSeq Cancer Hotspots v2 panel and Proton. Sequence validation was performed using RainDrop digital PCR (dPCR) for positive targets derived from both sequencing panels. Both targeted sequencing panels demonstrated comparable sequencing metrics and minor allele calls and frequencies. We found that the sensitivity and specificity of droplet-based PCR (ThunderBolts) enrichment outperformed traditional PCR methodology. The correlation of variant frequency by dPCR validation was more closely aligned to droplet-based enrichment and Illumina sequencing. Furthermore, the reproducibility and ease of use for droplet-based enrichment allowed for a more efficient and directed workflow for clinical laboratories We concluded that the combination of theRainDance ThunderBolts Cancer Panel and Illumina MiSeq System was the best available NGS solution for analyzing common somatic mutations on FFPE tissue, and Fluid Biopsy™ samples (circulating tumor DNA).

3391T

Molecular characterization of over growth syndromes using NGS reveals potential phenotype-genotype correlation. *F. Chang[1], L. Liu[1], E. Fang[1], G. Zhang[1], L. Emrick[1], M. Li[1,2].* 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dan Duncan Cancer Center, Baylor College of Medicine, Houston, TX.

Recent studies have shown that a group of overgrowth syndromes, such as congenital lipomatous overgrowth with vascular, epidermal, and skeletal anomalies (CLOVES), Proteus syndromes, and two overlapping disorders, megalencephaly-polymicrogyria-polydactyly-hydrocephalus (MPPH) and megalencephaly-capillary malformation (MCAP), are caused by somatic activating mutations in genes involved in the PI3K-AKT signaling pathway. Due to the low-abundance nature of these mutations, routine Sanger sequencing often yields negative results. We have developed a next generation sequencing (NGS) panel that targets all known mutations in multiple genes involved in the PI3K-AKT pathway for this group of overgrowth syndromes. Eight patients including two prenatal cases and six postnatal cases suspected of overgrowth syndromes were tested using the NGS sequencing panel. A somatic mutation in the PIK3CA gene was identified in all 8 cases including one G542K mutation, two H1047L mutations, two H1047R mutations, two G914R mutations and one E110del mutation with the mutant allele frequencies ranging from 4.65% to 38.36%. These PIK3CA mutations were only present in the affected tissues in the majority of the cases demonstrating causal role of the mutations in the development of the diseases and the importance of obtaining affected tissue for testing. Phenotype-genotype correlation analysis showed that mutations at amino acid residue 1047 are often associated with CLOVES syndrome or isolated overgrowth of extremities or fatty mass; while the G914R mutation is preferentially related to syndromes with over growth of brain, such as MPPH/MCAP. In addition, in vitro cell culture showed significant enrichment of the cells harboring mutant alleles, indicating that the activating mutation may render growth advantages to the mutant cells. In a prenatal case, the PIK3CA G542K mutation was positive in the DNA extracted from cultured amniocytes but negative in uncultured cells of the same amniotic fluid specimen, suggesting that cultured amniocytes can be used for prenatal diagnosis of these syndromes. Our experience demonstrates that NGS technology is highly sensitive for the detection of low-level mosaic mutations and can be used for the diagnosis of these overgrowth syndromes in both prenatal and postnatal settings.

3392S

Small clone of JAK2V617F positive chronic eosinophilic leukemia detected by realtime PCR. *N. Yu[1], Y.J. Kim[2], S.J. Park[1], Y.R. Kim[3], K.A. Lee[1], J.W. Kim*[4].* 1) Department of Laboratory Medicine, Yonsei University College of Medicine; 2) Samkwang medical laboratories; 3) Department of Internal Medicine, Yonsei University College of Medicine; 4) Department of Laboratory Medicine, Yonsei University Wonju College of Medicine*.

Chronic eosinophilic leukemia (CEL) is defined as clonal proliferation of eosinophil precursors resulting in persistently increased numbers of eosinophils. Evidence of clonality is crucial in making a diagnosis of CEL. Herein we report a case of CEL with JAK2V617F positive by allele specific PCR and realtime PCR but negative by pyrosequencing and direct sequencing. A 33-year-old Korean female patient was admitted to our hospital in March 2012 for evaluation of persistent abdominal pain, diarrhea, vomiting. The patient had no previous history of hematologic malignancies or cytotoxic therapy. Peripheral blood smear showed eosinophilia. Bone marrow aspiration and biopsy revealed normocellular marrow with increased number of eosinophils and its precursors (19.7% of all nucleated cells). Endoscopic gastric biopsy revealed eosinophilic infiltration. Conventional karyotyping showed 46,XX in all metaphases. In fluorescence in situ hybridization and gene rearrangement studies, no BCR-ABL, FGFR1 and PDGFR rearrangements were detected. Allele specific PCR for JAK2V617F mutation (Seeplex JAK2 Genotyping kit, Seegene, Seoul, Korea) revealed faint positive band. Direct sequencing for exon 14 of JAK2 gene (ABI3500 genetic analyzer, Applied Biosystems, CA, USA) did not reveal any point mutation that prompted further investigation. The pyrosequencing (JAK2 pyro kit, Qiagen, CA, USA) showed negative results while realtime PCR using Real-Q JAK2V617F detection kit (BioSewoom, Seoul, Korea) was confirmed positive for the mutation. To date, only 6 cases of JAK2V617F has been reported in the literature. However, JAK2V617F positive CEL cases always demonstrated negative results for other gene rearrangement; therefore, indicating that JAK2V617F could be an early event in the leukemogenesis. Although prognosis are not well defined due to paucity of CEL cases with JAK2V617F mutation, in our case, steroid alone led to a good response and outcome. In this case, the small mutated clone was detected using most sensitive methods such as allele specific PCR and realtime PCR which could detect a burden as low as 0.01%, while direct sequencing and pyrosequencing with low sensitivity failed to detect mutated clone, causing a false negative. Therefore, selection of a more sensitive molecular method would be helpful for accurate diagnosis. Investigation into more cases with JAK2V617F positive CEL are needed to gain a deeper understanding of the mechanism involved in leukemogenesis.

3393M

Approaches to Integrating Germline and Tumor Genomic Data in Cancer Research. *L.E. Mechanic[1], H. Spencer Feigelson[2], K.A.B. Goddard[3], C. Hollombe[3], S.R. Tingle[1], S.A. Nelson[1], E.M. Gillanders[1].* 1) Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD; 2) Institute for Health Research, Kaiser Permanente Colorado, Denver, CO; 3) Center for Health Research, Kaiser Permanente Northwest, Portland, OR.

Cancer is characterized by a diversity of genetic and epigenetic alterations occurring in both the germline and somatic (tumor) genomes. Hundreds of germline variants associated with cancer risk have been identified, and similarly large amounts of data have been generated identifying mutations in the tumor genome that play important roles in tumorigenesis. Increasingly, these two genomes are being explored jointly to better understand how cancer risk alleles contribute to carcinogenesis and whether they influence development of specific tumor types or mutation profiles. To understand how data from germline risk studies and tumor genome profiling is being integrated, we reviewed 160 articles describing research that incorporated data from both genomes, published between 2009 and 2012, and summarized the current state of the field. We identified three types of research questions being addressed using these data: (1) use of tumor data to determine the putative function of germline risk variants; (2) identification and analysis of relationships between host genetic background and particular tumor mutations or types; and (3) use of tumor molecular profiling data to reduce genetic heterogeneity or refine phenotypes for germline association studies. We also found several descriptive studies comparing germline and tumor variation in a gene, related genes, or gene families. We identified a large and expanding set of resources that can be used to analyze and integrate data from both genomes and discuss opportunities and challenges for this approach.

3394T

Multiplex detection of KRAS mutations in colorectal cancer FFPE samples using droplet digital PCR. *S. Cooper, W. Yang, D. Shelton, J. Berman, B. Zhang, S. Tzonev, E. Hefner, J. Regan.* Digital Biology Center, Bio-Rad Laboratories, Inc., Pleasanton, CA.

Targeted therapies in many cancers have allowed unprecedented progress in the treatment of disease. However, routine implementation of genomic testing is limited due to: 1) limited amounts of sample (pg-ng range) per biological specimen, 2) diagnostic turnaround time and workflow, 3) cost, and 4) difficulties in detection of mutational loads below 5%. KRAS is mutated in approximately 40% of colorectal cancers. The majority of these KRAS mutations are activating mutations in codons 12, 13, and 61, and are predictive of a negative response to αEGFR therapy. To optimize therapy strategies for personalized care, it is critical to rapidly screen patient samples for the presence of multiple KRAS mutations. We have developed a multiplexing strategy to screen clinically-actionable KRAS mutations in cancer samples using digital PCR. This multiplex assay targets seven KRAS point mutations prevalent at greater than 1% , resulting in an effective profiling tool for 98% of KRAS mutant colorectal cancers (Faulkner et al. 2010). No pre-amplification step is required. This KRAS screening assay was used to quantify KRAS mutational load in a panel of FFPE specimens from advanced metastatic colorectal cancer patients. KRAS mutations present at <1% fractional abundance were detected in multiple samples. This sensitive and inexpensive method reduces the risk of contamination and can be easily implemented in molecular diagnostic laboratories for rapid, routine screening of cancer patients.

3395S

Evidence of multiple independent *NF2* somatic inactivation and tumor initiation events in neurofibromatosis type 2-associated vestibular schwannomas. *A. Pemov[1], R. Dewan[2], J. Kim[3], K. Morgan[2], R. Vasquez[2], P. Chittiboina[2], X. Wang[2], A. Ray-Chaudhury[2], S. Chandrasekharappa[4], J. Butman[5], A. Asthagiri[6], D.R. Stewart[1].* 1) NIH/NCI, Bethesda, MD; 2) NIH/NINDS, Bethesda, MD; 3) NIH/NIODCT, Bethesda, MD; 4) NIH/NHGRI, Bethesda, MD; 5) NIH/CC, Bethesda, MD; 6) University of Virginia, Charlottesville, VA.

Background and hypothesis. Neurofibromatosis type 2 (NF2) is a tumor predisposition syndrome that results from mutation of the *NF2* tumor suppressor gene (chromosome 22q12). The hallmark of NF2 is the presence of bilateral vestibular schwannomas (VS). It is generally accepted that somatic inactivation of the second copy of *NF2* is the initiating event for neoplastic transformation in NF2-associated tumors including VS. Though NF2-associated and sporadic VS share identical histopathologic findings and underlying cytogenetic alterations, NF2-associated VS are less responsive to radiosurgery, are associated with worse surgical outcomes and often appear as a cluster of grapes. We hypothesized that the distinct multi-lobulated morphology of NF2-associated VS is determined by the genetics of the tumors and that the “grapes” in the cluster are in fact individual tumors that arose independently with distinct somatic mutations in the *NF2* gene and overall genomic architecture. **Study design and methods.** To test this hypothesis, we analyzed the mutation status of *NF2* by Sanger sequencing and genomic architecture by Illumina HumanOmniExpress SNP-arrays in 24 tumor samples procured from five VS, or 4.8 specimen per tumor on average, in four NF2 patients. **Results.** First, we identified *NF2* germline mutations in each of four NF2 patients. Two patients carried deleterious point mutations in *NF2* and the other two presented with large chromosomal deletions affecting the gene. Second, we identified 18 distinct *NF2* somatic mutations (13 point mutations and five chromosome 22 LOH events) in five tumors, or 3.6 average independent somatic *NF2* hits per tumor and ranging from 1 to 7 per tumor. Finally, SNP-array analysis revealed that a deletion or mitotic recombination affecting the entire chromosome 22 or its q-arm was the most frequent large-scale chromosomal aberration in the tumors (5/24, or 21 per cent). The rest of the genome in all but one tumor sample resembled that of normal diploid cells. **Conclusions.** These findings support our hypothesis that NF2-associated multi-lobulated VS are not single tumors, but rather are clusters of multiple smaller tumors, each arising from their own unique second hit event. These data suggest, for the first time, that the majority of NF2-associated VS arise from multiple independent tumor initiation events. These findings have important clinical implications and will be critical when establishing clinical trials endpoints.

3396M

MicroDNA (Extra Chromosomal Circular DNA) in Mammalian Tissues and Cancer Cell Lines. *P. Kumar, L.W Dillon, Y. Shibata, A. Dutta.* DEPT. OF BIOCHEMISTRY & MOLECULAR GENETICS, SCHOOL OF MEDICINE, UNIVERSITY OF VIRGINIA, CHARLOTTESVILLE, VA.

DNA in normal mammalian somatic cells primarily exists as long chromosomes or in mitochondria. However, we previously reported the presence of thousands of short (mostly 200-400 bps) extra chromosomal circular DNAs (microDNAs), single- or double-stranded, arising mainly from the GC rich regions in the genome, and enriched in genic regions in embryonic mouse brain, heart and liver tissues (Shibata, Y. et al. Science 336, 82-86, (2012)). More than 95% of microDNAs could be mapped uniquely to a genome. Chromosomal loci that are enriched sources of microDNA in the adult brain are somatically mosaic for microdeletions that appear to arise from the excision of microDNAs. MicroDNAs often have micro homology of 2-15 bases at the start and after the end of the corresponding chromosomal sequence suggesting that a micro homology mediated mechanism could be the mechanism of microDNA biogenesis. In the current research we studied the characteristics of microDNAs in chicken cell lines deleted in various DNA repair genes, including DNA ligase IV and Ku70 involved in non-homologous end joining (NHEJ); BRCA1, BRCA2, NBS1, and Rad54 involved in homologous recombination (HR); CtIP required for microhomology-mediated alternative end joining (MMEJ); and MSH3 involved in DNA mismatch repair (MMR). The MSH3-deleted cells had a striking change in their microDNA population, suggesting that the mismatch-repair pathway, and by extension replicative polymerase slippage is involved in the generation of a fraction of the microDNAs. To determine tissue specificity of microDNA generation, we examined a panel of mouse tissues (brain, heart, lung, liver, kidney, muscle, sperm, spleen, testis and thymus) and human prostrate (C4-2, LnCap and PC-3) and ovarian (ES2 and OVCAR-8) cancer cell lines. MicroDNAs were observed in all tissue types, confirming that microDNAs are universally present even in germ cells. We discovered that there are hot spots of microDNA generation distributed throughout the genome that are common between tissues and correlate with areas of high gene density and high GC content. However, hierarchical clustering on the basis of microDNA co-ordinates classified the prostate and ovarian cancer cell lines into two separate groups suggesting that at least some microDNAs are tissue-specific and therefore their sites of origin are affected by tissue-specific gene expression patterns or epigenetic marks.

3397T

Heterozygous mutations in *PALB2* predispose to breast cancer by causing DNA replication and damage response defects. *R. Winqvist[1,2], M. Bose[1,2], J. Nikkilä[1,2,8], A.C. Parplys[3], K. Borgmann[3], Y. Huo[4], K. Rapakko[2], N. Laurila[1,2], P. Nieminen[5], B. Xia[4], H. Pylkäs[1,2], H. Pospiech[6,7].* 1) Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, University of Oulu, Oulu, Finland; 2) Northern Finland Laboratory Centre NordLab, Oulu, Finland; 3) Laboratory of Radiobiology and Experimental Radiation Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 4) Department of Radiation Oncology, Rutgers Cancer Institute of New Jersey and Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA; 5) Medical Informatics and Statistics Research Group, University of Oulu, Oulu, Finland; 6) Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland; 7) Research group Biochemistry, Leibniz Institute for Age Research—Fritz Lipmann Institute, Jena, Germany; 8) Present address: Gene Function Team, Breakthrough Breast Cancer Research Centre, ICR, London, UK.

Breast cancer is the most common cancer among females. Although the majority of the breast cancer cases appear to be sporadic, it has been estimated that as much as 5-10% is significantly contributed by powerful hereditary genomic risk factors. Besides mutations in *BRCA1* and *BRCA2*, also defects in the *PALB2* gene occur worldwide and are important in hereditary predisposition to breast cancer. PALB2 interacts with BRCA1 and BRCA2 to regulate homologous recombination and to mediate cellular DNA damage response. We have previously shown that in the Finnish population a relatively common heterozygous mutation in *PALB2* increases the risk of malignancy in carrier individuals about sixfold. About 1% of the breast cancer patients in Finland carry this mutation. Unfortunately, however, the mechanistic details for how heterozygous mutations in cancer predisposing genes such as *PALB2* trigger disease development are currently largely obscure. By analysing lymphoblastoid cell lines from heterozygous female *PALB2* mutation carriers, we have recently been able to demonstrate that haploinsufficiency for PALB2 causes aberrant DNA replication/damage response. Mutation carrier cells show increased origin firing and shorter distance between consecutive replication forks. Carrier cell lines also show compromized S and G2/M cell cycle checkpoint functions and aberrant DNA repair capabilities, which result in gradual accumulation of various genomic lesions. Elevated chromosome instability was also observed in primary blood lymphocytes of heterozygous *PALB2* mutation carriers, indicating that the described mechanisms of genome destabilization operate also at the organism level. Thus the functional loss of one copy of the *PALB2* gene causes instability in the genome, even though the other allele is still intact and functioning. These exciting findings provide a new mechanism for early stages of breast cancer development that may also apply to other heterozygous homologous recombination signaling pathway gene mutations involved in hereditary cancer predisposition.

3398S

Investigation of *de novo* mutation rates in families with DNA Polymerase ε and Δ exonuclease domain mutations. *S.E.W. Briggs, O. Venn, C. Palles, G. McVean, I.P. Tomlinson.* The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom.

DNA polymerases ε and Δ have an essential role in eukaryotic DNA replication, proofreading newly synthesised DNA and identifying and excising mispaired nucleotides. Mutations in the exonuclease domains (EDMs) of DNA polymerases *POLE* (MIM #174762) and *POLD1* (MIM #174761) have been identified as causal mutations in a novel cancer predisposition syndrome, polymerase proofreading associated polyposis (PPAP [MIM #615083; #612591]). PPAP displays high penetrance and a variable phenotype including colorectal, endometrial and ovarian tumours. Sporadic colon and endometrial tumours with somatic *POLE* EDM mutations are defined by a hypermutated phenotype with a distinct mutational spectrum of increased transversion mutations.

To investigate the effect of germline mutations on *de novo* mutation rates we performed whole genome sequencing of two parent-offspring quintets. One quintet carries *POLE* L424V (c.1270C>G, NM_006231), and one *POLD1* S478N (c.1433G>A, NM_002691); in each family the mother and two of three offspring carry the mutation.

The whole genomes of each family member were sequenced to a read depth of 7.9x. We applied alignment based calling to each family without incorporating relatedness, detected 4.9 million variants with a transition:-transversion ratio of 2.16. To these calls we applied a modified Lander-Green algorithm to construct a transmission scaffold across chromosomes for each family. Probabilistic models were used to classify sites inconsistent with transmission, followed by filtering for false-positive errors. Candidate mutations were then validated using independent targeted sequencing to estimate the false positive rate, and direct paired-end phasing was conducted.

We sought to test whether there is an elevated incidence of *de novo* mutations in the carriers. Preliminary results suggest that in both pedigrees the transition:transversion ratio is decreased (1.2 in *POLE*; 1.5 in *POLD1*), with an abundance of indels (23% in *POLE*; 34% in *POLD1*) suggesting an impact of germline mutation in line with the biological expectation. Paternal bias in *de novo* mutation origin has been reported in humans and chimpanzees. If DNA polymerase mutations do impact germline mutation, we may observe an increase in the female contribution given that the mothers are the carriers in both families.

3399M

NBN gene expression and cytogenetic changes in irradiated cells with NBN gene mutations. *D. Januszkiewicz-Lewandowska[1, 2, 3], B. Swiatek-Koscielna[1], J. Rembowska[1], A. Dzikiewicz-Krawczyk[1], M. Zawada[1], J. Nowak[1].* 1) Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 2) University of Medical Sciences, Poznan, Poland; 3) Department of Medical Diagnostics, Poznan, Poland.

NBN gene product is part of the MRE11/RAD50/NBN complex, which plays essential role in genomic stability. Homozygous mutation of NBN gene c.657-661del causes Nijmegen breakage syndrome in the course of which there is a high susceptibility to cancer. Heterozygous NBN gene mutations are considered to be a minor cancer risk factor. The study tries to answer the following questions: Does mutations of NBN gene influence the DNA synthesis and expression of NBN gene in irradiated immortalized cells? What is the effect of NBN gene mutation and irradiation on the chromosome stability in lymphoblastic cell line with germinal NBN gene mutations? Radioresistant DNA synthesis measured by 3H thymidine incorporation assay, NBN gene expression and chromosomal aberrations were assessed in cultured cells before and after irradiation with doses 1-8 Gy. Our results indicated that only cells with homo- and heterozygous c.657-661del mutation showed higher 3H thymidine incorporation after irradiation, particularly in the 48-hour culture and after higher doses of irradiation. These results suggest that those cells showed significantly lower rates of DNA synthesis inhibition. Radiosensitivity measured as a numer of chromosomal aberrations per one metaphase was significantly higher in cells with all NBN gene mutations studied as compared to the control cell line and was clearly marked for homozygous and heterozygous c.657-661del cell lines. Cells with homozygous and heterozygous c.657-661del and p.R215W mutations possessed significantly higher relative NBN gene expression when compare to the control. After irradiation the relative expression of NBN was significantly higher only in homozygous c.657-661del and heterozygous p.R215W. Our observation may indicate on a compensatory increase in the expression of p70-NBN subunit in cells with homozygous c.657-661del and heterozygous c.657-661del and p.R215W mutations. Similarly, irradiation may in an alternative way in these cell lines stimulate significantly higher increase of NBN gene expression. Financial support - National Science Centre 2011/01/B/NZ5/04322.

3400T

Mechanism of formation of complex chromosomal aberrations in patients with myelodysplastic syndromes (MDS): clonal evolution or chromothripsis? *Z. Zemanova[1], H. Lhotska[1], J. Brezinova[2], I. Sarova[1, 2], L. Lizcova[1], S. Izakova[1], S. Ransdorfova[2], K. Svobodova[1], Z. Krejcik[2], M. Belickova[2], M. Siskova[3], A. Jonasova[3], R. Neuwirtova[3], J. Cermak[2], K. Michalova[1].* 1) Center of Oncocytogenetics, IMBLD, General University Hospital and First Faculty of Medicine, Charles University in Prague, Czech Republic; 2) Institute of Hematology and Blood Transfusion, Prague, Czech republic; 3) 1st Medical Department, General University Hospital and First Faculty of Medicine, Charles University.

Complex chromosomal aberrations (CCAs) are seen in approximately 20% of patients with newly diagnosed MDS and are associated with a poor prognosis. The mechanisms leading to its formation remain poorly understood. It is not clear whether CCAs arise by a gradual acquisition of genetic changes during the clonal evolution or by extensive chromosome fragmentation through a unique cellular crisis (chromothripsis). The aim of the study was to perform detailed genome wide analyses of bone marrow cells of patients with MDS and CCAs and to assess the frequency and clinical significance of chromothripsis. A comprehensive molecular cytogenetic analysis was performed of bone-marrow cells from 177 adults with CCAs (≥3 aberrations) identified with conventional G-banding at the diagnosis of MDS. The CCAs were studied by FISH with Vysis DNA probes (Abbott, Des Plaines, IL) and mFISH/mBAND methods (MetaSystems, Altlussheim, Germany). Genomic imbalances were identified with CytoChip Cancer SNP 180K (BlueGnome, Cambridge, UK) or with Illumina Human CytoSNP-12 arrays (Illumina, San Diego, CA). The findings in 86/177 patients corresponded to the gradual accumulation of random aberrations over clonal evolution. In remaining 91 patients, mFISH/mBAND and microarray assays showed breaks with a large number of chromosomal losses and gains, probably resulting of chromothripsis. The fragmentation or shattering of chromosomes into many pieces was also observed. Parts of the fragmented chromosomes were translocated or inserted elsewhere in the genome, leading to chaotically reassembled chromosomes. The most frequently shattered were Nos. 5, 7, 17, and 12. The OS of patients with shattered chromosomes did not differ from that of patients with no evidence of fragmentation (median OS in both groups, four months; $p = 0.224$). We observed signs of chromothripsis in 51.4% of MDS patients with CCA. Although initial studies have suggested that patients displaying a chromothripsis have more aggressive tumors and poor outcomes, in this cohort, we found no significant difference in the OS of patients with and without chromothripsis. We assume that in MDS, CCA can occur by either mechanism, i.e., as the result of chromothripsis or with gradual clonal evolution. The prognosis in both cases is very poor and deteriorates with increasing number of aberrations, regardless of mechanism of complex karyotype formation. Supported by RVO-VFN64165, GACR P302/12/G157/1, PRVOUK-P27/LF1/1, MHCR 00023736.

3401S

A new control mechanism for repair of DNA in human cells: MDC1 and ATR regulate DNA Double-Strand Break (DSB) resection independently of ATM. *P.S Bradshaw[1,2], M. Komosa[1], M.S Meyn[1,2,3,4].* 1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Division of Clinical and Metabolic Genetics, Department of Paediatrics, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) The Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada.

To maintain genome stability, human cells have developed strategies to detect, repair or induce cell death/apoptosis in response to DNA damage. DNA DSBs frequently arise from exposure to exogenous and endogenous sources and are one of the most lethal forms of DNA damage. Detection of DNA DSBs result in the activation of the PIK-related kinase ATM and its phosphorylation mediated signaling pathway. ATM is mutated in the autosomal recessive disorder Ataxia-Telangiectasia (A-T) in which defective DSB repair leads to genome instability, radiation sensitivity, cerebellar ataxia, immunological defects, radiation sensitivity and high cancer risk. Here we report that ATM is rapidly recruited to DNA breaks in the nuclei of human cells, where it forms distinctive spatially well-defined, non-diffusible, foci. Using interative deconvolution microscopy, we find that ATM belongs to a novel class of DSB repair proteins that rarely co-localize with Rad51 and RPA (proteins directly involved Homologous Recombination Repair - HRR), resected DSBs, or proteins that make up the mega-base region of γ-H2AX containing chromatin that surround DSBs. Human cells lacking ATM display reduced numbers of RPA and Rad51 foci following the generation of radiation-induced DSBs. RPA foci formation is dependent on CtIP a protein target for ATM phosphorylation and implicated in DNA resection and the generation of single stranded DNA (ssDNA). Our data indicate that ATM promotes the initiation of resection and subsequent HRR, through the CtIP dependent generation of ssDNA. In contrast, MDC1, a downstream member of the ATM-dependent signaling pathway, acts independently of ATM to restrain the initiation of DSB resection and prevent the formation of ssDNA. Interestingly, we find that MDC1 restrains DNA resection by inhibiting the ATR dependent activation of CtIP. Like ATM, ATR is a PIK-related kinase involved in DNA damage signaling. Mutations in ATR can cause Seckel Syndrome (SCLK1), an autosomal recessive disorder characterized by growth retardation, dwarfism, microcephaly, mental retardation and facial dysmorphology. Significantly, CtIP mutations can also cause Seckel Syndrome (SCLK2). Our work identifies a new MDC1 / ATR dependent pathway of DSB resection that functions independently of the classical ATM-mediated DNA damage signaling pathway. Further, our demonstration of MDC1/ATR control of CtIP mediated resection establishes a molecular link between SCLK1 and SCLK2 syndromes.

3402M

Breast cancer eQTLs from the Nurses' Health Study. *A. Hazra[1,5], A. Quiroz-Zarate[2], B. Harshfield[1], R. Hu[1], N. Knoblauch[3], A.H. Beck[3], S.E. Hankinson[4], V. Carey[1], R.M. Tamimi[1,5], D.J. Hunter[5], J. Quackenbush[2].* 1) Medicine, Brigham and Women's Hospital, Boston, MA; 2) Center for Computational Biology, Dana Farber Cancer Institute, Boston, MA; 3) Pathology, Beth Israel Deaconess Medical Center, Boston, MA; 4) Division of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA; 5) Program in Genetic Epidemiology and Statistical Genetics, Harvard School of Public Health, Boston, MA.

Genome-wide association studies (GWAS) of breast cancer have identified 71 single nucleotide polymorphisms (SNPs), majority of which are located in intergenic or intronic regions. To identify regulatory variants we conducted expression quantitative loci (eQTLs) analysis among invasive postmenopausal breast cancer cases in the Nurses' Health Study (NHS) diagnosed from 1990-2004 with GWAS data. RNA extracted from formalin fixed paraffin embedded (FFPE) breast tumor (tumor) and normal adjacent (normal) breast tissue was profiled using the Affymetrix Human Transcriptome Array (HTA 3.0v1). Multivariate linear regression was conducted separately for 376 tumor and 264 normal samples using 71 loci and 26,004 expression probes. We also developed the functional QTLs (fQTLs) method to gain pathway-level insight and evaluate the hypothesis that SNPs are associated with regulation of processes. We identified 12 trans-acting SNPs in normal tissue, 9 trans-acting SNPs in estrogen-receptor positive (ER+) tumors and 13 trans-acting SNPs associated ER- tumors, including 3 2 overlapping loci across normal adjacent, ER+ tumors and ER- negative tumors (false discovery rate (FDR)<10%). Although not statistically significant after FDR correction, we identified suggestive cis-acting loci. For the fQTL analyses, we tested the association of SNPs with genes as repeated measurements of a specific pathway. We used the Molecular Functions (MFs) of the Gene Ontology and identified 2 SNPs associated with 2 MFs in normal tissue, 1 SNP associated with 5 MFs in ER- tumors but no significant associations in ER+ tumors (FDR<10%). Using bioinformatic tools we identified trans-acting loci that alter transcription-binding motifs and were associated with transcripts in chromatin enhancer regions or non-coding RNA. The overlapping loci are associated with multiple unique transcripts in breast tissue and may be key regulatory SNPs in breast tissue. Further functional work is needed to elucidate the mechanism of these associations. Our results provide functional insights on the underlying biology of breast cancer loci in the specimen type that is most impactful in translation to clinical practice.

3403T

Utilization of Bioluminescence Resonance Energy Transfer (BRET) for functional evaluation of missense variants at the BRCA1-BARD1 heterodimeric RING-RING interface. *T. Kayoko[1], E.L. Young[1], B. Luo[2], S.V. Tavtigian[1].* 1) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT; 2) Drug Discovery Core, University of Utah, Salt Lake City, UT.

[Background] The clinical significance of BRCA1 RING domain missense substitutions has been underlined by multiple studies of familial breast/ovarian cancer cases. The structural interaction between BRCA1 and its binding partner, BARD1 (BRCA1-associated ring domain 1), is essential for their tumor suppressor function. To evaluate the impact of missense substitutions on BRCA1/BARD1 complex formation, we applied a Bioluminescence Resonance Energy Transfer (BRET) system to quantify protein-protein interactions. [Method] The BRET assay measures the efficiency of energy transfer from a donor to an acceptor moiety. NanoLuciferase-fused BRCA1 RING domain (donor construct) and HaloTag-fused BARD1 (acceptor construct) were generated as BRET biosensers to monitor the interactions. 24 hours after these were transfected into human embryonic kidney 293 cells, the reactive ligands were covalently attached to the HaloTag proteins; donor luminescence (460nm/40nm) and acceptor fluorescence (620nm/40nm) emission intensity were then measured with a microplate reader. Sample values were normalized by donor activity in each well, and BRET ratios (acceptor/donor emission) were calculated. [Results] We made an initial test of BRET assay suitability using wild-type and mutant constructs of the BRCA1 RING domain. Two cancer-predisposing mutations (BRCA1 p.C61G and p.C64G), known to disrupt the BRCA1/BARD1 dimerization, showed significantly decreased BRET ratios (p.C61G = 0.137±0.0004, p.C64G = 0.135±0.0052) compared to wild-type (WT = 0.374±0.0372), whereas BRCA1 p.I42V, which has been reported to retain the ability to bind BARD1, showed a BRET ratio that is comparable to wild-type (I42V = 0.330±0.0220). [Conclusions] Clinically significant missense mutations are often found in protein binding regions. The BRET-based protein interaction assay has several advantages over conventional methods: 1) protein interactions can be measured in living mammalian cells in a real-time manner, which enables monitoring of modification-dependent protein interactions, 2) no artifact prone washing steps are required before measurement, 3) high-throughput screening compatible. Additional BRCA1 and BARD1 RING domain missense substitutions will be evaluated to further assess the suitability of BRET for interpretation of unclassified variants in these two genes.

3404S

TERT Polymorphism rs2736100-C Is Associated with EGFR Mutation-Positive Non-Small Cell Lung Cancer. *W. Liu[1], R. Wei[1], L. Cao[2], H. Pu[3], H. Wang[4], A.E. Wasilk[1], Y. Zheng[4], X. Niu[5], X. Weng[2], H. Zhang[2], M. Favus[4], L. Zhang[6], W. Jia[3], Y. Zeng[3], S. Lu[5], H. Wang[3], Y. Liu[2].* 1) Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 47907, USA; 2) Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, P. R. China; 3) State Key Laboratory of Oncology in South China, Sun Yat-Sen University Cancer Center, Guangzhou, 510080, P. R. China; 4) Department of Medicine, The University of Chicago, Chicago, IL 60637, USA; 5) Department of Shanghai Lung Cancer Center, Shanghai Chest Hospital, Shanghai Jiaotong University, Shanghai 200031, P. R. China; 6) Department of Thoracic Surgery, Sun Yat-Sen UniversityCancer Centre, Guangzhou, 510080, P. R. China.

Somatic mutations in the epidermal growth factor receptor (EGFR) gene are oncogenic drivers for non-small cell lung cancer (NSCLC). It has been suggested that EGFR mutation-positive (EGFRmut+) NSCLC might be a unique orphan disease with genetic susceptibility. EGFR mutations are associated with lung adenocarcinoma histology, female gender and non-smoking history. NSCLC with these features has been consistently associated with an intronic polymorphism rs2736100 in the telomerase reverse transcriptase (TERT) gene. We therefore hypothesized that rs2736100 may be a risk factor for EGFRmut+ NSCLC. To test this, we conducted a genetic association study in Chinese NSCLC patients (n=714) and healthy controls (n=2,520). Among the cases, 42% (n=303) were EGFRmut+. We further tested the association between the EGFR mutation status and mean leukocyte telomere length (LTL). The potential function of rs2736100 in lung epithelial cells was also explored. We found that while the rs2736100-C allele was significantly associated with NSCLC as an overall phenotype (OR=1.24, 95%CI=1.1-1.39, $p=4\times10^{-4}$), it was more strongly associated with EGFRmut+ NSCLC (OR=1.52, 95%CI=1.28-1.80, $p=1.6\times10^{-6}$) compared to the EGFRmut- NSCLC (OR=1.07, 95%CI=0.92-1.24, p=0.4). There was also a significant difference in the allele frequency between EGFRmut+ and EGFRmut- populations (OR=1.42, 95%CI=1.15-1.76, $p=1.1\times10^{-3}$). The results remained significant after controlling for age, gender, smoking status and histology (corrected $p\leq0.035$). Further analyses demonstrated that while NSCLC patients as a whole have significantly longer LTL compared to healthy controls ($p\leq10^{-13}$), the EGFRmut+ patients have even longer LTL compared to EGFRmut- patients (p=0.008, corrected p=0.043). Meanwhile, rs2736100 was significantly associated with TERT mRNA expression in both normal and tumor lung tissues after controlling all covariates (corrected $p\leq0.047$). We also found that the rs2736100 DNA sequence has an allele-specific affinity to nuclear proteins extracted from lung epithelial cells. The polymorphism was also associated with an altered enhancer activity in vitro. Our study for the first time linked a TERT polymorphism to EGFRmut+ NSCLC. The data further revealed insight into the role of TERT and its polymorphisms in the carcinogenesis of EGFR mutation-driving lung cancer.

3405M

Genome Analysis of Latin American Cervical Cancer: Frequent Activation of the PIK3CA Pathway. *H. Lou[1], G. Villagran[2], J. Boland[3], K. Im[4], W. Zhou[3], U. Odey[5], E. Torres[6,7], I. Martinez[6,7], E. Basaure[7], J. Mitchell[3], D. Roberson[3], J. Sawitzke[1], L. Garland[1], M. Herrera[4], D. Wells[1], J. Troyer[1], F. Pinto[2], S. Bass[3], X. Zhang[3], M. Castillo[2], B. Gold[4], M. Yeager[3], J. Berumen[6], E. Alvirez[5], E. Gharzouzi[2], M. Dean[4].* 1) Leidos Biomedical Research, Inc., Frederick, MD USA; 2) Instituto de Cancerologia, Guatemala City, Guatemala; 3) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD USA; 4) National Cancer Institute, Frederick, MD USA; 5) Hospital Central Universitario "Dr. Antonio M Pineda", Barquisimeto, Lara State, Venezuela; 6) Unidad de Medicina Genómica, Facultad de Medicina, Universidad Nacional Autónoma de México, México, D.F. México; 7) Hospital General de México. México, D.F. México.

Cervical cancer is the most common cause of cancer mortality for women living in poverty, causing over 28,000 deaths annually in Latin America and 266,000 worldwide. We identified somatically mutated PIK3CA genes using exome and ultra-deep targeted sequencing of invasive cervical tumors from Guatemala, Venezuela, and Mexico. Analysis of 636 HPV positive cervical cancer revealed activation of PIK3CA genes in up to 34% of tumors. The novel PIK3CA mutations including delN107 (ABD domain); P134T (link between ABD and RBD domains); T229I and M232L (RBD); V952G and E1034Q (Kinase domain) and mutations, D350N, D454Y (C2) and E726K (Kinase domain) lacked function study, were identified in the study. To future study the effect of these mutations on PIK3CA catalytic activity, we expressed wild type and mutant constructs in vitro and measured levels of AKT (Thr308) and AKT (Ser473) phosphorylation, respectively. The result showed that these novel PIK3CA mutations led to significantly and slightly increased levels of phospho-AKT (Thr308) and phospho-AKT (Ser473), respectively. Furthermore PIK3CA and HPV E6 and E7 gene expression is significantly higher in PIK3CA mutation-positive tumors. Our study demonstrates that PIK3CA/AKT pathway activation contributes to the development of HPV positive cervical cancer in Latin America.

3406T

CARP Is a Potential Tumor Suppressor in Gastric Carcinoma. *F. Lu[1,2], J. Xue[2], Y. Hu[2], L. Gan[2], H. Yang[2], Y. Wei[2].* 1) Center for Molecular Genetics, Sichuan Provincial People's Hospital, chengdu, Sichuan, China; 2) State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, P.R. China.

Background The caspase-associated recruitment domain-containing protein (CARP) is expressed in almost all tissues. Recently, the tumor-suppressive function of CARP was discovered and attracted increasing attention. This study aimed to investigate the role of CARP in the carcinogenesis of human gastric carcinoma. Methods/Result Compared with normal gastric tissue, the downregulation of CARP expression was observed in gastric carcinoma tissue by cDNA array and tissue microarray assay. In vitro, the gastric carcinoma cell line (BGC-823) was stably transfected with pcDNA3.1B-CARP or plus CARP siRNA, and we used MTT, flow cytometry, cell migration on type I collagen, cell-matrix adhesion assay and western blot analysis to investigate the potential anti-tumor effects of CARP. Our data showed that overexpressing CARP suppressed the malignancy of gastric carcinoma BGC-823 cell line, including significant increases in apoptosis, as well as obvious decreases in cell proliferation, migration, adhesion ability, and tumor growth. The tumor-suppressive effects of CARP were almost restored by siRNA-directed CARP silence. In addition, overexpression of CARP induced G1 arrest, decreased the expressions of cyclin E and CDK2, and increased the expressions of p27, p53 and p21. In vivo, the tumor-suppressive effect of CARP was also verified. A single-nucleotide polymorphism (SNP) genotype of CARP (rs2297882) was located in the Kozak sequence of the CARP gene. The reporter gene assay showed that rs2297882 TT caused an obvious downregulation of activity of CARP gene promoter in BGC-823 cells. Furthermore, the association between rs2297882 and human gastric carcinoma susceptibility was analyzed in 352 cases and 889 controls. It displayed that the TT genotype of rs2297882 in the CARP gene was associated with an increased risk of gastric carcinoma. Conclusions CARP is a potential tumor suppressor of gastric carcinoma and the rs2297882 C>T phenotype of CARP may serve as a predictor of gastric carcinoma.

3407S

Association of Platelet Derived Growth Factor-B (PDGF-B) and Human Epidermal Growth Factor Receptor -2 (HER-2/neu) Single Nucleotide Polymorphisms (SNP's) with Gallbladder Cancer (GBC). *K. Mishra[1], V.K. Kapoor[1], A. Behari[1], S. Khan[2], S. Agrawal[1].* 1) Surgical Gastroenterology, Sanjay Gandhi Post-Graduate Institute of Medical Sciences (SGPGIMS), Lucknow, Uttar Pradesh, India; 2) Integral University, Kursi Road, Lucknow, Uttar Pradesh, India.

Purpose: Gall bladder cancer (GBC), a highly malignant gastrointestinal tumour, is very common in north India. Platelet derived growth factor-B (PDGF-B) at chromosome 22q12.3-q13.1 and Her-2/neu at 17q12-q21 play an important role in tumour angiogenesis. PDGF-B and Her-2/neu overexpression has been found in many cancers. We studied PDGF-B and Her-2/neu single nucleotide polymorphisms (SNPs) in GBC and gall stone associated benign diseases viz. chronic cholecystitis (CC) and xanthogranulomatous cholecystitis (XGC). Methods: DNA was extracted from blood in patients with GBC (n=195), CC (n=140), XGC (n=47) and normal controls (n=300). PDGF-B polymorphisms were investigated using ARMS PCR for +286A/G and +1135A/C, and for Her-2/Neu Ile655Val by PCR-RFLP method. Results: +286A/G polymorphism homozygous GG genotype, and +286G allele was found to be risk associated for GBC (OR=5.25 P= 0.0001 and OR=2.02 P= 0.0001). Recessive model (GG vs. AA+GA) of +286A/G polymorphism was risk associated (OR=4.78, P=0.0001) whereas dominant model (AA vs. GG+GA) was risk protective (OR=0.56, P=0.003) with GBC. +1135A/C polymorphism CC genotype and +1135C allele was risk associated (OR=3.19, P=0.0001 and OR=1.81, P=0.0001) with GBC. Recessive model (CC vs. AA+AC) was risk associated (OR=2.75, P=0.0003) whereas dominant model (AA vs. CC+AC) showed protective association (OR=0.56, P=0.0024) with GBC. In HER-2 Ile655Val polymorphism, dominant model and Val allele of this polymorphism was risk protective for XGC. Allele frequencies of GBC were compared with CC and XGC. Homozygous GG genotype, recessive model and +286G allele of +286A/G genotype were risk associated but dominant model of this genotype was risk protective for GBC as compared with CC. In case of GBC vs XGC, GG genotype, recessive model and +286G allele were risk associated for GBC. In +1135A/C polymorphism, homozygous CC genotype, recessive model and 1135C allele were risk associated whereas dominant model was risk protective when GBC was compared with CC. In HER-2 Ile655Val polymorphism, Val allele was risk associated when GBC was compared with XGC. In haplotype analysis, haplotypes ACIle (OR= 1.48), GAVal (OR=1.70) and GAIle (OR=2.00) were risk associated with GBC. Conclusion: PDGF-B +286A/G and 1135A/C are risk susceptibility markers for GBC and need further evaluation.

3408M

Highly Sensitive Fusion Transcript Detection and Quantification in Cancer. *L.C. Watson, S.M. Gross, I. Khrebtukova, F. Schlesinger, S. Pathak, T. Hill, T. Singer, G. Schroth.* Functional Genomics, Illumina, Inc., San Diego, CA.

Gene fusion detection in cancer samples can provide tumor-specific information for cancer research, clinical diagnosis and targeted treatment. Common fusion detection methods such as qPCR and FISH are restricted to known fusion junctions and limited in the number of genes that can be detected in parallel. In contrast, RNA sequencing is a powerful approach for simultaneous discovery of all possible fusion junctions in a single reaction. But, the sequencing depth required for sensitive detection of fusions from whole-transcriptome libraries can be cost-prohibitive. Here we describe a cancer-specific capture-based approach for fusion detection by RNA sequencing that requires only a fraction of the sequencing depth of whole-transcriptome methods. We designed biotinylated oligo probes that densely target coding regions of ~1500 cancer-associated genes including more than 300 genes associated with clinically relevant fusions, 80 alternatively spliced transcripts derived from TCGA RNA-Seq data, and more than 400 genes that show expression differences in tumor/normal pairs across TCGA subtypes. This panel was used to capture cancer-specific fusions from total RNA-Seq libraries. We used commonly studied cancer cell lines—including MCF-7, K562, PC-3, T47D, U20S—and Universal Human Reference RNA (UHRR) to compare the sensitivity of fusion detection across three RNA-Seq library prep methods: (1) cancer gene library capture (2) whole-transcriptome library capture and (3) PolyA selection. We show that probes targeting individual exons can robustly capture well-characterized cancer gene fusions such as *BCR-ABL* and *BCAS3-BCAS4.* Furthermore, these comparisons demonstrate the sequencing efficiency of using highly targeted and specific panels. By including *in vitro*-transcribed RNA gene fusions as spike-in controls, we also establish that fusion detection is quantitative. We show that selective enrichment of RNA-Seq libraries with cancer-specific capture probes enables high-resolution mapping of genomic rearrangements in patient cancer samples, even those derived from FFPE, facilitating sequencing studies that were not previously possible.

3409T

Charting gene regulatory networks for interpreting prostate cancer GWAS results. *G. Wei, P. Gao, Q. Huang, Y. Yang, R. Lin.* Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, University of Oulu, P.O. Box 5000, Oulu FIN-90014, Finland.

Prostate cancer represents one of the most common cancers in males and the second most common cause of cancer-related death among men in developed world. Driver transcription factors (TFs) including androgen receptor, TMPRSS2-ERG, FOXA1 and HOXB13 are known to be important in prostate cancer progression, and frequently overexpressed in human prostate tumors. Particularly, genetic evidence shows that HOXB13 also affects prostate cancer susceptibility but with unknown genetic and functional mechanisms. To understand underlying function for these TFs, we chart gene regulatory networks driven by them using chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq) and define up to 50,000 binding sites in average for each TF. Through systematic localization analyses, we find that the prostate cancer risk-associated single nucleotide polymorphisms (SNPs) identified by genome-wide association studies (GWASs) are significantly enriched in these TF binding sites, revealing potential causal roles for these non-coding genetic variants. Intriguingly, this suggests that GWAS-identified non-coding SNPs are more likely to affect chromatin binding affinity for disease cell-type-specific TFs. We further provide compelling evidence to substantiate the causality for several prostate cancer risk-associated SNPs including rs339331. We find that rs339331 lies within a HOXB13 functional binding site and causes upregulation of RFX6 gene, promoting prostate cancer cell growth and metastasis. Together, our study suggests a gene regulatory network approach in interpreting GWAS results of prostate cancer and other diseases.

3410S

Functional analysis of mutations in polymerase epsilon gene that predispose to polymerase proofreading associated polyposis (PPAP). *E. Heitzer[1], C. Palles[2], P. Ulz[1], M.R. Speicher[1], S. Kearsey[3], I. Tomlinson[2].* 1) Institute of Human Genetics, Medical University Graz, Graz, Austria; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 3) Department of Zoology, University of Oxford, UK.

Germline mutations in the exonuclease domain of genes encoding the catalytic subunits Polymerase δ and ε, i.e. POLD1 and POLE, predispose to "polymerase proofreading associated polyposis" (PPAP) resulting in multiple colorectal adenomas and carcinoma with high penetrance and dominant inheritance. Moreover, somatic mutations in the ED of POLE have been frequently found in sporadic CRCs and endometrial carcinoma. Tumors, with both inherited and somatic EDMs, were microsatellite stable and showed a mutator phenotype with a dramatic increase of base substitutions indicating impaired proofreading. To assess the functional consequences of POLE germline and somatic mutations, we characterized exonuclease domain mutant alleles in S.pombe. We therefore generated constructs encoding the equivalent changes in the fission yeast protein and determined the effect of this change on reversion of the ade6-485 allele, 5-fluoroorotic acid and canavanine resistance. The somatic mutations including P286R, S297F, and S459F showed a dramatically increased mutation rate (up to more than 100-fold) compared to the wild type strain. V411L was only slightly increased (4-7 fold). Also the germline variant L424V showed increased mutation rates (2-11-fold) compared the wild type strain. As expected whole genome sequencing of POLE mutated strains and strains with a deficient MMR background revealed an elevated rate of base substitutions. This data indicate that EDMs in POLE indeed lead to proofreading deficiency and can induce replication errors during synthesis of oncogenes and tumor suppressors resulting in tumor formation.

3411M

From GWAS to therapy: Fatty acid synthase in uterine leiomyomata. *Z. Ordulu[1], M. Hayden[1], S. Eggert[2], M. Shinohara[3], C. Serhan[3], B.J. Quade[2], C.C. Morton[1,2].* 1) Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Boston, MA; 2) Pathology, Brigham and Women's Hospital, Boston, MA; 3) Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, MA.

Uterine leiomyomata (UL) pose a major public health problem given their high prevalence (>70%) in women of reproductive age and their indication for >200,000 hysterectomies in the U.S. annually. Genome-wide association studies for fibroid predisposition performed in white women identified a candidate SNP (rs4247357) under a linkage peak in 17q25.3 that spans the gene fatty acid synthase (*FASN*). By immunohistochemistry we found FAS protein levels elevated (3-fold) in UL in comparison to matched myometrial tissue. It has been reported that FAS is upregulated in various neoplasms and implicated in tumor cell survival. We assessed effects of FAS inhibitors Orlistat and C75 in primary cell cultures of 17 matched myometrium and UL by cell proliferation assays. Cell counts are reduced in a dose-dependent manner with a half maximal inhibitory concentration ranging between 50-75 uM (Orlistat) and 30-45 uM (C75). Further, we showed that Orlistat treatment reduced C16:0, C18:0 free fatty acids and increased CoA-ester of C14:0, C16:0 and C18:0 by gas chromatography-mass spectrometry analysis, supporting its role through inhibition of the thioesterase domain of FAS. *In vivo* experiments are underway in an immunocompromised mouse model with subcutaneous fibroids derived from primary human fibroid cells. *FASN* represents the first potential UL predisposition gene identified in white women with responsiveness of UL to inhibition of FAS a potential model for targeted therapy to mitigate risks and complications of surgical management.

3412T

Evaluation of miR-338-3p role in the progression of Esophageal Squamous Cell Carcinoma. *H. Mollaei[1,2], M. Shafiee[1,3], S.A. Aleyasin[3], S.J. Mowla[4], M. Moghaddam-Matin[5,6], †The two first authors (M. Shafiee and H. Mollaei) contributed equally to this work.* 1) Golestan Research Center of Gastroenterology and Hepatology (GRCGH), Golestan University of Medical Sciences, Gorgan, Iran; 2) Department of Biology, Faculty of Natural Sciences, University of Tabriz, Tabriz, Iran; 3) Department of Medical Genetics, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran; 4) Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran; 5) Cell and Molecular Biotechnology Research Group, Institute of Biotechnology, Ferdowsi University of Mashhad, Mashhad, Iran; 6) Department of Biology, Faculty of Sciences, Ferdowsi University of Mashhad, Mashhad, Iran.

Objective: Esophageal cancer is one the most lethal form of human malignancies, mostly due to its late diagnosis. Golestan province in Northeastern Iran, as a part of Caspian littorals, has been identified as one of the highest risk areas in the world. This high incidence, highlighting the necessity of identifying novel and potent diagnostic and prognostic biomarkers for early detection of the disease. Alterations in the expression of miRNAs have been widely reported in numerous diseases including almost all types of cancers. Act as oncogenes (oncomiRs) or tumor suppressors, miRNAs are playing prominent roles in cancer-related processes such as proliferation, apoptosis, metastasis and angiogenesis. Motivated by our recent data on downregulation of miR-338-3p expression in esophageal squamous cell carcinoma, we aimed to investigate its potential causative role in tumorigenesis of an Esophageal Carcinoma cell line, KYSE-30. Materials and Methods: The human KYSE-30 cell line was obtained from Pasteur Institute in Tehran, Iran. After seeding the cells, they were transfected with either pEGFP-C1 vector containing the miR-338-3p precursor sequence or the mock pEGFP-C1 vector with Lipofectamin 2000 transfection reagent. 48 h after transfection, the cells were harvested for RNA isolations and flow cytometry analysis. Experiments were repeated at least twice and cell cycle alterations were analyzed using FlowJo software. Results: In contrast to the cells transfected with a mock vector, as a negative control, the KYSE-30 cells overexpressing miR-338-3p showed a remarkable increase (15.5 times) in miR-338-3p expression 48 hours after transfection. Over-expression of miR-338-3p caused a dramatic alteration in cell cycle distribution of transfected cells, including around two folds increase in the number of cells distributed in the sub-G1 phase of the cell cycle. Conclusion: The data suggests that miR-338-3p may act as an anti-apoptotic role in epithelial cells of esophagus. A similar observation has been reported in other cancer cells such as pancreatic interepithelial neoplasia, gastric cancer and colorectal carcinoma. A functional causative role of miR-338-3p in ESCC is also supported by bioinformatic analyses using miRwalk and DIANALAB. These softwares predicted MMP2, MMP9, SMO, n-Ras, c-Myc and Cyclin D as potential targets of this miRNA. These findings and further studies may suggest miR-338-3p as a good biomarker for ESCC diagnosis and treatment target.

3413S

Assessment of the clinical relevance of variants of uncertain significance in BRCA2 by functional and computational approaches. *L. Guidugli[1], N.M. Lindor[2], V.S. Pankratz[1], D.L. Masica[3], R. Karchin[3], F.J. Couch[1].* 1) Lab Med and Path, Mayo Clinic, Rochester, MN; 2) Health Sciences Research, Mayo Clinic Arizona, Scottsdale, AZ; 3) Institute for Computational Medicine, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD.

Genetic testing of individuals with a family history of breast and/or ovarian cancer has led to the identification of many unique BRCA2 missense variants of uncertain significance (VUS). Current methods for classification of BRCA2 mutations depend heavily on the availability of information on segregation of VUS with breast and ovarian cancer within families, and on predicted probabilities of pathogenicity based on evolutionary sequence conservation. The failure to assess the clinical relevance of VUS, owing to insufficient genetic information, may deprive patients of the benefits of risk assessment and enhanced risk management available to those found to carry pathogenic mutations in BRCA2. In order to classify greater numbers of VUS that may or may not have available family data for segregation analysis, we have developed a homology directed DNA repair (HDR) assay for evaluation of the pathogenicity of BRCA2 VUS. This assay measures the influence of VUS in the DNA binding domain of BRCA2 on the HDR activity of the protein and has high sensitivity and specificity for established pathogenic missense mutations. Here we report on the assessment of 106 VUS located inside the BRCA2 DNA Binding Domain (DBD) (amino acid 2459 to 3190) using this assay. Whereas 47 variants had a 99% probability of being non-pathogenic, and were classified as non-pathogenic, 41 variants had a 99% probability of being pathogenic, and were classified as pathogenic by the HDR assay. The results from the assay were subsequently compared to predictions from the Align-GVGD sequence based prediction model (r=0.49). The categories of Align-GVGD associated with high probabilities of pathogenicity (p=0.66 to 0.81) were significantly different from the results of the HDR assay (p<0.05). In contrast, the Phenotype Optimized Sequence Ensemble (POSE) sequence model, trained on functional properties, displayed a moderate to strong correlation with the HDR assay (r=0.69). Overall, the results suggest that the prior probability of pathogenicity assigned to a number of VUS by the Align-GVGD model is too high and that the POSE model may be a useful method for selecting BRCA2 VUS for further evaluation by functional or family-based methods.

3414M

The role of the cilia protein Arl13b in activated-Smoothened medulloblastoma oncogenesis. *S.N. Bay[1,2], R.C. Castellino[3], T. Caspary[2].* 1) Genetics and Molecular Biology Program; 2) Dept of Human Genetics; 3) Dept of Pediatrics, Emory University, Atlanta, GA.

The proper balance of Shh signaling is critical to both development and adult homeostasis. Too little Shh signaling causes birth defects such as holoprosencephaly, while too much Shh signaling results in tumors, as occurs repeatedly in Gorlin syndrome patients. Indeed, overactive Shh signaling in a variety of tissues contributes to the formation of cancers including basal cell carcinoma and medulloblastoma - a tumor of the cerebellum that is the most common pediatric malignancy in the central nervous system. Many available treatments are inadequate; surgery, chemotherapy, and radiation are invasive and can cause long-term cognitive defects in patients, and current molecular therapies are prone to fail due to tumor resistance. The drawbacks of current treatments mean that new approaches are needed. Recent research has shown that Shh-derived medulloblastomas are “addicted” to cilia since cilia are required for proper Shh signaling. We study a ciliary GTPase called Arl13b that uniquely regulates Shh signaling, both upstream of Smoothened (Smo) and between Smo and Gli activation. Loss of Arl13b results in ligand-independent constitutive low-level pathway activation but prevents maximal signaling. Due to this unique relationship to Shh signaling, we predict that the loss of Arl13b will reduce the high levels of pathway output caused by constitutive activation of Smo. To determine whether loss of Arl13b can prevent or delay tumor formation, we deleted *Arl13b* in a mouse model of medulloblastoma. We also use mouse embryonic fibroblasts to define how the loss of Arl13b affects Smo localization and Shh pathway output in an activated-Smo signaling context. Together, these experiments will define whether Arl13b is a viable target for molecular therapies for the treatment of medulloblastoma and will reveal mechanistic details of Arl13b’s role in the regulation of normal Shh signaling and a disease-relevant activated Shh signaling context.

3415T

Modeling cancer in zebrafish embryos. *L. Francescatto, N. Katsanis.* Center for Human Disease Modeling, Duke University, Durham, NC.

Zebrafish has emerged as an invaluable model organism to study cancer proliferation, metastasis and invasiveness. Many characteristics make this model organism attractive to research experimentation, such as external embryonic development, transparency, high fecundity rate, and short generation time. In particular, the lack of an adaptive immune system during the first four weeks of development facilitates cancer xenotransplantation acceptance and survival. Chemical screenings have also enabled studies of cancer maintenance and proliferation. To date, several types of cancer cells have been successfully transplanted into different stages of the zebrafish embryo. The feasibility of this model prompted us to ask whether we could take primary cancer cells from patients and test them for tumorigenicity. We have successfully observed proliferation and migration of colon cancer cells from two patients, further confirming the potential of zebrafish in studying cancer. Using this model of cancer, we have undertaken a two-pronged approach that harnesses the power of this system. First, we have compared the exomes of distally attached human cells to non-migrating cell populations as a surrogate to identify candidate metastatic drivers. In parallel, we have initiated a small molecule screen to identify potential drugs that can inhibit either proliferation or migration of cancer cells. Zebrafish cancer xenotransplants provide an in vivo model that is cost and time efficient, and can be a first line of evidence for anticancer therapies.

3416S

Lifestyle Issues of BRCA Mutation Carriers that May Affect and Health Outcomes. *A. Caceres, R. McLamara, O. Ivanov, K. Wiercinski, C. Buffington.* Florida Hospital Celebration Health, Celebration, FL.

Rationale and Objectives. The BRCA protein helps to prevent cancer by repair of damaged or mutated cells and by enhancing apoptosis. Obesity, sedentary activity, a poor diet, life stressors and sleep loss cause metabolic and hormonal changes that increase the risk for mutagenesis, tumor growth and progression. Therefore, a healthy lifestyle is likely to be particularly important for reduction of cancer occurrence among BRCA mutation carriers. The purpose of this study was to study the lifestyles of BRCA mutation carriers and identify conditions that could increase their risk for primary or secondary cancer. Methods. The study population included 38 BRCA mutation cancer survivors, 45 BRCA mutation previvors, and 26 controls. Diet was analyzed by 3-day food records; physical activity, anxiety, quality of life were assessed by standardized questionnaires; and body size and composition were determined by 4-lead bioelectric impedance. Results. No significant differences between BRCA previvors, survivors or the non-BRCA mutation controls were found as regards diet (calories, macronutrient, micronutrient content), blood pressure, heart rate, body mass index, education, use of coffee, cigarettes or alcohol. Body weight and BMI also did not differ between groups although the BRCA survivors, as compared to previvors and controls, had a greater % of body fat and lower lean tissue mass in association with reduced physical activity and function. BRCA carriers (previvors and survivors, in comparison to controls) exhibited significantly (p<0.05) greater psychological distress (high anxiety, reduced emotional quality of life) than non-BRCA controls and in association to poor sleep quality and reduced sexual activity and libido. Conclusion. BRCA mutation carriers display several lifestyle conditions that could adversely affect health and quality of life including psychological distress, reduced sexual activity and libido, poor sleep quality, and adverse body composition relative to low physical activity and function.

3417M

Small molecule of natural origin has potential to activate TLR3 and aid in immunotherapy for cancer. *A. Das, J. Kaur, S. Sharma.* Biotechnology, Delhi Technological University, Delhi, Delhi, India.

Cytotoxic T lymphocytes (CTLs) and Natural killer (NK) cells mediate cellular immunity against tumor cells and virus infected cells.. CTLs recognize the peptides presented by MHC class I molecules present on the surface of target cells and kill them. But certain tumor and virus infected cells downregulate their MHC class I expression to escape this T cell response. Hence, NK cell mediated cytotoxicity, which is not MHC restricted becomes important for response against such MHC class I deficient target cells. Viral dsRNA triggers NK cell response by binding to their surface TLR3 receptors. In the presence of dsRNA the ectodomains of two TLR3 molecules dimerize which brings their cytoplasmic domains in close proximity to start the relay of intracellular signaling. So far only synthetic oligonucleotides have been used for studying the activation of NK cells. Here, we have proposed small molecules of natural origin which could dimerize the TLR3 ectodomains in a manner similar to dsRNA. These small molecules showed high affinity for TLR3 receptor. Hence, with the potential to activate NK cells and capability to selectively eliminate transformed cells; these compounds can be used as a combinatorial immunotherapy for cancer.

3418T

Accurate and inexpensive sequencing of *BRCA1* and *BRCA2*: Application to a US-wide study of breast cancer in Latinas. *M. Dean[1], J.F. Boland[2], M. Yeager[2], K.M. Im[1], J. Mitchell[2], D. Roberson[2], K. Jones[2], J. Sawitzke[3], S. Bass[2], X. Zhang[2], H. Lee[2], R. Eggebeen[2], V. Robles[4], C. Hollis[5], C. Barajas[5], E. Rath[6], C. Arentz[7], J.A. Figeroa[7], D.D. Nguyen[7], Z. Nahleh[6].* 1) Lab Experimental Immunology, NCI-FCRDC, Frederick, MD; 2) Cancer Genetics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD; 3) Basic Science Program, Leidos Biomedical Research, Inc., Frederick, MD; 4) Nueva Vida, Richmond, Richmond, Virginia; 5) Latino Community Development Agency, Oklahoma City, OK; 6) Texas Tech University Health Sciences Center, El Paso, TX; 7) Texas Tech University Health Sciences Center, Lubbock, TX.

Germline mutations in the *BRCA1* and *BRCA2* genes are the most common cause of inherited breast cancer and are found in 5-10% of all cases. Detection of BRCA mutation carriers can lead to therapeutic interventions such as mastectomy, oophorectomy, hormonal prevention therapy and improved screening; as well as targeted therapies such as PARP-inhibition. African Americans and Hispanics are 3-4 times less likely to receive BRCA screening, despite having approximately the same mutation frequency as non-Jewish European Americans. Furthermore, both of these minorities have a higher mortality for breast cancer. To begin addressing this health disparity, we initiated a nationwide trial of BRCA testing of Latinas with breast cancer. Patients were recruited through community organizations, clinics, public events, and through the mail and internet. Subjects completed the consent process and questionnaire over the phone, and/or email contact, and provided a saliva sample by mail. DNA from 117 subjects was extracted from saliva and used to sequence the entire *BRCA1* and *BRCA2* coding regions and splice sites using a community designed panel at a total supply cost of $80/subject. A newly developed, more accurate polymerase (HiQ) allowed for the detection of mutations and elimination of many false positive results. Subjects ranged in age from 23-81 years (mean of 51 years), 6% had bilateral disease, 57% were ER/PR+, 23% HER2+; and 17%; had triple-negative disease. A total of 6 different predicted deleterious mutations, one newly described and the rest rare, were identified along with 4 variants of unknown effect. An alignment of all available primate *BRCA1* and *BRCA2* sequences was useful in analysis of unknown variants. Application of this strategy on a larger scale could lead to improved cancer care of minority and under-served populations.

3419S

How choriocarcinoma DNA identification can interfere in treatment decision? A report of two unexpected cases. *G.JF. Gattas[1], P. Exman[2], VD. Cantagalli[1], FT. Gonçalves[1], FN. Aguiar[2], REO. Ramos[2], MDPE. Diz[2].* 1) Dept of Legal Medicine, Bioethics and Occupational Health, Faculty of Medicine, University of Sao Paulo, Sao Paulo, Brazil; 2) Cancer Institute of Sao Paulo, Faculty of Medicine, University of Sao Paulo, Sao Paulo, Brazil.

Choriocarcinoma is a rare, highly malignant tumor, mainly with gestational origin, and rarely germ cell origin. The clinical presentation and histological characteristics of both tumors are identical but the origin differentiation is critical for treatment decision. DNA polymorphisms (STRs) used for forensic pourposes can be useful to identify paternal cells in paraffin wax embedded choriocarcinoma tumor cells, compared to patient's healthy cells, to characterize gestational origin. We here describe two patients that were evaluated after histologic diagnosis of choriocarcinoma. The first patient was a 21 years old woman with previous abortion history and clinical presentation extremely suggestive of gestational disease. After no favorable treatment results the tumor DNA analysis, using MiniFiler Kit (Applied Biosystems), was realized and revealed identical profile when compared to the patient's blood sample. Thus, the treatment was changed for BEP (Bleomicin, Ethoposide and Cisplatin) that is considered the first line treatment for an ovarian germ cell tumor. The patient achieved pathologic complete response, with no disease evidence in 2 year follow-up. The second case was a 54 years old woman with initial presentation of metastasis within the pulmonary artery and absence of other disease site with biopsy confirmation as choriocarcinoma. The main hypothesis was gestational origin and she was firstly treated with BEP with partial results and no serum marker response. DNA STRs analysis showed the presence of additional genetic material, beyond the maternal one for all 8 STRs evaluated, including Y chromosome. The results indicated paternal contribution (confirmed with the husband DNA) and the treatment was then changed to dactinomycin, and patient achieved complete radiologic and of serum marker response. These two cases suggest that DNA STRs could be used as a necessary tool to confirm the gestational origin of choriocharcinomas even when the clinical case is not indicative of it. Financial Support: LIM-40 (HC-FMUSP); FAPESP (09/54868-6).

3420M

A simple mainstreamed, oncogenetic pathway delivers fast, affordable routine BRCA testing for ovarian cancer (OC) patients. *H. Hanson[1,2], D. Riddell[1], F. Smith[1], V. Cloke[1], M. Gore[3], S. Banerjee[3], N. Rahman[1,2], MCG consortium.* 1) Division of Genetics & Epidemiology, The Institute of Cancer Reserach, London, United Kingdom; 2) Cancer Genetics Unit, Royal Marsden NHS Foundation Trust, London, United Kingdom; 3) Gynaecology Unit, Royal Marsden NHS Foundation Trust, London, United Kingdom.

Background ~15% of ovarian cancer (OC) patients carry a BRCA mutation, yet historically are under-referred to Cancer Genetics (CG) services. NICE (2013) recommended that BRCA testing be offered to individuals with ≥10% chance of mutation detection. This and development of targeted agents e.g PARP inhibitors make it essential that OC patients can access genetic testing within an appropriate time frame. We developed an "oncogenetic model" to facilitate testing in routine oncology care. Methods Patients with non-mucinous OC <65 years at Royal Marsden Hospital were offered BRCA gene testing at their oncology appointment by cancer team members who had completed 30min online training. The test results were interpreted by genetics and returned to the patient by the cancer team. All mutation carriers attended a Genetics appointment for detailed discussions and to organise testing for relatives. Any patient could contact genetics at any time, if required. Patients and clinicians were sent a questionnaire to assess their experience. Results 119 OC patients were tested in 6 months. Patients were tested during initial treatment (n=32), relapse (n=45) or follow-up (n=42). 20 (17%) carried a mutation: 8 *BRCA1*, 12 *BRCA2*. 12 had no family history, 8 met current genetic referral criteria. BRCA results changed management in 9/20 (45%) carriers. No patient requested additional Genetics input before testing. Questionnaires were sent to 71 patients; 46/57 non-carriers and 11/14 carriers responded. All were pleased to have had testing, and to have accessed testing within Oncology. All understood the results potentially had implications for themselves and their family. 25 clinicians undertook training. All welcomed the opportunity to offer testing and felt confident in consenting and giving results. Conclusions The Oncogenetic testing pathway provides flexible, patient-centred, equitable, high-throughput gene testing with considerable time and cost savings compared to traditional pathways. It is now the standard BRCA testing pathway for non-mucinous OC patients at Royal Marsden and roll-out to other NHS centres is planned.

3421T

Inhibition of STAT3 and RelA expression levels Bortezomib treated K-562 leukemic cells and indiction of apoptosis. *N. Selvi Günel[1], B. Tazcanli Kaymaz[1], S. Kipçak[1], A. Dalmizrak[1], B. Kosova[1], G. Saydam[2].* 1) Department of Medical Biology, Ege University, Izmir, Turkey; 2) Department of Hematology. Ege University, Izmir, Turkey.

Signal transducer and activator of transcription (STAT) proteins; especially STAT3, are crucial for signalling pathways leading to discover the underlying cancer development mechanism. Activated STAT3 increases leukemic cell proliferation and survival; also suppresses anti-tumour immunity and apoptosis by promoting pro-oncogenic pathways, including nuclear factor-κB (NF-κB). RelA is one of the gene products of NF-κB that gives rise to increase in STAT3 activation. Bortezomib (BOR) is a first class of proteasome inhibitor, and used in of multiple myeloma and mantle cell lymphoma treatment. In this study, we aimed to identify the cytotoxic and apoptotic effects of BOR upon chronic myelogenous leukemia (CML) cell model K562, with determining STAT3 and RelA expressions both at mRNA and protein levels. Cell proliferation was assessed by XTT assay in order to determine cytotoxicity of BOR upon leukemic cells. While mRNA expression levels of STAT3 and RelA were analyzed by qRT-PCR; protein expressions were detected via western-blot method. IC50 was calculated as 17.7 nM for 72th hour. Number of apoptotic cells were increased by 37% and 38% for 72th-96th hours (p=0.0041, p=0.0038). While STAT3 mRNA expression was significantly decreased by 66% [(2.86 fold; p= 0.0043)] at 72th hour, Rel A was downregulated by 45.8% and 83.63% [(1.83 fold; p= 0.003), (6.1 fold; p=0.0002)] for 72th and 96th hours. As for protein results, both Rel A and STAT3 protein expressions were highly inhibited in a time dependent manner.

3422S

Real-time and sequential profiling of cancer through concurrent somatic mutation and gene amplification analysis of cancer by digital sequencing of cell-free DNA from patients with metastatic solid tumors.

A. Talasaz[1], S. Mortimer[1], B. Schiller[1], G. Mei[1], D. Sebisanovic[1], L. Siew[1], A. Zapanta[1], S. Huang[2], D. Hoon[2], H. Eltoukhy[1]. 1) R&D, Guardant Health, Redwood city, CA; 2) Dept Molecular Oncology, John Wayne Cancer Institute, Santa Monica, CA 90404.

Introduction: Genomic profiling of cell-free DNA (cfDNA) through a simple blood test is non-invasive, sequential and can potentially be more predictive than tumor profiling. Due to high concordance of tumor and cfDNA, this blood test can be considered when biopsies are not an option due to the risk, cost, or time. **Methods:** We used single-molecule digital sequencing technology, Guardant360 for high-fidelity and comprehensive profiling of 54 actionable genes in cfDNA of more than 500 metastatic cancer patients. This technology allows concurrent profiling of somatic mutations, gene amplifications and fusions. The longitudinal blood samples from patients have been processed to assess residual disease post surgical resection or targeted treatments. **Results:** The overall detection rate of somatic alterations in cfDNA approached 90% for all indications (breast, lung, colorectal, melanoma and prostate). When blood samples were concurrent with biopsies, the concordance of alterations was 93%. Residual disease in 30 colorectal and melanoma patients post multi-surgical resection has been studied (2-7 draws). The alterations were positively correlated with clinical status of the patients. In 70 colorectal patients, plasma samples post targeted treatment were processed. A combination of previously reported and novel resistance-related alterations were found in cfDNA from the post targeted treatment samples. **Conclusions:** Comprehensive sequencing of patient's cancer in real-time through simple blood test can empower oncologists in making more informed treatment decisions, especially when tissue biopsy is not an option.

3423M

Constitutive mismatch repair deficiency syndrome:clinical description in a French cohort.

C. Colas[1,17], N. Lavoine[2], G. Sebille[3], O. Cabaret[4], C. Charpy[5], T. Frebourg[6], N. Entz-Werle[7], Q. Wang[8], S. Lejeune[9], D. Leroux[10], G. Couillault[11], G. Leverger[12], JP. Fricker[13], R. Guimbaud[14], M. Mathieu-Dramard[15], F. Bourdeaut[16], M. Muleris[17], O. Caron[18], L. Brugieres[2]. 1) Laboratoire d'oncogénétique et d'angiogénétique, Département de génétique, GH Pitié-Salpêtrière, APHP, Paris, France; 2) Départemment de cancérologie de l'enfant et de l'adolescent, Institut Gustave Roussy, Villejuif,France; 3) Département de dermatologie, Institut Gustave Roussy, Villeejuif, France; 4) Département de biologie et pathologie médicales, Service de génétique, Institut Gustave Roussy,Villejuif, Frannce; 5) Département d'anatomo-pathologie, Institut Gustave Roussy, Villejuif, France; 6) Laboratoire de génétique, Hôpital universitaire, Institut pour la recherche biomédicale etl'innovation, Rouen, France; 7) Département d'oncologie pédiatrique, Centre hospitalier universitaire, Strasbourg, France; 8) Plateforme mixte de génétique constitutionnelle des cancers fréquents HCL-CLB, Centre Léon-Bérard, Lyon, France; 9) Département de génétique clinique, Hôpital Jeanne de Flandre, Lille, France; 10) Département de génétique, Hôpital universitaire, Grenoble, France; 11) Département de pédiatrie, Hôpital universitaire, Dijon, France; 12) Département d'hématologie et d'oncologie pédiatriques, Hôpital d'enfants Armand Trousseau,Paris, France; 13) Département d'oncogénétique, Centre Paul Strauss, Strasbourg, France; 14) Département d'oncologie digestive, Institut Claudius Régaud et hôpital universitaire de Toulouse,Toulouse, France; 15) Unité de génétique médicale, hôpital universitaire d'Amiens, Amiens, France; 16) Département d'oncologie pédiatrique, Institut Curie, Paris, France; 17) Centre de Recherche Saint-Antoine INSERM / UMR S938, Equipe "Instabilité des microsatellites etcancer" Paris, France; 18) Département d'oncogénétique, Institut Gustave Roussy, Villejuif, France.

Introduction. Constitutive mismatch repair deficiency syndrome (CMMR-D) is a recently described childhood cancer predisposition syndrome involving biallelic mutation of MMR genes (MLH1, MSH2, MSH6 and PMS2). More than 140 cases have been previously reported but only as case reports. Methods. We performed a retrospective review of all 31 cases of CMMR-D from 23 families diagnosed in French genetics laboratories in order to describe clinical characteristics, treatment and outcome of malignancies, and biological diagnosis data of an unselected series of patients. Results. Overall, 67 tumors were diagnosed in these 31 patients, 17 (25%) hematologic malignancies, 22 (33%) brain tumors, 25 (37%) Lynch syndrome-associated malignancies, and 3 (5%) other tumors. Median age of onset of first tumor was 6.98 years [1.23-33.53]. 23 (74%) patients had NF1-unrelated CALMs or hypopigmented macules and 4 (13%) had brain malformative features. Colorectal adenomas were found in all 16 patients who have had colonoscopy with synchronous colorectal cancer or advanced adenoma in all of them. Overall, 18 patients died, 7 (39%) due to the primary tumor. Median survival after diagnosis of the primary tumor was 23 months [0.26-213.2]. Among the patients who survived after their first malignancy, 20 developed a second malignancy. No obvious excess of toxicity to treatment was reported. A familial history of LS-associated cancer was found in only 6 families, and consanguinity in 43% of cases. PMS2 mutations (18 patients) were more frequent than mutations of MLH1 (4 pts), MSH2 (3 pts) and MSH6 (6 pts). Conclusion. CMMR-D is a severe condition associated with multiple malignancies in childhood. Its rarity warrants international collaboration to define diagnosis criteria and guidelines for surveillance and prevention in order to decrease tumor-related mortality.

3424T

Germline *TP53* mutation analysis in HER2-positive breast cancer patients from Southern Brazil. *M. Fitarelli-Kiehl[1,2], J. Giacomazzi[1,2], P. Santos-Silva[1,3], V.B. Sempé[1], D. de Mendonça Uchôa[4], L.F. Rivero[4], M.S. Graudenz[4], P. Ashton-Prolla[1,2,3,5].* 1) Experimental Research Center, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; 2) Post-Graduate Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 3) Post-Graduate Program in Medicine: Medical Sciences, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 4) Department of Pathology, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; 5) Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Germline *TP53* mutations are associated with Li-Fraumeni Syndrome (LFS), an autosomal dominant disorder characterized by predisposition to multiple early-onset cancers including breast cancer (BC), the most prevalent tumors among LFS women. Recent reports show that *TP53* mutation carriers have a significantly higher likelihood of developing a BC overexpressing HER2 (63-83%) when compared to mutation-negative BC cases (16-25%). Germline *TP53* mutations are estimated to occur in 1 to 8.6% of patients with early onset BC unselected for family history of cancer, and HER2-positive status supposedly increases the odds of having a germline *TP53* mutation by nearly 7-fold. In this work, we determined the prevalence of germline *TP53* mutations in a cohort of women with HER2 overexpressing BC, diagnosed before age 60, unselected for family history of cancer. We retrospectively reviewed the immunohistochemistry data of BC from patients diagnosed at Hospital de Clínicas de Porto Alegre between years 2007 and 2012, and 106 women who fullfilled inclusion criteria were recruted and periferic blood was colected. Germline *TP53* sequence variants were identified by High Resolution Melting (HRM) analysis followed by Sanger DNA sequencing of the entire coding sequence and flanking intronic regions. Among 106 patients analyzed, 2 (1.9%, 95% CI 0.5-6.6%) were found to carry a *TP53* mutation with known pathogenic effect, and other 5 patients (4.7%, 95% CI 2.0-10.5%) shown sequence variants with unknown effect on p53 function (all variants with MAF<0.01). One patient with pathogenic mutation met classic LFS criteria. Although we found a low prevalence of germline *TP53* pathogenic mutations in HER2 overexpressing BC patients, *TP53* testing should be considered in these cases even in the absence of family history of cancer.

3425S

The *DICER1* Leiden Open Variation Database (LOVD). *N. Hamel[1], W.D. Foulkes[1,2,3].* 1) Dept Med Gen, McGill Univ Hlth Ctr, Montreal, PQ, Canada; 2) Lady Davis Res Inst Jewish Gen Hospital, Montreal, PQ, Canada; 3) Depts Oncology & Hum Genet, McGill University, Montreal, PQ, Canada.

Over the past several years, we and others have worked to describe the phenotypes associated with the *DICER1* pleiotropic tumor syndrome (OMIM 601200). While some tumors are strongly indicative of an inherited *DICER1* mutation even in the absence of other phenotypes (e.g. pituitary blastoma), many tumor types must be seen in combination with other *DICER1*-associated tumors or tell-tale phenotypes such as multinodular goiter in the patient or his or her relatives in order for the presence of a *DICER1* mutation to be expected. The Leiden Open Variation Database (LOVD) is an open source database framework that provides a flexible, freely available tool for Gene-centered collection and display of DNA variations. The *DICER1* LOVD database, https://grenada.lumc.nl/LOVD2/mendelian_genes/home.php?select_db=DICER1, is curated by our research group and contains an inventory of index patients with reported germ-line mutations in *DICER1* along with the associated phenotypes observed in the carrier family. Data compiled from the database to date shows that the majority of inherited mutations are frameshift and nonsense mutations predicted to result in protein truncation, though some pathogenic missense mutations are also observed. Some mutations are observed more than once in apparently unrelated families, suggesting either a common ancestry dating back several generations or a hot spot area in the gene where these specific mutations are more likely to arise. No case of homozygous or compound heterozygous mutation carrier has been reported to date, suggesting that carrying two mutated copies of *DICER1* in the germ-line may be incompatible with life. Associated tumor types reported in the database include pleuropulmonary blastoma, Wilms tumor, cystic nephroma, Sertoli-Leydig cell tumor and embryonal rhabdomyosarcomas, among others. It is our goal to make the *DICER1* LOVD database a central repository for all germ-line *DICER1* mutations. We believe that compiling all genotype and phenotype data relating to this syndrome in one publicly available location will provide an increasingly clearer picture of this syndrome and will be a useful resource for anyone attempting to assess whether phenotypes observed in particular patients are consistent with a *DICER1* inherited mutation. We therefore invite all research and clinical teams who have mutation and phenotype data from affected families to contribute to this group effort and submit their information to the database.

3426M

Breast Cancer in *PTEN* Hamartoma Tumor Syndrome: Can a Predictive Fingerprint Be Identified? *A. Machaj[1], J. Mester[2], C. Eng[2].* 1) Case Western Reserve University, Cleveland , OH; 2) Genomic Medicine Institute, Cleveland Clinic, 9500 Euclid Ave. NE50 Cleveland, OH 44195.

Breast cancer is predicted to occur in up to 85% of women with PTEN Hamartoma Tumor syndrome (PHTS) in a lifetime. PHTS includes patients with Cowden Syndrome (CS) and other conditions with germline mutation of the *PTEN* tumor suppressor gene. Little is known about the histopathologic features and molecular profile of PHTS-related breast cancers. Previous work found that CS-related breast cancer is associated with a molecular apocrine subtype (Banneau et al. 2010). Our study sought to validate these pilot observations in a large independent series and investigate other predictors of germline *PTEN* mutation status in a breast cancer series. A total of 1844 adult women with CS/CS-like features and invasive breast cancers were enrolled. Patients had germline *PTEN* PCR-based mutation analysis and MLPA/qPCR. Outside pathology reports and materials were requested for re-review and immunohistochemical analysis (IHC) of PTEN, pAKT, and androgen receptor (AR). Blood-PTEN/pAKT protein levels and *a priori* mutation risk via the Cleveland Clinic PTEN risk calculator (CC score) were available for 1258 patients. Nominal descriptive statistics, chi-square bivariate analyses, and multivariate logistic regression were used to find associations between potential predictors and germline *PTEN* mutation status. 43 female adults had germline *PTEN* mutation; pathology materials were available for 25. Most breast cancers had ductal histology (80%), and were ER/PR+ (64%), AR+ (75%) and HER2+ (14%). Apocrine features, atypical apocrine adenosis, and atypical ductal hyperplasia were found more often in background breast tissue of PHTS patients compared to published frequencies. Of 12 cases with complete biomarker data, 1 (8%) fit the molecular apocrine profile criteria. All tumors expressed PTEN and pAKT on IHC. Low blood-PTEN protein (p=0.01), high blood-pAKT (p=0.04), and higher CC score (p=0.01) were predictive of germline *PTEN* mutation. Breast cancer histology and molecular profile in these PHTS patients were comparable to general population patterns and not consistent with a molecular apocrine profile. Distinctive PTEN-associated features were noted in surrounding non-malignant breast tissue. Blood PTEN/pAKT protein levels and CC score were significant predictors of germline *PTEN* mutation, and if replicated, could be applied in a clinical setting to rapidly identify breast cancer patients who may benefit from genetics referral.

3427T

Further Defining the Polyposis Phenotype Associated with *PTEN* Mutations. *L. Panos[1], E. Weltmer[1], H. LaDuca[1], R. McFarland[1], E. Chao[1, 2].* 1) Ambry Genetics, Aliso Viejo, CA; 2) University of California, Irvine, School of Medicine, Irvine, CA.

PTEN hamartoma tumor syndrome (PHTS) is associated with an increased risk for colonic polyposis, particularly hamartomatous polyps. With the advent of multi-gene panel testing, *PTEN* mutations have been identified in patients with other polyp histologies, potentially widening the spectrum of gastrointestinal disease burden in these patients. We sought to define the polyp spectrum amongst *PTEN* mutation carriers identified by multi-gene panels. In a review of over 14,000 results of five multi-gene panels (*BRCA*plus, BreastNext, ColoNext, OvaNext, and CancerNext) that include full gene sequencing and gross deletion/duplication analyses of *PTEN* resulted between March 2012 and March 2014, 23 (0.15%) *PTEN* positive cases were identified. Clinical histories provided by ordering clinicians were reviewed, and clinicians were contacted to confirm history. Family history of polyposis was not assessed. Of the *PTEN* positive cases, 11/23 (47.8%) had a reported personal history of colonic polyps of varying type and quantity. Four patients met criteria for attenuated or classic familial adenomatous polyposis, with one having 20-99 adenomas, one having 20-99 adenomas and 20-99 hamartomas, and two having greater than 100 adenomas. Two patients presented with less than 10 adenomas, but had other polyp types including mucosal, inflammatory, hyperplastic, and a rectal lipoma. One patient presented with 10-19 hamartomas. In 4 individuals polyps were reported without further information. These results reveal that the polyp burden in *PTEN* positive patients expands beyond hamartomatous polyps. However, our study was limited to information reported by ordering providers, which may be influenced by patient recall, inability to access prior medical records, or lack of prior screening colonoscopy. Furthermore, subjective pathological review of polyps may complicate the characterization of CS-associated gastrointestinal manifestations. Additional studies focused on polyp histologies of *PTEN* mutation carriers would help further delineate the gastrointestinal disease burden in these individuals. Our findings suggest that *PTEN* mutations should be part of the differential diagnosis for individuals with adenomatous and mixed polyposis and support the use of multi-gene panel testing in patients with colonic polyposis.

3428S

Mutation and uncertain variant findings in ethnic minority patients undergoing multi-gene panel testing for cancer risk assessment at a safety-net public hospital. *CN. Ricker, JD. Sturgeon, SB. Gruber.* USC Norris Comprehensive Cancer Center 1441 Eastlake Ave. Los Angeles California 90033.

Rapid integration of multi-gene testing into clinical cancer genetics has lead to an emerging body of literature that has not yet focused on individuals from diverse racial/ethnic minority groups. We report on a retrospective IRB approved chart review of 170 underserved minority patients who underwent a 25 cancer gene test (Myriad Genetics) from September 2013 to May 2014 at Los Angeles County + USC Medical Center. The majority were female (n=155, 91.2%) and race/ethnicity was; 81.2% Latino (n=138); 11.8% Asian (n=20), and 7.1% Black (n=12). Eighty-one percent had a cancer diagnosis (n=137) with an average age at diagnosis of 44 years (SD=9), 60.6% (n=83) with breast cancer and 19.7% (n=27) with colorectal cancer. Overall, 15.9% (n=27) had a deleterious mutation identified and rates were comparable across race/ethnic groups. Eighteen percent of cancer patients had a mutation detected (n=25); 17 in high-risk genes (*APC* n=1; *BRCA1* n=6; *BRCA2* n=3; *CDH1* n=2; *MLH1* n=3; *MSH2* n=1; *TP53* n=1), 5 in moderate-risk genes (*ATM* n=2; *CHEK2* n=1; *RAD51C* n=1; *RAD51D* n=1) and 3 *MUTYH* heterozygotes. Two mutations were detected (*BRIP1* and *MUTYH*) in 33 unaffected women tested. A total of 133 variants of uncertain significance (VUS) were identified in 90/170 patients (52.9%), 41.1% with two or more VUSs and 12.1% with three or more. While not statistically significant, there was a marginally higher proportion of VUSs in individuals with cancer (0.54) compared to those without cancer (0.48). When stratified by race/ethnicity, the VUS rate was 83.3% for Blacks, 65.0% for Asians, and 48.6% for Latinos. Several variants were seen at least four times; *ATM* (c.2289T>A), *CDKN2A* (c. -2G>A), and *PTEN* (c.802-51_802_del14). The proportion of VUSs was compared to four published cohorts and was statistically significantly higher (p=0.001). Deleterious mutations were identified in 18% of cancer patients tested, 40.0% (10/25) in genes that would not have been interrogated prior to multi-gene panel testing, given the limited resource setting of this safety net hospital. Over 50% of individuals had at least one VUS, emphasizing the need for greater attention to the classification of VUSs in racial/ethnic minority patients. Further studies are necessary to expand understanding of implementation and utilization of genetic advances across broad clinical settings. Supported by Norris Cancer Center Core Grant NCI P30CA014089; ACS RSGT 1020301; Avon Foundation 052011057; Lynne Cohen Foundation.

3429M

Majority of *PTEN* mutations identified on multi-gene panel tests are in non-classic patients: Expanding clinical phenotype or incomplete clinical history? *E.C. Weltmer[1], L. Panos[1], H. LaDuca[1], R. McFarland[1], E.C. Chao[1,2].* 1) Ambry Genetics, Aliso Viejo, CA, USA; 2) University of California, Irvine, School of Medicine, Irvine, CA, USA.

PTEN hamartoma tumor syndrome (PHTS), including Cowden syndrome (CS), Bannayan-Riley-Ruvalcaba syndrome (BRRS), and *PTEN*-related Proteus syndrome (PS), is a phenotypically diverse but well-described clinical condition. Pathogenic *PTEN* mutations are detected in most patients meeting clinical criteria for CS and BRRS (up to 85% and 65%, respectively). However, individuals harboring *PTEN* mutations without meeting criteria for PHTS have been reported. In this study, we sought to characterize the clinical phenotypes of individuals who were found to harbor a pathogenic *PTEN* mutation through hereditary cancer multi-gene panel testing. A retrospective analysis of 14,897 hereditary cancer multi-gene panels including comprehensive analysis of *PTEN* reported at our laboratory from March 2012 to March 2014 identified 23 cases with pathogenic *PTEN* mutations (0.15%). Multi-gene tests included: high risk breast cancer panel (6 genes), moderate/high risk breast cancer panel (18 genes), colon cancer panel (14 genes), ovarian cancer panel (23 genes), and general cancer panel (28 genes). Clinical histories were reviewed for these cases, with emphasis on whether a clinical diagnosis of CS, BRRS, or other PHTS condition was suspected. Six of the 23 *PTEN*-positive cases (26.1%) met National Comprehensive Cancer Network (NCCN) CS/PHTS testing criteria. Most of the *PTEN*-positive cases in our study population (n=17, 73.9%) did not meet diagnostic or testing criteria for CS. Of these 17 cases who did not meet criteria, histories were confirmed in 15 by additional consultation with the ordering clinician. Seven individuals reportedly met one major and one minor criterion, five met one major criterion only, one met two minor criteria only, and two individuals met no CS criteria. Ordering clinician bias toward cancer history may result in incomplete clinical data and explain why most positive patients did not appear to meet CS criteria. It is also possible that a broader PHTS phenotype is being revealed through multi-gene panel testing, which is often ordered for individuals whose clinical features are atypical or of lower penetrance. Follow-up analyses of PHTS features in *PTEN*-positive individuals identified through panel testing, who have a broader spectrum of referral indications, will provide additional insight regarding the lack of CS in a panel-based PHTS-positive cohorts.

3430T

DNA methylation profiling to assess pathogenicity of *BRCA1* unclassified variants in breast cancer. *K. Flower[1], N.S. Shenker[1], M. El-Bahrawy[2], D.E. Goldgar[3], M.T. Parsons[4], A.B. Spurdle[4], J.R. Morris[5], R. Brown[1,6], J.M. Flanagan[1], KConFab Investigators, AFFECT study group.* 1) Epigenetics Unit, Department of Surgery and Cancer, Imperial College London, London, UK; 2) Department of Histopathology, Hammersmith Hospital, Imperial College London, UK; 3) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA; 4) QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia; 5) Genome Stability Unit, School of Cancer Sciences, University of Birmingham, UK; 6) Section of Molecular Pathology, Institute for Cancer Research, Sutton, UK.

Germline pathogenic mutations in *BRCA1* increase the risk of developing breast cancer, and can inform clinical decisions on treatment of breast cancer cases as well as risk reduction for mutation carriers. However, screening for mutations in the *BRCA1* gene frequently identifies sequence variants of unknown pathogenicity (unclassified variants) and recent work has aimed to develop methods for determining variant pathogenicity. We have previously shown that tumour DNA methylation profiles can differentiate *BRCA1* mutated tumours from *BRCA1* wild type tumours from high-risk families with no known mutations (BRCAx). We hypothesised that we could predict pathogenicity of variants based on DNA methylation profiles of tumours that had arisen in carriers of unclassified variants. We selected 150 FFPE breast tumour DNA samples (47 *BRCA1*, 65 BRCAx, 38 *BRCA1* test variants) and analysed a subset of these (n=60) using the Illumina 450K methylation platform. We identified 23 probes associated with mutation status (fdr p<0.05), and apparently independent of ER status and grade. Bisulphite pyrosequencing of the remaining 90 samples validated four independent novel loci (*BACH2*, *C8orf31*, *C17orf108* and *LOC654342*) to be associated with mutation status, but three of these were shown to be at least partially correlated with ER status and grade. These methylation markers were combined into a logistic regression model to predict pathogenicity of 27 unique variants from the independent test set, and LRs determined from methylation status used to calculate a posterior probability of pathogenicity. There were no major inconsistencies in predictions based on tumour methylation compared to those determined using standard multifactorial likelihood analysis that included information from segregation, co-occurrence, family history and ER-grade histopathological predictions. In particular, prediction of pathogenicity was consistent for *BRCA1* IVS 19-12 G>A with high probability of pathogenicity (P>0.99), whereas 15 variants were considered not pathogenic or likely not pathogenic using both approaches. A further two variants that are currently of uncertain classification using current methods show low probability of pathogenicity using our methylation prediction model. We conclude that tumour DNA methylation data has potential to be used in prediction of *BRCA1* variant pathogenicity.

3431S

Overexpression of MicroRNA-200c predicts poor outcome in patients with PR-negative breast cancer. *K. Luostari[1], M. Tuomarila[1], Y. Soini[1], V. Kataja[2], V.-M. Kosma[1], A. Mannermaa[1].* 1) Institute of Clinical Medicine, Clinical Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland; 2) Institute of Clinical Medicine, Oncology, University of Eastern Finland, Kuopio, Finland.

Micro-RNAs are small, noncoding RNAs that can act as tumor suppressors or oncogenes. MiR-200c is a member of the miR-200 family; it is known to be dysregulated in invasive breast carcinoma. MiR-200c maintains the epithelial-mesenchymal transition and inhibits cell migration and invasion. Recent studies have shown miR-200c to regulate steroid hormone receptors, estrogen receptors (ER), and progesterone receptors (PR). The present study detected miR-200c in 172 invasive breast carcinoma cases selected from a prospective cohort enrolled in Kuopio, Eastern Finland, between 1990 and 1995. MiR-200c expression was determined with relative qPCR, and results were compared to clinicopathological variables and patient outcome. We found that PR status combined with miR-200c expression was a significant marker of outcome. High miR-200c expression was associated with reduced survival in PR-negative cases (n = 68); low miR-200c expression indicated reduced survival in PR-positive cases (n = 86) (Cox regression: P = 0.002, OR = 3.433; and P = 0.004, OR = 4.176, respectively). In addition, in PR-negative cases, high miR-200c expression was associated with shortened relapse-free survival (Cox regression: P = 0.001, OR = 3.613); increased local/distant recurrence (Logistic regression: P = 0.006, OR = 3.965); and more frequent distant metastasis (Logistic regression: P = 0.015, OR = 3.390). We also found that high grade and low stage tumors were positively correlated with high miR-200c expression (Logistic regression: P = 0.002, OR = 2.791 and P = 0.013, OR = 3.911, respectively). Our results indicate that miR-200c may play a role in invasive breast carcinoma. Furthermore, miR-200c combined with PR status provides a refined predictor of outcome. This data may provide a basis for new research target - progesterone receptor - regulated microRNAs in breast cancer.

3432M

PDGFB hypomethylation is a favorable prognostic biomarker in primary myelofibrosis. *M. Miozzo[1, 2], C. Augello[1], F. Rossella[1], S. Tabano[1], E. Bonaparte[1, 2], M. Ciboddo[1], L. Paganini[1], D. Cattaneo[3], A. Iurlo[3], A. Cortellezzi[1, 3], S. Bosari[1, 2], S.M. Sirchia[4].* 1) Department of Pathophysiology and Transplantation, Università di Milano, Milano Milano, Italy; 2) Unit of Pathology, Fondazione IRCCS Ca' Granda Ospedale Maggiore, Milano; 3) Hematology and Transplantation UnitFondazione IRCCS Ca' Granda Ospedale Maggiore, Milano; 4) Department of Health Sciences, Università di Milano.

Primary myelofibrosis (PMF) is a myeloproliferative condition characterized by the clonal proliferation of the hematopoietic precursors and progressive development of bone marrow fibrosis. This stromal alteration is an important clinical issue and specific prognostic markers are not today available. JAK2, MPL and CALR genes are frequently mutated in PMF and are mutually exclusive. Mutations of TET2, ASXL1, DNMT3A and IDH1/2 are present in about 20% of PMF cases. It is intriguing that these loci are involved in the epigenetic regulation of cells and are associated with epigenetic alterations in cancer, suggesting that epigenetic defects could be also present in PMF. In 65 bone marrow biopsies from 58 PMF patients, stratified by clinical classification and JAK2/CALR and IDH mutations, we explored the methylation pattern of genes encoding for cytokines involved in the stromal reaction: PDGF, TGFB and FGF2. We also evaluated the methylation profile of the LINE-1. The methylation was analyzed using the pyrosequencing approach. In agreement with the European consensus of bone marrow fibrosis grading, the cases were categorized at diagnosis as cellular phase (MF-0 n=12) and fibrotic phases (MF-1 n=30; MF-2 n=18 and MF-3 n=5). The follow-up ranged from 6 to 287 months (average 76 months). As controls, we included 20 cases of iliac bone fragments from patients who underwent surgery for iliac bone prosthesis. We found that PDGFB, FGF2 and LINE-1, but not TGFB, methylation levels were heterogeneous and dynamic in the different phases of PMF compared to controls. The distribution of PDGFB and FGF2 methylation percentages in the subgroups of PMF cases showed that for both genes the methylation values in MF0 are significantly higher compared to controls (PDGFB, FGF2: $p < 0.0005$) and MF1 and MF2 cases showed a subgroup of cases with hypomethylation. PDGFB hypomethylation (<15%) was correlated with a favorable PMF prognosis ($p=0.03$, $p=0.01$ and $p=0.02$ for fibrosis, the International Prognostic Scoring System and the Dynamic International Prognostic Scoring System, respectively). In addition, low PDGFB methylation levels were mainly present in cases from young patients with normal karyotype and LINE-1 hypermethylation, genetic features associated with global genome stability.

3433T

MiR-145 regulates stem cell characteristics of human laryngeal squamous cell carcinoma Hep-2 cells. *M. Ozen[1,2], O.F. Karatas[2,3], I. Suer[2], B. Yuceturk[2], M. Yilmaz[4], H. Cansiz[4], M. Ittmann[1], M. Solak[5].* 1) Department of Pathology & Immunology Baylor College of Medicine, Houston, TX, 77030, USA; 2) Department of Medical Genetics, Istanbul University Cerrahpasa Medical School, Istanbul, Turkey; 3) Molecular Biology and Genetics Department, Erzurum Technical University, Erzurum, Turkey; 4) Department of Otorhinolaryngology, Cerrahpasa Medical School, Istanbul University, Istanbul, Turkey; 5) Afyon Kocatepe University Medical School Department of Medical Genetics.

The Cancer Stem Cells (CSCs) are tumorigenic cells promoting initiation, progression and spread of the tumor. Accumulating evidences has proven the existence of CSCs in a variety of tumors including lung, brain, breast, prostate, colon, head and neck and ovarian cancers. They are identified by their tissue specific stem cell-like properties including self-renewal and having potential to differentiate. miRNAs have been proposed as significant regulators of carcinogenesis and CSCs. A recent study that utilized Dicer or Dgcr8 knockout mice, which lacks global miRNA processing, showed that cells failed in self-renewal due to lack of downregulation of stem cell specific markers, pointing the importance of miRNAs in establishing stem cell identity. Furthermore, a number of miRNAs have been proposed to have direct roles in survival of CSCs. Understanding the contribution of miRNAs as regulators in these processes will help providing the opportunity to develop miRNA-based therapeutic approaches. In this study, we aimed to explore the role of miR-145, which is downregulated in laryngeal squamous cell carcinoma (LSCC), on cancer stem cell potency of laryngeal cancer cells. First of all, we demonstrated the downregulation of miR-145 expression in CD133+ cells directly isolated from freshly resected LSCC tumor tissues (n=20). qRT-PCR analysis of miR-145 transfected Hep-2 cells demonstrated the inhibition of stem cell markers such as SOX2, OCT4, KLF4 and ABCG2 upon miR-145 upregulation. We further investigated the stem cell features of miR-145 overexpressing Hep-2 cells by sphere formation assay, single cell cloning assay, colony formation assay, and Aldehyde Dehydrogenase assay, which all demonstrated the inhibition of stem cell potency upon miR-145 overexpression. In conclusion, we have demonstrated the regulatory role of miR-145 in stem cell characteristics of Hep-2 cells. Based on these results, we propose that mir-145 might carry crucial roles in LSCC tumorigenesis, prognosis, metastasis, chemoresistance and recurrence through regulating stem cell properties of tumor cells.

3434S

Methylation of *MLH3* Promoter: new recurrent finding in low grade gliomas. *H. Lhotska[1], Z. Zemanova[1], H. Cechova[2], L. Lizcova[1], S. Ransdorfova[2], F. Kramar[3], Z. Krejcik[2], K. Svobodova[1], D. Bystricka[1], K. Michalova[1,2].* 1) Center of Oncocytogenetics, General University Hospital and 1st Faculty of Medicine, Charles University, Prague, Czech Republic; 2) Institute of Hematology and Blood Transfusion, Prague, Czech Republic; 3) Department of Neurosurgery, Central Military Hospital and 1st Faculty of Medicine, Charles University, Prague, Czech Republic.

Astrocytomas and oligodendrogliomas (WHO grade II) are the most common histological subtypes of low-grade gliomas (LGGs). Several markers such as 1p/19q codeletion, mutations of *TP53*, *IDH1* and/or *IDH2* genes and epigenetic biomarkers, i.e. methylation of *MGMT* promoter have been used to predict patient's prognosis or response to treatment. The aim of this study was to investigate genetic and epigenetic background of LGGs in order to search for new markers that might play a role in tumor progression or patient's response to the treatment.

Biopsies from 41 patients with LGGs (25 astrocytomas, 16 oligodendrogliomas) were analysed using iFISH with the panel of VYSIS locus specific and centromeric probes (Abbott Molecular), SNP array (HumanCyto SNP-12 Bead Chip, Illumina), and MLPA (P370-A1 kit, MRC-Holland) to assess copy number variations and *IDH1*/*IDH2* mutations status in glial cells. Methylation of promoters was investigated using methylation-specific MLPA (ME011 kit, MRC-Holland).

In oligodendrogliomas, 1p/19q codeletion was frequent finding observed in 81% of patients. In astrocytomas, recurrent aberration was uniparental disomy of 17p (60%). Somatic mutation R132H of *IDH1* gene was detected in 81% of oligodendrogliomas and in 72% of astrocytomas. Methylation of *MGMT* promoter was present in 100% of oligodendrogliomas and in 84% of astrocytomas. *MLH3* (DNA mismatch repair gene) promoter methylation was observed in 69% of oligodendrogliomas and 28% of astrocytomas.

Methylation of *MGMT* promoter, 1p/19q codeletion and mutated *IDH1* were the most recurrent aberrations found in oligodendrogliomas. On the contrary, the findings in astrocytomas were more diversified. To our knowledge, this is the first time that methylation of *MLH3* gene promoter was described in LGGs. The loss of *MLH3* function leads to microsatellite instability which may contribute to tumor progression. Moreover, *MLH3* mutations have been associated with susceptibility to endometrial and colorectal cancers. The larger study of patients have to be investigated in order to reveal the role of methylated *MLH3* promoter as well as other detected aberrations in tumorigenesis of LGGs.

Supported by IGA MZ CR NT/13212-4, PRVOUK-P27/LF1/1, RVO-VFN64165.

3435M

A Custom 5m-Seq™ Cancer NGS Panel to Detect Epigenetic Signatures of Various Cancer Types. *J. Alexander, A. Meyer, J. Zhou, R. Drennan, M. Reddy, M. Poulin, L. Yan.* EpigenDx, Hopkinton, MA.

The number of genes that show some type of dys-regulation of DNA methylation in cancer cells continues to increase. In order to develop a panel for the determination of a DNA methylation pattern in cancer cells, we started with a candidate gene approach, in contrast to genome-wide methylation screening. Genes from different functional groups were selected. The cancer panel includes tumor suppressor genes (BRCA1, RASSF1, RUNX3), DNA methyltransferases (DNMTs and MGMT), DNA repeats (Line1 and Sat2), oncogenes (MYC,ERBB2), receptor genes (EGRF, ESRRA, NR2E1), Homeobox genes (HOXA10, HOXA11, EN1) and detoxification genes (GSTM1, GSTM2) and others. A total of 80 genes covering over 1500 CpG sites were selected and assays for these regions were developed and validated using PCR/Pyrosequencing, individually. The development of 5m-Seq™ Cancer NGS Panel using an Ion Torrent PGM™ system requires that the PCR amplicons be grouped into different pools based on the amplicon size, amplicon GC contents, and PCR conditions. DNA methylation controls (0%, 5%, 10%, 25%, 50%, 75%, and 100%) were sequenced on an Ion 314 Chip to validate each target region. The correlation between the calculated methylation levels and the detected methylation levels were examined. Additionally, 10 pairs of DNA samples (tumor vs adjacent normal) for each cancer type (ovarian cancer, breast cancer, cervical cancer, and colorectal cancer) were further tested using both an Ion Torrent PGM™ system and Pyroseqencing. The comparison of the results obtained by Pyrosequencing and Next-Gen Sequencing using PGM shows that PCR pools with common amplicons gave linear correlation. PCR pools with extreme conditions, i.e. AT-rich (less than 20% GC contents) or GC-rich (greater than 70% GC contents) required further optimization. In summary, high-throughput 5m-Seq™ NGS has the capability to generate equivalent or better results compared to Pyrosequencing with the potential of using less DNA for the analysis at lower cost. In addition, the panel gave different methylation signatures for each cancer type.

3436T

Differential DNA Methylation Patterns in Hereditary Non-polyposis Colorectal Cancer with or without Germline MLH1/MSH2 Mutation. *C.H. Chen[1,2], S.S. Jiang[1], L.L. Hsieh[3], R. Tang[4], I.S. Chang[1,2,5], C.A. Hsiung[5], H.J. Tsai[5].* 1) National Institute of Cancer Research, National Health Research Institutes, Zhunan, Taiwan; 2) Taiwan Bioinformatics Core, National Health Research Institutes, Zhunan, Taiwan; 3) Department of Public Health, Chang Gung University, Guieshan, Taoyuan County, Taiwan; 4) Colorectal Section, Chang Gung Memorial Hospital, Guieshan, Taoyuan County, Taiwan; 5) Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Taiwan.

Introduction: In spite of several DNA methylation studies reported in sporadic colorectal cancer, the aberrant global methylation patterns in hereditary non-polyposis colorectal cancer (HNPCC) remain unclear. It has been suggested that the global methylation patterns may vary among subtypes of HNPCC patients. Methods: A total of 40 HNPCC patients with tissue samples were incuded in this study. Global DNA methylation patterns were measured using Illumina Infinium HumanMethylation27 BeadChip in 40 HNPCC patients for both tumor and adjacent normal tissues. Paired t-test was used to identify CpG sites differentially methylated between tumor and adjacent normal tissues among HNPCC patients, among those with mismatch repair deficiency (MMP) and among those without such deficiency (MMN). Student's t test was applied to identity the differential methylated CpG sites between HNPCC patients with MMN and with MMP. Pathway analysis was applied to explore the global DNA methylation patterns between HNPCC patients with MMN and with MMP. Results: At false discovery rate 0.01 (q-values<0.01), 65.6% and 46.6% of CpG sites are hypomethylated in tumor tissues, compared with adjacent normal tissues, among MMN and MMP respectively. Our agnostic approach also confirmed the previous candidate gene approach reporting that hypomethylation in RUNX3, MLH1, NEUROG1, SOCS1 and CACNA1G in HNPCC MMN patients, compared with MMP patients. Conclusions: The findings provided suggestive evidence that global hypomethylation in tumor tissues are more prominent in HNPCC patients with MMN (Familial Colorectal Cancer Type X) than those with MMP (Lynch syndrome). The underlying pathogenesis between HNPCC patients with MMN and with MMP may be differentially regulated through methylation. Further investigation will be needed to gain more understanding in regulatory mechanisms of DNA methylation on HNPCC development and progression.

3437S

A Custom 5m-Seq™ Immunology NGS Panel to Detect Epigenetic Signatures in Various Cancer Types. *A. Meyer, J. Alexander, J. Zhou, R. Drennan, M. Reddy, M. Poulin.* EpigenDx, Hokinton, MA.

The study of the immune system in different cancers is of interest for the role it may be playing in the onset and development of the disease and in the role immunotherapy might play in the treatment of different cancer types. The epigenetic regulation of immunologically relevant genes in different cancer types may provide important information for the role they may play in cancer etiology. In order to develop an immunology panel for the determination of a DNA methylation pattern in cancer, we started with a candidate gene approach, in contrast to genome-wide methylation screening. Genes from different functional groups were selected as candidates for methylation analysis. The immunology panel includes interleukins and their receptors (i.e. IL1B, IL6, IL8, IL27RA), TNFs and their receptors (i.e. TNFA, TNFS11, TNFRSF25), Interferon-gamma, chemokines and their receptors (i.e. CXCL10, CXCR1), and CD families (i.e. CD5, CD21), cell adhesion molecules (i.e. BCAM, ICAM1) and others such as the insulin gene and its receptor (INS, INSR) and PDCD1. A total of 50 genes with over 100 amplicons covering over 500 CpG sites were selected and assays for these regions were developed and validated using PCR/Pyrosequencing, individually. The development of a 5m-Seq™ Immunology NGS Panel using an Ion Torrent PGM™ system requires that the PCR amplicons be grouped into different pools based on the amplicon size, amplicon GC contents, and PCR conditions. DNA methylation controls (0%, 5%, 10%, 25%, 50%, 75%, and 100%) were sequenced on an Ion 314 Chip to validate each target region. The correlation between the calculated methylation levels and the detected methylation levels were examined. Additionally, 10 pairs of ovarian cancer DNA samples (tumor vs adjacent normal) were further tested using both an Ion Torrent PGM™ system and Pyroseqencing. The comparison of the results obtained by Pyrosequencing and Next-Gen Sequencing using PGM shows that PCR pools with common amplicons gave linear correlation. PCR pools with extreme conditions, i.e. AT-rich (less than 20% GC contents) or GC-rich (greater than 70% GC contents) required further optimization. In summary, high-throughput 5m-Seq™ NGS analysis has the capability to generate equivalent or better results compared to Pyrosequencing with the potential of using less DNA at lower cost. These results support future use of a custom 5m-Seq™ panel for efficient and sensitive epigenetic detection and analysis of cancer samples.

3438M

HES1 gene expression in patients with Medullary Thyroid Cancer is independent of its promoter methylation. *M.G. Cardoso[1], M.M.L. Kizys[2], S.C. Lindsey[2], J.H. Lee[3], D.F. Ierardi[4], C.P. Camacho[2], R. Delcelo[3], J.M. Cerutti[5], R.M.B. Maciel[2], M.G. Jasiulionis[4], M.R. Dias-da-Silva[2].* 1) Biochemistry, Universidade Federal de São Paulo, São Paulo, Brazil; 2) Medicine, Universidade Federal de São Paulo, São Paulo, Brazil; 3) Pathology, Universidade Federal de São Paulo, São Paulo, Brazil; 4) Pharmacology, Universidade Federal de São Paulo, São Paulo, Brazil; 5) Genetics, Universidade Federal de São Paulo, São Paulo, Brazil.

Abnormal methylation seems to play an important role in differentiated thyroid carcinomas, in which several of the tumor suppressor genes are epigenetically silenced. Little is known about the methylation profiles of genes that are associated with medullary thyroid carcinoma (MTC [MIM 155240]). Studies have reported that aberrant Notch signaling in MTC results in the downregulation of the HES1 [MIM 139605] gene, increase in expression of the tumor markers chromogranin A and calcitonin. Therefore, we aimed to determine if the downregulation of HES1 is caused by hypermethylation at its regulatory region. From 128 patients followed for MTC, we obtained formalin-fixed, paraffin-embedded (FFPE) tumor tissues from 29 patients, and peripheral blood samples from 18 patients with persistent disease. We also studied a human MTC cell line (TT). For HES1 analysis in tumor DNA we used bisulfite sequencing and methylation-specific PCR. HES1 protein expression was evaluated by immunohistochemistry , and mRNA levels were evaluated by RT-qPCR of peripheral blood. We observed increased expression of HES1 in peripheral blood samples of patients with persistent MTC compared with healthy individuals (p=0.007). Likewise, immunohistochemical analysis revealed moderate staining for HES1 in MTC tissues. Hypermethylation of a CpG island previously implicated in HES1 silencing was assessed, but no evidence of methylation at this promoter was found in sporadic or familial MTC tissue, the TT cell line or nontumorigenic tissues. The finding of increased expression of HES1 observed in persistent MTC is not mediated by methylation of DNA in its promoter region. This expression of HES1 suggests that the Notch signaling pathway may be involved in the cellular response to the constitutively active MAPK signaling in MTC tumorigenesis. Funding: FAPESP.

3439T

Genetic characterization of near-haploid and low hypodiploid acute lymphoblastic leukemia. *S. Safavi, B. Johansson, K. Paulsson.* Division of Clinical Genetics, Department of Laboratory Medicine, Lund University, Sweden.

Near-haploid (23-29 chromosomes) and low hypodiploid (HoL; 30-39 chromosomes) acute lymphoblastic leukemia (ALL) are two rare subtypes associated with a dismal prognosis. The aim of the present study was to investigate the underlying genetic mechanisms in these cases, in order to detect genetic aberrations and imbalances specific for near-haploidy/HoL. We performed methylation array analysis, whole transcriptome sequencing and whole exome sequencing on 9 near-haploid and HoL cases. Results from methylation array and gene expression analyses show that near-haploidy forms a group with an expression profile separate from HoL, emphasizing the divergent genetic nature of these two subtypes. Methylation array analysis detected 100 genes that showed significantly different methylation in near-haploid/HoL group compared with the control group ($P < 0.05$, false discovery rate (FDR) < 0.1). Forty-five genes showing a significantly different expression between the near-haploid/HoL group and the control group ($P < 0.05$, FDR < 0.2) were detected by gene expression analysis. *PARD3B*, involved in cell polarization, was one of four genes found to be upregulated in both subtypes and downregulated in the control group. By genetically characterizing near-haploidy and HoL ALL we can potentially identify therapeutic targets that may be used in a clinical setting for an improved prognosis.

3440S

Functional HPSE gene SNP rs4693608 modifies heparanase expression and thereby affects the responsiveness to broad number of treatments. *O. Ostrovsky[1], P. Baryakh[1], A. Shimoni[1], Y. Margulis[1], M. Mayorov[1], I. Vlodavsky[2], A. Nagler[1].* 1) Bone Marrow Transplantation, Sheba Medical Center, Ramat Gan, Israel; 2) Cancer and Vascular Biology Research Center, Rappaport Faculty of Medicine, Technion, Haifa, Israel.

Heparanase is an endo-glucuronidase, which responsible for heparan sulfate degradation, plays important roles in processes of angiogenesis, tumor metastasis, inflammation and autoimmunity and regulates histone methylation by binding to target gene control regions. Our previous studies indicated that HPSE gene SNPs (rs4693608, rs11099592 and rs4364254) significantly correlated with expression level of heparanase among healthy persons. SNP rs4693608 was the most prominent. Analysis 414 patients with hematological malignancies and their donors revealed a highly significant correlation of rs4693608 with the risk of developing graft vs. host disease (GVHD). The discrepancy in this SNP between recipients and donors was found to be a more significant factor for the risk of aGVHD than patient genotype. Expression of the HPSE gene was increased following pre-transplantation treatment and correlated with the rs4693608 both before and after conditioning. In the present study we analyzed HPSE gene expression in response to LPS treatment in 128 umbilical cord blood (CB) and 104 peripheral blood (PB) samples. LPS was found to up-regulate HPSE gene expression ($P<10^{-7}$) through TLR4. Post-treatment heparanase expression correlated with rs4693608 in both PB and CB MNCs. RQ (relative quantification) in PB MNCs of AA carriers was 16.3 (7.5-34.5), while RQ in GG possessors was 6.4 (3.3-19.9), P=0.014. Analysis of HPSE gene expression in CB MNCs showed similar results (24.9 (15.2-40.5) *versus* 5.4 (3.1-13.0), P=0.0006). In addition, we examined association between HPSE gene SNPs and $CD34^+$ cell mobilization in 279 healthy individuals receiving granulocytes colony-stimulating factor (G-CSF). We found significant correlation between rs4693608 and total $CD34^+$ cells$\times 10^6$ (754 *versus* 502; P=0.02). Opposed effect of rs11099592 to mobilization of $CD34^+$ cells was identified. Allele A of rs11099592 was found in linkage disequilibrium with allele G of rs4693608. However, the mobilization of $CD34^+$ cells was higher in GG-AA carriers in comparison to GG-GG possessors (1022 *versus* 506; P=0.005). The present study indicated that the level of heparanase strongly correlates with the rs4693608 and depend on cell type and it activity. Functional HPSE gene SNPs may be used as a marker for prediction of acute inflammation, useful key for analysis of heparanase involvement in various biological processes and may lead, in the future, for correct heparanase level modification by potential inhibitors.

3441M

Distinct molecular signature of Giant Cell Tumor occurring in pagetic or non-pagetic patients suggests distinct pathologic entities. *F. Gianfrancesco[1], G. Divisato[1], D. Formicola[1], D. Rendina[2], M. De Lucia[1], L. Pazzaglia[3], L. Michou[4], D. Merlotti[5], M.S. Benassi[3], T. Esposito[1], L. Gennari[5].* 1) Institute of Genetics & Biophysics, National Research Council of Italy, Naples, Italy; 2) Department of Clinical and Experimental Medicine, Federico II University, Naples, Italy; 3) Laboratory of Experimental Oncology, Rizzoli Orthopaedic Institute, Bologna, Italy; 4) Department of Medicine, Faculté de Médecine de l'Université Laval, Quebec City, Canada; 5) Department of Clinical, Surgical and Neurological Sciences, University of Siena, Siena, Italy.

Giant cell tumor of bone (GCT) is an aggressive bone tumor caused by the uncontrolled proliferation of the spindle-like stromal cells which promote osteoclast-like giant cells formation responsible for the osteolitic lesions. PDB patients who develop GCT, show the polyostotic form of disease (93%), an increased prevalence of familial PDB (59%) and an earlier age at diagnosis of PDB (52±12 yrs). In contrast, in non-pagetic patient GCT mainly occurs in Asiatic subjects while pagetic GCT affect white Caucasian patients in up to 83% of cases and is rarely multifocal (<1% vs 25% in pagetic GCT). The skeletal localization of GCT is also different between pagetic and non pagetic patients, with involvement of the axial skeleton in 75% vs 15% of cases, respectively. A recent study showed that GCT is due to recurrent somatic mutations in H3F3A gene in the stromal cells. We analysed a cohort of giant cell tumor of bone for the presence of H3F3A mutations identifying somatic mutations in 38 out of 44 cases (86%). In contrast, the analysis of patients with Paget's disease of bone (PDB) associated with giant cell tumor did not show any mutation in H3F3A gene, at both somatic and germline level, suggesting a different genetic background. Recently we reported an extended Italian family in which 4 out of 14 PDB affected members developed multiple GCTs at pagetic skeletal sites. Clinically, all affected members had polyostotic PDB, but subjects developing giant cell tumors showed an increased disease severity with a reduced clinical response to bisphosphonate treatment and an increased prevalence of bone pain, deformities, and fractures. Whole exome sequencing, in this family identifies a missense mutation in a novel uncharacterized gene. Additional genetic analysis in 7 independent affected patients with the same clinical phenotype, discloses the same mutation in all patients, strongly suggesting that this clinical phenotype is due to a founder effect. Therefore, GCT associated or not with PDB is due to mutations in different genes, suggesting different molecular signatures.

3442T

miR-22 supresses cell proliferation in zoledronic acid treated glioblastoma cells by inducing autophagy and targeting mTOR. *C. Caliskan[1], B. Goker[1], Z. Mutlu[1], C. Kayabasi[1], S. Yilmaz Susluer[1], T. Balci[1], B. Ozmen Yelken[1], N. Oktar[2], C. Gunduz[1], C. Biray Avci[1].* 1) Ege University Medical School Department of Medical Biology, Izmir, Turkey; 2) Ege University Medical School Department of Neurosurgery, Izmir, Turkey.

Glioblastoma multiforme (GBM) is the most common and malign type tumors of central nervous system in elderly people. It is non-responsive to chemotherapy in many cases and carries the lowest chances of survival. Despite various treatment, the median survival time of patients with GBM is 12-16 months. FDA approved for treatment of metastatic bone disease is one of nitrogen-containing bisphosphonate is zoledronic acid (ZA). It demonstrates anticancer activity in various cancers. MicroRNAs are small (19-24 nucleotides) and non-coding RNA that regulate post-transcriptional gene expression via disrupting the stabilization of target transcripts or inhibiting protein translation. miRNAs serve as oncogenes and tumor suppressors through unlighted mechanisms in human. Autophagy is a physiological process that triggered degradation of cellular components by lysosomal mechanism and controlled by genetically. The aim of the study was to evaluate the expressions of miRNAs that are associated with GBM genetics after treatment with zoledronic acid; to investigate target genes of miRNAs which show significant expression alterations and molecular mechanisms of zoledronic acid treatment. U87-MG cell line (human glioblastoma-astrocytoma) is used as an in vitro model to investigate the cytotoxic and apoptotic effect of ZA in glioma cells. U87-MG cells were treated with 25 μM (IC50) ZA during 72 hours and cytotoxicity was evaluated by using WST-1 assay. Apoptosis assays were performed by using ApoDIRECT In Situ DNA Fragmentation Assay. The RT-qPCR is used for miRNA expression analysis. Results showed that IC50 dose of ZA induced apoptosis 4,25 fold when compared to control cells that untreated with ZA. Also IC50 dose of ZA of miRNA expression results showed that; miR-22 expression level was upregulated 3,08 fold according to control group. Also RT-qPCR results showed that; Bcl-2 gene expression decreased 2,14 fold according to control cells. Inhibition of Bcl-2 expression showed that ZA induces autophagy. In additional, Akt1 and mTOR genes' expressions decreased 3,7 and 3,36 fold, respectively according to control cells. Inhibition of AKT/mTOR cascade can allow the phosphorylation/inactivation of Bcl-2. These novel findings showed that ZA is very important in glioma progression and it is necessary to question whether it can be used as a drug candidate in glioma treatment with further research on miR-22 and its target gene expression.

3443S

MicroRNA in biofluids - Robust biomarkers for disease. *D. Andreasen[1], A.R. Thomsen[1], T. Blondal[1], J. Krummheuer[1], M.W. Teilum[1], N. Tolstrup[1], M. Borre[2], C. Haldrup[3], T.F. Ørntoft[3], K.D. Sørensen[3], P. Mouritzen[1].* 1) Exiqon A/S, Vedbaek, Denmark; 2) Dept of Urology, Aarhus University Hospital, Denmark; 3) Dept of Molecular Medicine, Aarhus University Hospital, Denmark.

microRNAs constitute a class of small cellular RNAs (typically 19-23 nt) that function as post-transcriptional regulators of gene expression. Current estimates indicate that more than one third of the human cellular transcriptome is regulated by this small class of RNA (~2000 miRNA). The study of extracellular microRNAs and their potential as pathophysiological markers has greatly expanded in the last couple of years. microRNAs have been shown to be actively exported from tissues into the circulation through a variety of mechanisms including exosome and microvesicle transport, and complexing with RNA binding proteins or HDL. The high relative stability of microRNAs in common clinical source materials (FFPE blocks, plasma, serum, urine, saliva, etc.) and the ability of microRNA expression profiles to accurately classify discrete tissue types and specific disease states have positioned microRNAs as promising new biomarkers for diagnostic application in cancer. We have applied Exiqons highly sensitive LNA™-based qPCR platform for detection of microRNAs, which has enabled microRNA profiling in biofluids where levels are extremely low. The platform uses a single RT reaction to conduct full miRNome profiling and allows high-throughput profiling of microRNAs without the need for pre-amplification. Thousands of biofluid samples including serum/plasma and urine have been profiled to determine normal reference ranges for circulating microRNAs as well as to identify biomarkers of disease. Extensive data qualification and analysis methods have been developed and are central parameters to secure high quality data from biofluids. The methods can quickly and robustly be applied in biomarker discovery and validation projects. We will present examples from our collaborative cancer diagnostic projects. Also we recently developed a new exosome enrichment method and will present a characterization of the exosome fraction obtained with this method. We will also present a comparison of microRNA profiles obtained with this method to profiles obtained with different commercially available exosome isolation methods. For plasma and serum we will furthermore compare to standard profiles of non-fractionated samples.

3444M

Resveratrol up-regulates tumor suppressor mir-31 expression via inhibiting histon deacetylase 1 gene expression in chronic myeloid leukemia. *C. Biray Avci[1], Z. Mutlu[1], C. Caliskan[1], B. Goker[1], S. Yilmaz Susluer[1], F. Sahin[2], G. Saydam[2], C. Gunduz[1].* 1) Ege University Medical School Department of Medical Biology, Izmir, Turkey; 2) Ege University Medical School Department of Hematology Izmir, Turkey.

Chronic myeloid leukemia (CML) is characterized by a reciprocal translocation between chromosomes 9 and 22. The Philadelphia chromosome is available as a result of this translocation and causes BCR-ABL1 fusion gene which produces actively tyrosine kinase. Resveratrol (RES) is a natural phytoalexin found in grapes and induces apoptosis, erythroid differentiation and autophagy in leukemic cells. MicroRNAs are small (~22 nucleotides), single strand, non-coding RNA molecules that regulate post-transcriptional gene expression. miRNAs cause modulation of oncogenic or tumor suppressive pathways in different cancer types. In this study we aimed to determine cytotoxic effect of RES in K562 human CML cell line and to evaluate the expressions of miRNAs that are related with leukemogenesis after the treatment with RES. Also we analyzed target genes of miRNAs which show notable expression levels. K562 cells were treated with 100 μM (IC50 dose) RES during 72 hours and cytotoxicity was evaluated by using WST-1 assay. The RT-qPCR is used for miRNA and gene expression analysis. miRNAs and gene expression levels were evaluated by using miScript miRNA PCR Array and RT2 Profiler PCR Array, respectively. Results showed that; RES up-regulated tumor suppressor miR-31 level 3.60 fold and significantly down-regulated HDAC1 gene expression (p=0.003), according to the control cells. Our findings showed that Resveratrol acts as a HDAC inhibitor targeting HDAC1 gene expression level. HDACs play a key role in the regulation of gene expressions in cell progressions such as tumorigenesis and cell proliferation. Downregulation of HDAC1 provides post-translational modification for expression of tumor supressor genes and leads to cell cycle arrest and increases apoptosis. These results provide that Resveratrol could be a therapeutic candidate as a HDAC inhibitor for chronic myeloid leukemia treatment.

3445T

Association of Ile655Val polymorphism of the HER2 gene with Neutropenia toxicity in breast cancer patients treated with trastuzumab chemotherapy. *D.I. Carrillo-Moreno[1, 2], L.E. Figuera[3], L. Gómez Flores[1], A. Ramos[1, 4], R. Ramírez[1, 2], O. Soto[1, 4], I. Gutíerrez[1, 2], G.M. Zúñiga[5], M.P. Gallegos[1].* 1) Laboratorio de Genética Molecular, División de Médicina Molecular, CIBO, IMSS, Guadalajara, Jal,. Mex, Mexico; 2) Doctorado en Genética Humana, Centro universitario de Ciencias de la Salud, Universidad de Guadalajara,; 3) División de Genética, CIBO, IMSS; 4) Doctorado en Farmacología, Centro universitario de Ciencias de la Salud, Universidad de Guadalajara; 5) Loboratorio de Mutagenesis, División de Médicina Molecular, CIBO, IMSS, Guadalajara,.

Background: The influence of *Ile655Val* polymorphisms in the *HER2* gene involved in trastuzumab metabolism has been studied in breast cancer. *HER2* (*erbB-2, neu*) is a proto-oncogene which encodes a transmembrane protein with tyrosine kinase activity but with no identified physiological ligand. The *HER2* gene is amplified in 30% of invasive breast cancers and correlated with reduced patient survival. Our aim was to evaluate the association of Ile655Val polymorphisms in the *HER2* gene with toxicity effects in breast cancer patient's treated with trastuzamab chemotherapy. Methods: DNA genomic samples from 175 patients (UMAE gynecology and obstetrician Hospital, CMNO, IMSS), that received trastuzumab chemotherapy; were included in the study. The *Ile655Val* polymorphism was determinate by polyacrylamide gels electrophoresis, previously PCR and BsmAI restriction enzyme analysis. The association was determinate by odds ratio. Results: The genotype 677*ValVal-IleVal* was associated with gastrointestinal [3.1(IC95% 1.11-9.138), p=0.030] and hematological [3.2(IC95% 1.12-9.08), p=0.029] toxicity (diarrhea and neutropenia grade II- III) respectively in breast cancer patient. Conclusion: The *Ile655Val* polymorphisms in the *HER2* gene could be a good marker of toxicity in breast cancer patients treated with trastuzamab in the analyzed sample.

3446S

The BIM Deletion Polymorphism Cannot Account for Intrinsic TKI Resistance of Chinese CML Patients. *X. Chen[1], H. Liu[2], Y. Zhang[2], F. Wang[2], W. Teng[2], P. Zhu[1], C. Tong[2], D. Lu[2].* 1) Department of Hematology, Peking University First Hospital, Beijing, beijing, China; 2) Medical Laboratory Division, Ludaopei Hematology & Oncology Center, Beijing.

Background: BIM is a pro-apoptotic protein of the Bcl-2 family. Recently it has been reported that a common deletion polymorphism in the BIM gene mediates intrinsic resistance to imatinib mesylate (IM), a tyrosine kinase inhibitor (TKI). Chronic myeloid leukemia (CML) patients with the deletion experienced inferior responses to IM. **Method and Cases:** To identify the influence of this deletion on TKI responses in Chinese individuals, we collected 141 newly diagnosed patients with chronic phase CML whose first-line therapy was a standard dose of IM and 200 healthy individuals. Patients were classified as resistant or sensitive to IM according to the European Leukemia Net (ELN) criteria. The 2,903bp deletion of BIM was detected by PCR. **Results:** 1) 20 patients were heterozygous for the deletion and the other 121 did not harbor the deletion. 2) Of the patients with the deletion, 15 (75%) were sensitive to IM and only 5 (25%) were resistant. For patients without the deletion, 86 (71%) were sensitive to IM and 35 (29%) were resistant. 3) 41 of 200 healthy individuals carried the deletion. One was homozygous and the other 40 were heterozygous. The carrier frequency was 20.5%, which was much higher than previously reported (12.3%) and indicated a different genetic background. 4) Logistic regression analysis shows the overall odds ratio for resistance among patients with the deletion compared to those without it was 1.221 (P=0.719, 95% CI 0.412-3.616). **Conclusion:** The BIM deletion polymorphism cannot account for intrinsic TKI resistance of Chinese patients with CML.

3447M

Association of CD44 expression before, during and after treatment in patients with head and neck cancer in comparison with healthy controls. *K. Chukka[1], Z. Vishnuvardhan[2], S. Dasari[3], U. Radhakrishna[4].* 1) Department of Biotechnology, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, India; 2) Department of Botany & Microbiology, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, India; 3) Department of Biotechnology, Dravidian University, Kuppam, India; 4) Green cross Pathology and Molecular Biology laboratory, Paldi, Ahmedabad, India.

Cancer Stem Cells (CSCs) are presumed to be responsible for recurrence of cancer even after treatment. CSCs have been identified in various types of cancers, including head and neck (HNC) cancer, breast cancer, prostate cancer and pancreatic tumors. The lack of reliable markers to identify CSCs is the major problem in the development of new strategies to tackle cancer. CD44 molecule is one of the promising surface markers of CSCs in HNC patients to identify and assess the stage of cancer diagnosis before and later treatment. We have measured the levels of CD44 expression in 40 HNC cases undergoing treatment and compared with post treatment levels of marker protein. The expression levels of CD44 in normal controls (healthy individuals), cases before, during and after treatment was significantly correlated. Low levels correlated with healthy condition and positive response to cancer treatment. A lower CD44 expression was observed in those from the healthy controls (269.3±46.3) compared to cases. Also observed lower levels of CD44 expression after treatment (475.24±7.27) compared to during (549.38±9.77) and before treatment (583.64±8.82) respectively. There is a significant difference between HNC cases and controls as well as cases before, during and after treatment at 5% level. CD44 appears to be a suitable marker of head and neck squamous cell carcinoma having potential diagnostic value, since its detection is easy.

3448T

PCA3 prostate cancer biomarker long non-coding transcription unit: Transcriptional interference of overlapping genes. *R. Clarke[1,2], M. Lavin[2], R. Gardiner[2,3], C. Chow[2], R. Stirling Richards[2], Z. Fang[1].* 1) Ingham Institute University of Western Sydney School of Medicine, Liverpool, NSW, Australia; 2) University of Queensland Centre for Clinical Research, Brisbane, Qld, Australia; 3) Department of Urology, Royal Brisbane & Women's Hospital.

PCA3 is arguably the most specific diagnostic biomarker for prostate cancer with great potential to improve clinical screening of men at risk. Given the over-riding need for prognostic and therapeutic biomarkers for prostate cancer the question remains as to whether PCA3 has a functional role in the disease and whether it could serve as a target or facilitator of therapy. The PCA3 gene appears to be a long non-coding transcription unit which is transcribed at high levels in prostate cancer. The PCA3 gene recently emerged in primates to give rise to a processed transcript with weak conservation and no detectable PCA3 protein product suggesting that PCA3 may have no direct functional role in prostate cancer. However, PCA3 is nested within another gene, BMCC1, and the transcription of PCA3 is discordantly regulated with BMCC1 in prostate cancer cells and after androgen treatment. BMCC1 is involved in Rho signalling and vesicular trafficking and may play roles in cellular transformation and metastasis. We apply novel technology to the BMCC1/PCA3 overlapping gene complex that suggests that the nested status of PCA3 within BMCC1 modulates transcription in a discordant fashion. These results suggest that PCA3 transcription may be a surrogate biomarker for BMCC1 and useful not only for the early detection of prostate cancer but to better understand prostate cancer initiation and development. If functional, interference of PCA3 transcription may represent a possible therapeutic application for early prostate cancer. Alternatively, novel PCA3 transcription products specific for prostate cancer could be used to selectively trigger a lethal exogenous target vector.

3449S

Creation of an open data sharing exchange to optimize BRCA clinical variant assessment. *N. Conti[1], C. Strom[2].* 1) Quest Diagnostics, Madison, NJ; 2) Quest Diagnostics Nichols Institute, San Juan Capistrano, CA.

Following the Supreme Court of the United States (SCOTUS) decision invalidating certain BRCA gene patents, several clinical laboratories in the US began to offer BRCA sequence testing. As with any sequence based test, some patients will have sequence variants that require clinical assessment Clinical assessment encompasses literature review, population frequency determination, database searches, segregation and coinheritance analyses, computer modeling, and functional studies. Databases aid in this process by summarizing non-published data (such as segregation analyses) and previous interpretive conclusions, and serving as a gateway to published literature. Careful review of databases and confirmation, in general, are required. The Universal Mutation Database (UMD) collects data from 16 contributing laboratories throughout France. The UMD is actively curated. As of May 2014, UMD contained 6902 BRCA1 (MIM 113705) variants and 8082 BRCA2 (MIM 600185) variants. UMD is comparatively richer in BRCA content than US databases, because European labs have been doing BRCA testing for more than a decade. Until the recent SCOTUS decision, BRCA testing in the US had been limited to a single laboratory that no longer contributes variants to public databases. Our objective was to optimize patient care through facilitated access to UMD and to create a mechanism to assure contributions of new variant information to the database. Quest Diagnostics obtained a commercial use license for UMD with the stipulation that any research laboratory would continue to have free access to the database. Any commercial laboratory willing to share their data and provide user group fees through a user group will also have access to UMD. The fees will allow UMD to expand access with both hardware and software upgrades, develop processes for incorporation of new variant data, and increase the number of curators for the expected increase in uploading of new variants. The user group agreement also allows participating laboratories to request functional studies for families with potentially clinically relevant variants. The planned transparency of the processes will permit independent assessment of the level of evidence for any variant in the UMD database and provide the scientific and medical communities with a valuable resource to optimize research and patient care.

3450M

Towards the minimal breast cancer genome and its relevance to chemotherapy. *S.N. Dorman[1], J.H.M. Knoll[2], K. Baranova[1], C. Viner[3], P.K. Rogan[1,3].* 1) Biochemistry, University of Western Ontario, London, Ontario, Canada; 2) Pathology, University of Western Ontario, London, Ontario, Canada; 3) Computer Science, University of Western Ontario, London, Ontario, Canada.

We are investigating whether the variable response to chemotherapy in breast cancer (BC) can be explained by the diversity of somatic mutations among patients. Our laboratory proposed that there is a minimal genome required for BC cell survival. A stable set of 5,804 genes was derived by comparing regions of BC genomes unaltered in copy number (CN) with genes normal in expression levels in tumours (*Mol. Oncol.* 6: 347-359). A subset of these genes are targets of well-established therapies (ie. paclitaxel, gemcitabine). We hypothesize that that the effectiveness of these or other chemotherapy treatments may rely on the integrity of stable genes. We analyzed integrated genomic data for BC tumours from The Cancer Genome Atlas to re-define and/or narrow the minimal BC genome. Consistent with the design of our previous study, we identified genes stable in CN and gene expression, but now also exclude genes with deleterious somatic point/indel mutations (in ≥ 90% tumours). An analysis of 318 tumours revealed 6,994 stable genes, of which 2,364 genes were identified in the previous stable gene set. There are 3,856 unstable genes with deleterious somatic point/indel mutations, of which 32 genes exhibit ≥ 10 mutations in different tumours. Removing somatic point/indel mutations reduced the stable gene set by 122 genes. This is because unstable genes were often mutated by different mechanisms: 3,353 genes with somatic mutations also showed either abnormal CN or gene expression. Dysregulated pathways present in both stable gene sets (n = 265) are enriched for transcription (n = 37), mRNA processing and transport (n = 41), translation (n = 16), signaling (n = 19), and the formation of protein complexes (n = 19). Growth inhibition studies (GI50 values) of 27 BC cell lines show that chemotherapy sensitivity is not consistent across all cell lines (*Genome Biol.* 2013, 14:R110). Paclitaxel and gemcitabine metabolic pathways include 33 (9 stable) and 14 (7 stable) genes, respectively. We related GI50 of BC cell lines, to CN, gene expression, and somatic mutation data for these genes. Paclitaxel sensitivity was correlated with *MAPT*, *BCL2*, and *CSAG2* expression levels, and somatic mutations in *CYP2C8*, *FGF2*, and *TWIST1*. *BCL2* and *MAPT* are direct targets, and the remaining genes have previously been associated with paclitaxel resistance. Growth inhibition by gemcitabine was inversely correlated with *RRM2* and *RRM2B* expression, which interact with known drug targets.

3451T

Whole exome sequencing approach in sib pairs identifies oligogenic germline mutations predisposing to early lung adenocarcinoma in non-smokers. *E. Frullanti[1], M.A. Mencarelli[1], F. Cetta[2], M. Baldassarri[1,3], F. Mari[1,3], S. Furini[4], P. Piu[5], T.A. Dragani[6], F. Ariani[1,3], A. Renieri[1,3].* 1) Azienda Ospedaliera Universitaria Senese, Siena, Siena, Italy; 2) IRCCS MultiMedica, Milan, Italy; 3) Medical Genetics, University of Siena, Siena, Italy; 4) Department of Medical Biotechnology, University of Siena, Siena, Italy; 5) Department of Medicine, Surgery & Neuroscience, University of Siena, Siena, Italy; 6) Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.

A polygenic model is commonly assumed for the predisposition to common cancers. Despite the great bulk of research to identify genetic susceptibility genes in lung cancer by genome-wide association studies, only three loci have been identified at 15q25, 5p15.33, and 6p21, respectively. It is noteworthy that findings have not been replicated consistently in subsequent studies. In addition to confer a very low risk, they have been associated with lung cancer in smokers, but not in non-smokers. The polygenic nature of common cancers has frequently been suggested, but its biological basis still remains elusive. We tested the hypothesis that genetic susceptibility may rely on a restricted number of genes, disrupted by germ-line mutations. A combination between an advanced technical tool, i.e. the exome sequencing, and a new patient selection strategy was used. The strategy relies on the selection among 964 lung adenocarcinoma patients those with early onset disease (mean age 43) in absence of cigarette smoking, and having a first degree healthy sibling available for genome comparison with an age difference of at least 7 years. Germ-line truncating mutations were detected in 8 and 5 different cancer predisposing genes in each affected subject, respectively, but not in the healthy sib of two pairs (p=0.0026). Some of them are well known cancer players in lung tumors and others are genes previously identified in other cancer tissues. This study demonstrated for the first time that never-smoker patients with lung adenocarcinoma carry a specific and private oligogenic combination of germ-line mutations in cancer predisposing genes. These findings, if replicated with further studies, support the hypothesis of an oligogenic nature of common cancers. At the moment, we are performing exome sequencing in other pairs of siblings.

3452S

Epigallocatechin-3-gallate induces apoptosis and autophagy via up-regulation of TNF and GABARAPL2 gene expression in chronic myeloid leukemia cells. *B. Goker[1], C. Caliskan[1], Z. Mutlu[1], B. Erbaykent Tepedelen[2], M. Korkmaz[3], G. Saydam[4], C. Gunduz[1], C. Biray Avci[1].* 1) Department of Medical Biology, School of Medicine, Ege University, Izmir, Turkey; 2) Department of Molecular Biology and Genetics, Faculty Of Science And Letters, Avrasya University, Trabzon, Turkey; 3) Department of Medical Biology, School of Medicine, Celal Bayar University, Manisa; 4) Department of Hematology, School of Medicine, Ege University, Izmir, Turkey.

Chronic myeloid leukemia (CML) is a clonal hematopoietic stem cell disease. Philadelphia chromosome (Ph) is the result of a reciprocal (9;22) translocation is responsible for developing leukemic phenotype of CML. Epigallocathecin-3-gallate (EGCG) which has significant therapeutic effects on different human tumor types is a major flavonoid of green tea. Autophagy known as programmed cell death is one of the main response of cell to stimulus. Raise of autophagic activity in cancers is novel therapeutic target with using different active substance. In this study we aimed to determine cytotoxic and autophagic changes in K562 CML cell line treated with EGCG compared to untreated control group, and detect the expression changes ofautophagy related genes with the treatment of EGCG. The cytotoxic effect of EGCG on K562 cells was determined in time and dose dependent manner by using WST-1 analysis. Total RNA was isolated from K562 cells treated with EGCG and untreated cells as control group. Reverse transcription procedure was performed for cDNA synthesis. Apoptotic effect of EGCG was evaluated with ApoDIRECT and gene expressions were shown by RT-qPCR. IC50 (24th hours) dose of EGCG was determined as 50μM. EGCG induced apoptosis 10.9 fold compared to control cells. EGCG up-regulated TNF and GABARAPL2 genes 13.1775 and 9.5137 fold according to the control cells, respectively. Increase of TNF, most significiant cytokine is placed in apoptotic and autophagic pathways, and GABARAPL2, responsible for autophagic vacuol formation, expressions are associated with apoptotic and autophagic cell death in several human cancer cells. Our current findings manifested that EGCG treatment increases cancer cell death via up-regulation of genes are part of both apoptotic and autophagic signalling pathways. Therefore EGCG suggests a nontoxic herbal agent for chronic myeloid leukemia treatment.

3453M

Hypomorphic *CYP2C9* *2 and *3 alleles associate with improved non-small-cell lung cancer (NSCLC) prognosis. *L.N. Gordon[1], A. Pozzi[2,3], J.H. Capdevila[2,3], S.M. Williams[1,4].* 1) Dartmouth College, Department of Genetics Hanover, NH 03755; 2) Vanderbilt University, Department of Medicine Nashville, TN 37232, USA; 3) Vanderbilt University, Department of Nephrology Nashville, TN 37232, USA; 4) Dartmouth College, Institute for Quantitative Biomedical Sciences Hanover, NH 03755.

Lung cancer is the leading cause of cancer death in the United States, despite the higher incidences of other cancers. Given this disproportionate mortality, more effective treatments are needed. In mice knockout of the Cyp2c44 epoxygenase is associated with reduced angiogenesis and tumor growth. Interestingly, expression of the Cyp2c44 gene can be downregulated by treatment with PPARα ligands with a corresponding decrease in angiogenesis tumor growth, and distant metastasis. The human homologue of Cyp2c44, CYP2C9 epoxygenase, and its variants *CYP2C9*2* and *CYP2C9*3* have been linked to breast and colon cancer. However, whether *CYP2C9*, influences survival in non-small-cell lung cancer (NSCLC) patients and if it is linked to PPARα is not known, but could constitute an excellent drug target. We genotyped *CYP2C9* alleles *2* and *3* along with tag SNPs in *CYP2C9* in a sample of 398 Caucasian NSCLC patients. Possession of either *CYP2C*2* or *3* in female patients associated with increased survival (HR = 0.383, p-value = 0.035) adjusted for cancer staging, chemotherapy, and tumor resection. However, not only is this genetic association limited to females, but these alleles may interact with gender because the interaction between the hypomorphic alleles and gender borders the significance threshold (p-value = 0.100). Together, these results suggest not only that genetic variation in *CYP2C9* influences NSCLC survival, but modulation of this gene with PPARα ligands, such as clinically used fibrates, may be a viable clinical treatment.

3454T

Evaluation of miR-27a, miR-181a, and miR-570 Genetic Variants with Gallbladder Cancer Susceptibility and Prognosis in North Indian Population. *A. Gupta, B. Mittal, A. Yadav, A. Sharma, K.L. Sharma, N. Rastogi, S. Agrawal, A. Kumar.* Sanjai Gandhi Post Graduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India.

Purpose of study: microRNAs are small endogenously expressed short non-coding RNAs. They appear to be critical regulators of tumour biology as their aberrant expression is well characterized in cancer progression. Many miRNAs genetic variants have been found to play important role in many cancers but their contribution in gallbladder carcinoma has not been fully explored. So, in present study, we investigated the role of cancer-specific miR-27a, miR-181a, and miR-570 genetic variants with gallbladder cancer (GBC) susceptibility, therapeutic response and toxicities to chemo-radiotherapies. Methods: The present study was carried out in 515 GBC patients and 200 healthy controls in North Indian population. Among them, 126 patients receiving adjuvant or neoadjuvant therapies as per requirement were followed up for treatment outcome. Treatment response was recorded according to Response Evaluation Criteria in Solid Tumors (RECIST) criteria. Hematological and gastrointestinal toxicities profile were recorded as per CTCAE version 3.0 2006. Genotypes were determined by TaqMan probes. Statistical analysis was done by SPSS ver. 16. Results: Logistic regression analysis showed no significant association of miR-27a, miR-181a, and miR-570 polymorphism with GBC susceptibility ($p > 0.05$). On stratifying data on the basis of gall stone status, the [AG+GG] genotypes of miR-27a rs895819 (A>G) were significantly associated with increased risk of GBC in patients without stone (*p-value*=0.003; p_{corr} =0.006; OR=1.83; CI=1.23-2.72). The genetic risk by miR-27a, rs895819 (A>G) was modulated by tobacco consumption as the heterozygote's (AG) were at higher risk p-value=0.005, p_{corr} =0.01, OR=1.94; CI=1.22-3.08]. Univariate analysis revealed that AG+GG genotype of miR-27a and CT+TT genotype of miR-181a were significantly associated with non-responsiveness to chemo-radiotherapies (*p-value*=0.03 and 0.04 respectively) whereas miR-570 variants did not show any association. None of the studied polymorphism had any influence on haematological and gastrointestinal toxicities. Conclusion: We found significant association of miR-27a rs895819A>G with gallbladder cancer risk through gallstone independent pathway and tobacco usage. Our results are also suggest that variant allele of miR-27a (rs895819 A>G) and miR-181a (rs12537 C>T) may be associated with poor therapeutic response in GBC patients. However study needs to be validated in independent cohorts.

3455S

A case-control study of SNPs affecting microRNA binding sites in chronic myeloid leukemia. *H. Gutiérrez-Malacatt[1,2], M. Ayala-Sánchez[3], X. Aquino-Ortega[3], A. Martínez-Tovar[4], I. Olarte-Carrillo[4], J.A. Rebollar-Sánchez[2], U. Hernández-Guzmán[2], E.J. Córdova-Alarcón[2], L. Orozco[2].* 1) Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Ciudad Universitaria 3000, C.P. 04510, Coyoacán, Distrito Federal, México; 2) Laboratorio de Inmunogenómica y Enfermedades Metabólicas, Instituto Nacional de Medicina Genómica, Tlalpan 14610 México, D.F; 3) Servicio de Hematología, Hospital de Especialidades, CMN La Raza, IMSS; 4) Laboratorio de Biología Molecular, Servicio de Hematología. Hospital General de México, Cuauhtémoc, 06726 México, D.F.

Chronic myeloid leukemia (CML [MIM 608232]) is one of the most frequent hematopoietic malignancies in the elderly population. However, genetic factors associated with an increase risk to CML development are unknown. Recent studies have shown that SNPs affecting miRNA biogenesis or mRNA:miRNA interaction are important risk factors in the development of different types of cancer. Thus, we carried out a case-control study to test the association of SNPs located in the microRNA machinery genes AGO1 (rs636832), and GEMIN4 (rs2740348), as well as SNPs in the microRNA binding sites of the genes BRCA1 (rs799917) and KRAS (rs61764370) with CML susceptibility. We genotyped 828 Mexican-mestizo individuals (497 healthy subjects and 331 CML cases) using TaqMan probes. We found a significant association between the minor homocygote of the KRAS rs61764370 SNP (G/G) genotype and an increased risk for CML susceptibility (OR = 13.39, p = 0.015). After gender stratification, this association remains only in female individuals (OR = 13.41, p =0.016). None of the other studied SNPs showed a significant association with CML, even after gender stratification. In addition, the minor allele (C) of GEMIN4 rs2740348 SNP was significantly associated with an early age at CML diagnosis (OR= 2.14, p = 0.034) and with a high-risk Sokal score, (OR= 3.15, p = 0.01254). To the best of our knowledge this is the first study to show a significant association of the KRAS rs61764370 SNP and the GEMIN4 rs2740348 SNP with CML susceptibility and early age of CML diagnosis, respectively. To further determined the participation of these SNPs in CML susceptibility it is necessary to replicate our findings in different populations.

3456M

The role of Stem Cell Markers in Prostate Cancer Recurrence. *E. Guzel[1,2], OF. Karatas[1,3], MB. Duz[1], M. Ittmann[4,5], M. Solak[6], M. Ozen[1,4].* 1) Istanbul University, Cerrahpasa Medical School, Istanbul, Turkey; 2) Department of Electroneurophysiology Biruni University, Istanbul, Turkey; 3) Molecular Biology and Genetics Department, Erzurum Technical University, Erzurum, Turkey; 4) Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX, 77030, USA; 5) Baylor College of Medicine, Michael E. DeBakey VAMC, Houston, TX, 77030, USA; 6) Afyon Kocatepe University Medical School Department of Medical Genetics.

Prostate cancer (PCa) is one of the most common tumor types related to mortality in males in the developed countries. Studies have demonstrated that therapeutic tools mostly ineffective to give positive outcome especially for PCa. Cancer stem cells are composed of a small cell population, which are supposed to have roles in tumorigenesis, metastasis, and tumor recurrence after chemo-radiotherapy. The aim of this study is to explore the differential expressions of stem cell markers in recurrent PCa and non-recurrent PCa tumors as well as in adjacent normal prostate tissues. Here we compared the expression of important stemness regulators like SOX2, OCT4, KLF4, and ABCG2 genes in recurrent, non-recurrent PCa and adjacent normal tissue samples using quantitative real-time polymerase chain reaction. Our results demonstrated that SOX2 and OCT4 are strongly over-expressed in PCa samples. Recurrent PCa samples are markedly positive for stem cell markers SOX2, OCT4 and KLF4. Furthermore, non-recurrent PCa samples presented low levels of ABCG2, a multidrug resistance protein, compared to both normal and recurrent samples, which might be associated with chemo-sensitivity. Enhanced expression of ABCG2 and stem cell markers in the recurrent PCa tissues postulates the suggestion that enrichment for cells with stem cell characteristics in these tissues might be critical for enhancement of chemoresistance and recurrence of cancer.

3457T

***KEAP1* genetic polymorphisms associate with breast cancer risk and survival outcomes.** *J.M. Hartikainen[1], M. Tengström[2, 3], R. Winqwist[4, 5], A. Jukkola-Vuorinen[6], K. Pylkäs[4, 5], V.-M. Kosma[1, 7], Y. Soini[1, 7], A. Mannermaa[1, 7].* 1) School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, and Cancer Center of Eastern Finland, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Kuopio, Finland; 2) School of Medicine, Institute of Clinical Medicine, Oncology, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Finland; 3) Cancer Center, Kuopio University Hospital, P. O. Box 1777, FI-70211 Kuopio, Finland; 4) Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, P. O. Box 5000, FI-90014 University of Oulu, Oulu, Finland; 5) Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre NordLab, Oulu University Hospital, Oulu, Finland; 6) Department of Oncology, University of Oulu, Oulu University Hospital, P.O. Box 5000, FI-90014 University of Oulu, Finland; 7) Imaging Center, Clinical Pathology, Kuopio University Hospital, P. O. Box 1777, FI-70211 Kuopio, Finland.

Defects in response to oxidative stress may increase cancer susceptibility. In tumors the function of the rescue mechanisms may cause chemo- and radioresistance and affect the outcome of patients. We previously showed that genetic variation in nuclear factor erythroid 2-related factor 2 (NRF2) gene *NFE2L2* associates with breast cancer risk and prognosis. Here we studied further this pathway by investigating the Kelch-like ECH-associated protein 1 (KEAP1). Five tagging single nucleotide polymorphisms in *KEAP1* gene were genotyped in 996 breast cancer cases and 880 healthy controls from two Finnish case-control sets. KEAP1 protein expression was studied in 373 invasive breast cancer tumors by immunohistochemistry. Statistical significance of the associations between genotypes, protein expression, clinicopathological variables and survival were assessed. The genotype *TT* of rs34197572 associated with increased risk of breast cancer (P_{Trend}= 0.015). The minor allele *A* of rs11085735 associated with lower KEAP1 protein expression (*P*=0.040) and high extent nuclear NRF2 protein expression (*P*=0.009). It also associated with worse survival in all invasive cases (*P*=0.023). When treatment data were included rs11085735 associated with recurrence-free survival (RFS) (*P*=0.020) and breast cancer-specific survival (BCSS) (*P*=0.016), and rs34197572 with overall survival (OS) (*P*=0.045). rs34197572 and rs8113472 associated with OS also among radiation therapy-treated cases (*P*=0.018 and 0.025, respectively). rs9676881 and rs1048290 associated with RFS among all invasive cases (*P*=0.024 and 0.020, respectively), and among ER positive tamoxifen treated cases (*P*= 0.018 and 0.015, respectively). These associations possibly indicate the effect of the SNPs in response to oxidative stress by cancer treatment, thus providing further support to the involvement of the NRF2-KEAP1 pathway in breast cancer susceptibility and patient outcome.

3458S

Rare and Common Variants Contribute to Lung Cancer Survival in African Americans. *C.C. Iverson[1], W.S. Bush[1], D.C. Crawford[1], H.H. Dilks[2], J. Long[3], W.J. Blot[3,4], E.L. Grogan[5,6], M.C. Aldrich[1,3,5].* 1) Center for Human Genetics Research, Vanderbilt University Medical School, Nashville, TN; 2) VANTAGE, Vanderbilt University, Nashville, TN; 3) Division of Epidemiology, Vanderbilt University Medical School, Nashville, TN; 4) International Epidemiology Institute, Rockville, MD; 5) Department of Thoracic Surgery, Vanderbilt University Medical School, Nashville, TN; 6) Tennessee Valley Health System Veterans Affairs, Nashville, TN.

Lung cancer is the leading cause of cancer-related mortality in the U.S. Survival rates differ by race, with blacks experiencing poorer survival than whites, yet few studies have focused on blacks. Germline genetic variation may influence overall lung cancer survival. A total of 305 incident non-small cell lung cancer African American cases were identified from the prospective Southern Community Cohort Study through linkage with 12 state cancer registries. Vital status was determined by linking with the National Death Index or Social Security Administration. After 8.6 years of follow-up, 87% of lung cancer cases were deceased. We performed genotyping using the Illumina HumanExome BeadChip. After standard quality control, 301 individuals (60% male) and 274,438 variants remained for analysis. We identified variants previously associated with lung cancer survival from the NHGRI GWAS catalog on the ExomeChip array. For each of these SNPs, we ran a Cox proportional hazards model adjusted for age, sex, percent African ancestry (estimated from ancestry informative markers), stage at diagnosis, and treatment. We found the C allele at rs1878022, in the chemokine receptor-like 1 (*CMKLR1*) gene, was associated with reduced mortality [hazard ratio (HR): 0.72, 95% confidence interval (CI): 0.54-0.97, p=0.03]. The improved survival is in contrast to a prior lung cancer survival GWAS of similar sample size conducted in whites, suggesting ethnic-specific associations. We then sought to identify rare variants in protein coding regions associated with lung cancer survival. We identified variants with a MAF < 5% (n=114,646) and mapped them to 9,725 genes. We used the sequence kernel association test (SKAT) and burden test, adjusting for age, sex, and African ancestry, to examine associations between rare variants and lung cancer survival. The strongest association was with the *MICAL-L2* gene (HR =1.84, 95% CI: 1.42-2.37, $p=1.04\times10^{-6}$) that interacts with actinin-4, which in turn harbors variants associated with poor lung cancer survival. A significant association was also found for the vascular epidermal growth factor B (*VEGFB*) gene ($p=4.0\times10^{-6}$), which may play a role in tumor survival. We identified 3 genes associated with overall survival, which if validated could suggest potential new targets for lung cancer treatments. Fine-mapping in larger populations and functional studies are required to understand the role of these variants in lung cancer survival.

3459M

Cytogenetic abnormalities of 50 AML patients by FISH detection and conventional karyotype analysis. *E. Karaca, A. Aykut, B. Durmaz, A. Durmaz, I.M. Tekin, O. Cogulu, H. Akin.* Ege University Faculty of Medicine, Department of Medical Genetics, Izmir, Turkey.

Cytogenetic analyses in acute myeloid leukemia (AML) have come to light a great number of chromosome abnormalities. The cytogenetic analysis is an important laboratory tool for diagnosis, prognosis, in clinical decision-making and in follow-up for pediatric and adult patients with AML. Fluorescence in situ hybridization (FISH) is increasingly being used in the cytogenetic diagnosis of AML. In this study the detectable rate of cytogenetic abnormalities including t(8;21)(q22;q22), t(15;17)(q24;q21), t(9;22)(q34;q11), inv(16)(p13q22) and MLL rearrangements by FISH was evaluated. The aim was to investigate the ratio of cytogenetic abnormalities and compare it with the literature. FISH analysis for mentioned cytogenetic abnormalities were performed in 50 bone marrow samples obtained from adult patients with AML diagnosed according to the WHO criteria. The results showed that the cytogenetic abnormalities were identified in 12 (24 %) of the patients. The most frequent aberration rate was 10 % for inv(16)(p13q22). The second most frequent aberration was 6 % for t(9;22)(q34;q11). t(8;21)(q22;q22), t(15;17)(q24;q21), and MLL rearrangements detection rates were 4 %, 2 %, 4 % respectively. The repetitive analyses by the aim of follow up and cytogenetic results were also evaluated. It was noted a substantially lack of an algorithm at diagnosis and follow up request for AML patients. A more reliable approach like wider FISH panels are necessary in order to detect more anomalies and using an algorithm to reach effective diagnose and follow up in AML patients.

3460T

DNA methylation profiling reveals novel diagnostic biomarkers in renal cell carcinoma. *B.N. Lasseigne[1, 2], T.C. Burwell[1], M.A. Patil[3], D.M. Absher[1], J.D. Brooks[3], R.M. Myers[1].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, Al; 2) Department of Biological Sciences, University of Alabama in Huntsville, Huntsville, AL; 3) Department of Urology, Stanford University, Stanford, CA.

Renal cell carcinoma (RCC) is the 10th most commonly diagnosed cancer in the United States. While it is usually lethal when metastatic, RCC is successfully treated with surgery when tumors are confined to the kidney and have low tumor volume. Because most early stage renal tumors do not result in symptoms, there is a strong need for biomarkers that can be used to detect the presence of the cancer as well as to monitor patients during and after therapy. We examined genome-wide DNA methylation alterations in renal cell carcinomas of diverse histologies and benign adjacent kidney tissues from 96 patients. We observed widespread methylation differences between tumors and benign adjacent tissues, particularly in immune, G-protein coupled receptor, and metabolism-related genes. Additionally, we identified a single panel of DNA methylation biomarkers that reliably distinguishes tumor from benign adjacent tissue in all of the most common kidney cancer histologic subtypes (area under receiver operating characteristic curve, 0.991), and a second panel does the same specifically for clear cell renal cell carcinoma tumors (area under receiver operating characteristic curve, 0.990). This set of biomarkers were validated independently with excellent performance characteristics in more than 1,000 tissues in The Cancer Genome Atlas clear cell, papillary, and chromophobe renal cell carcinoma datasets (area under receiver operating characteristic curves, >0.97). These DNA methylation profiles provide insights into the etiology of renal cell carcinoma and, most importantly, demonstrate clinically applicable biomarkers for use in early detection of kidney cancer.

3461S

Association of LEP rs7799039 (G-2548A) polymorphism with obesity in breast cancer patients. *A. Méndez'Hernández[1, 2], JA. Espinosa[1], R. Pérez[3], MP. Gallegos[2].* 1) Laboratorio de Investigación, Facultad de Medicina, Universidad Juárez del Esta, Durango, Mexico; 2) Laboratorio de Genética Molecular, División de Medicina Molecular, CIBO, IMSS, Guadalajara. Jalisco; 3) Facultad de Ciencias Químicas, , Universidad Juárez del Estado de Durango, Gómez Palacio, Dgo.

Background: *LEP* rs7799039 polymorphism (*G-2548A*) has been correlated with changes in leptin levels, the degree of obesity and is involved in the evolution of breast cancer. The *LEP -2548AA* genotype was associated with increased leptin concentrations, increased expression of leptin mRNA in adipose tissue, and more high body mass index, thereby increasing the risk for overweight and obesity and evolution breast cancer. Our goal is to evaluate the association of polymorphism rs7799039 *LEP* (*G-2548A*) in obese patients with breast cancer. Methods: DNA genimic of 245 samples (UMAE Hospital Gyneco-Obstetrician, CMNO, IMSS) were included in the study. The *LEP* rs7799039 (*G-2548A*) genotyping was determinate by Taq-Man® probe and allelic discriminated was determinate by StepOne Software v.2., of Applied Biosystems. The association was determinate by odds ratio (OR). Results: The genotype *LEP- 2548AA* was associated with overweight and obesity [OR = 2.1 (95% CI 0.96 - 4.71) , p = 0.05] in patients with breast cancer, while the *LEP- 2548GG* genotype was found as a protective factor to overweight and obesity [OR = 0.47 (95% CI 0.20 . 0.82) , p = 0.011]. Conclusion: *LEP* rs7799039 polymorphism (*G-2548A*) may be an indicator of risk for overweight and obesity and explain the possible mechanism of obesity in the progression of breast cancer.

3462M

Changes in Colorectal Carcinoma Genomes under Anti- EGFR Therapy Identified by Whole-Genome Plasma DNA Sequencing. *S. Mohan[1], E. Heitzer[1], P. Ulz[1], I. Lafer[1], S. Lax[2], M. Auer[1], M. Pichler[3], A. Gerger[3], F. Eisner[3], G. Hoefler[4], T. Bauernhofer[3], J.B Geigl[1], M.R Speicher[1].* 1) Institute of Human Genetics, Medical University of Graz, Graz, Austria; 2) Department of Pathology, General Hospital Graz West, Graz, Austria; 3) Division of Oncology, Medical University of Graz, Graz, Austria; 4) Institute of Pathology, Medical University of Graz, Graz, Austria.

Monoclonal antibodies targeting the Epidermal Growth Factor Receptor (EGFR), such as cetuximab and panitumumab, have evolved to important therapeutic options in metastatic colorectal cancer (CRC). However, almost all patients with clinical response to anti-EGFR therapies show disease progression within a few months and little is known about mechanism and timing of resistance evolution. Here we analyzed plasma DNA from 10 patients treated with anti-EGFR therapy by whole genome sequencing (plasma-Seq) and ultra-sensitive deep sequencing of genes associated with resistance to anti-EGFR treatment such as *KRAS*, *BRAF*, *PIK3CA*, and *EGFR*. Surprisingly we observed that the development of resistance to anti-EGFR therapies was associated with acquired gains of *KRAS* in 4 patients (40%), which occurred either as novel focal amplifications (n=3) or as high level polysomy of 12p (n=1). In addition, we observed focal amplifications of other genes recently shown to be involved in acquired resistance to anti-EGFR therapies, such as *MET* (n=2) and *ERBB2* (n=1). Overrepresentation of the *EGFR* gene was associated with a good initial anti-EGFR efficacy. Overall, we identified predictive biomarkers associated with anti-EGFR efficacy in 7 patients (70%), which correlated well with treatment response. In contrast, ultra-sensitive deep sequencing of *KRAS*, *BRAF*, *PIK3CA*, and *EGFR* did not reveal the occurrence of novel, acquired mutations. Thus, plasma-Seq enables the identification of novel mutant clones and may therefore facilitate early adjustments of therapies that may delay or prevent disease progression.

3463T

Zoledronic acid treatment up-regulates miR-15a via targeting antiapoptotic BCL2 gene expression in chronic myeloid leukemia. *Z. Mutlu[1], C. Caliskan[1], B. Goker[1], C. Kayabasi[1], B. Tezcanli Kaymaz[1], G. Saydam[2], C. Gunduz[1], C. Biray Avci[1].* 1) Ege University Medical School Department of Medical Biology, Izmir, Turkey; 2) Ege University Medical School Department of Hematology, Izmir, Turkey.

Chronic myeloid leukemia (CML) is associated with a characteristic chromosomal translocation called the Philadelphia chromosome (Ph). This genetic alteration for the diagnosis of CML causes the formation of the BCR-ABL1 fusion gene which produces actively tyrosine kinase. Apoptosis is a form of programmed cell death and autophagy is a lysosomal degradation pathway essential for homeostasis that contributes to cell death. Zoledronic acid (ZA), a nitrogen-containing bisphosphonate, is a potential inhibitor of osteoclast-mediated bone resorption and also remarkable for anticancer activities. Autophagic effect of ZA in leukemia is still unknown. miRNAs cause regulation of autophagic and apoptotic pathways in different cancers. In this study, we aimed to evaluate the cytotoxic, autophagic and apoptotic effects of ZA and to examine expression levels of miRNAs treatment with ZA in K562 CML cells. Also we investigate the roles of miRNA target genes during leukemogenesis. K562 cells were treated with dose range of 2 μM to 80 μM ZA during 24, 48, 72 hours and cytotoxicity was performed by WST-1 analysis. Apoptosis assays were evaluated by AnnexinV and by ApoDIRECT In Situ DNA Fragmentation Assay Kits. The RT-qPCR is used for miRNAs and genes expressions analyses. miRNAs and genes expression levels were evaluated by using miScript miRNA PCR Array and RT2 Profiler PCR Array, respectively. According to results, ZA upregulated expression level of miR-15a 9.88 fold and downregulated expression level of BCL2 gene 3,16 fold which is a miR-15a target gene, according to control cells. The results provide that ZA can be played a role as a kinase inhibitory in CML. Because, apoptomirs, such a miR-15a were down-regulated and this downregulation of miR-15a was dependent on BCR-ABL kinase activity in ZA-untreated CML cells. Consequently, ZA upregulated miR-15a expression and miR-15a directly targeted BCL2 gene and induced apoptosis or autophagy progressions in CML. These data indicated that downregulation of miR-15a could be a pioneering target in CML treatment. With these results further experiments are required for the usage of ZA in CML.

3464S

Towards a national implementation of DNA-based personalized cancer treatment in the Netherlands. *I.J. Nijman on behalf of the Netherlands Center for Personalized Cancer Treatment.* Center for Personalized Cancer Treatment.

In the past 30 years, cancer treatment has improved enormously, partly due to the development and implementation of targeted therapies. These therapies are designed to tackle specific characteristics of a cancer cell, such as growth factor receptors. Unfortunately, the presence or absence of such characteristics only paints a partial picture of the tumor's responsiveness. Instead, treatment outcome can be better predicted by a combination of genetic mutations. As each tumor has its own genetic characteristics, the appropriate treatment-decision differs per patient and is therefore called "personalized treatment". This approach can improve treatment outcome for patients with metastatic cancer, decrease the number of patients unnecessarily exposed to toxic agents, and reduce costs. Being able to classify patient populations into groups with different likelihoods to benefit from a particular treatment is key in this process. The Center for Personalized Cancer Treatment brings this concept to the patient today by forging collaborations between all academic cancer centers in the Netherlands. The CPCT focusses on two DNA analysis tracks. First, rapid screening for actionable mutations and amplifications to select the appropriate standard targeted treatment or to allocate patient into investigator and pharma-driven Phase1 trials. At the moment over 1200 patients have been analyzed through this route. Secondly, cohorts of patients with evaluable response to targeted treatments are sequenced and analyzed in a system biology perspective to identify predictive biomarkers. Detailed follow up data for over 700 patients is now being collected in an electronic clinical response form (eCRF) and integration with genetic, pathology and pharmacological is ongoing. Combined with public datasets, this integrated dataset provides a huge and growing national resource that can be used for biomarker- or target evaluation, systems biology analyses, as well as for improving care for individual patients.

3465M

Integration of microarray meta-analysis with RNASeq and genome-wide genetic data to identify variants associated with endometrial cancer histological subtype. *T.A. O'Mara, A.B. Spurdle.* Genetics and Computational Biology Dept, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.

Non-endometrioid endometrial cancer comprises approximately 20% of new endometrial cancer diagnoses, but is responsible for an estimated 50% of deaths from this disease. Microarrays have been successfully used to measure mRNA expression in biological samples to identify differentially expressed genes. However, it is recognized that results from individual microarray studies are often not reproducible. Public availability of gene expression microarray data has facilitated research assessing the increased reliability of combined datasets. Furthermore, the availability of genome-wide scale "-omics" data on the same sample sets, such as that provided by The Cancer Genome Atlas (TCGA), allows for a powerful, integrative approach to identify genes and genetic variants important in cancer initiation and/or progression. We have used patient derived microarray data from multiple datasets to identify differences between endometrioid (EEC) and non-endometrioid (NEEC) histological subtypes of endometrial cancer, and then validated these differences using an independent endometrial cancer set with RNASeq data from the TCGA. A literature review and repository search conducted in April 2014 identified 15 endometrial cancer microarray studies. Raw microarray data was accessed from publication supplementary data, the NCBI Gene Expression Omnibus (GEO) and the TCGA data portal, or otherwise requested by directly contacting authors. Overall raw microarray data was accessed for 9 of 15 studies, with a maximum of 79 EEC and 12 NEEC cases in one study. Following QC and meta-analysis using the MetaOmics package in R, 1,910 genes from 6 studies were identified by Fisher's test as displaying significant differential expression between the two subtypes (FDR < 5%). The expression of these genes was then analyzed using normalized RNASeq data for 328 EEC and 97 NEEC non-overlapping samples from the TCGA. For the 1,186 genes captured by the RNASeq analysis, moderated t-test identified differences in expression for 912 genes at $P < 0.05$ (77% of genes analysed), of which 261 were at $P < 1\times10^{-19}$. These findings demonstrate a strong relationship between microarray meta-analysis and RNASeq results, and identify candidate genes differentially expressed in aggressive endometrial cancer histological subtypes. Future analysis will investigate genetic variation associated with differential gene expression in the TCGA dataset, to prioritize genetic variants for downstream prognostic studies.

3466T

The type II transmembrane serine proteases hepsin and TMPRSS3 are associated with breast cancer survival. *M. Pelkonen[1, 2, 3], K. Luostari[1, 2, 3], H. Ahonen[1, 2, 3], B. Berdel[1, 2, 3], V. Kataja[4, 5], Y. Soini[1, 2, 3], V-M. Kosma[1, 2, 3], A. Mannermaa[1, 2, 3].* 1) Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Finland; 2) Biocenter Kuopio and Cancer Center of Eastern Finland, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Finland; 3) Imaging Center, Clinical Pathology, Kuopio University Hospital, P.O. Box 1777, FI-70211 Kuopio, Finland; 4) Institute of Clinical Medicine, Oncology, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Finland; 5) Cancer Center, Kuopio University Hospital, P.O. Box 1777, FI-70211 Kuopio, Finland.

Background: Hepsin, (also called TMPRSS1) and TMPRSS3 are type II transmembrane serine proteases (TTSPs) that are involved in cancer progression. TTSPs can remodel extracellular matrix (ECM) and, when dysregulated, promote tumor progression and metastasis by inducing defects in basement membrane and ECM molecules. This study investigated whether the gene and protein expression levels of these TTSPs were associated with breast cancer characteristics or survival. Methods: Immunohistochemical staining was used to evaluate hepsin levels in 372 breast cancer samples and TMPRSS3 levels in 373 samples. *TMPRSS1* mRNA expression was determined in 125 invasive and 16 benign breast tumor samples, and *TMPRSS3* mRNA expression was determined in 167 invasive and 23 benign breast tumor samples. The gene and protein expression levels were analyzed for associations with breast cancer-specific survival and clinicopathological parameters. Results: Low *TMPRSS1* and *TMPRSS3* mRNA expression levels were independent prognostic factors for poor breast cancer survival during the 20-year follow-up (*TMPRSS1*, P = 0.023; HR, 2.065; 95% CI, 1.106-3.856; *TMPRSS3*, P = 0.013; HR, 2.106; 95% CI, 1.167-3.800). Low expression of the two genes at the mRNA and protein levels associated with poorer survival compared to high levels (log rank P-values 0.015-0.042). Grade III tumors, large tumor size, and metastasis were associated with low mRNA and protein expression levels. Conclusions: The results suggest that the TTSPs hepsin and TMPRSS3 may have similar biological functions in the molecular pathology of breast cancer and that they represent potential novel therapeutic targets. Low mRNA and protein expression levels of the studied TTSPs were prognostic markers of poor survival in breast cancer.

3467S

Genetic variants in the gene *ARID5B* associated with susceptibility to childhood acute lymphoblastic leukemia. *A. Reyes-León[1], C. Salas-Labadía[1], J. García-Cruz[1], M. Zapata-Tarrés[1], M. Ramírez-Martínez[1], R. Paredes-Aguilera[1], R. Rivera-Luna[1], S. Jiménez-Morales[2], L. Orozco-Orozco[2], M. Dean[3], P. Pérez-Vera[1].* 1) Instituto Nacional de Pediatría, Mexico city, Mexico; 2) Instituto Nacional de Medicina Genómica, Mexico city, Mexico; 3) National Cancer Institute, Frederick, Maryland, USA.

In Mexico the acute lymphoblastic leukemia (ALL) is the most common cancer in pediatric population. According to the National Popular Medical Insurance Program, in our country the incidence is 60 cases per million per year. Some individual genetic variations might contribute to the increased incidence, and simultaneously influence susceptibility to the disease development. Recently, some authors have reported single nucleotides polymorphisms (SNPs) of *ARID5B* gene that could be associated with susceptibility to childhood ALL in Latin American population; however these observations are not representative of the real situation in Mexico. The purpose of this study was to determine the presence of 2 SNPs of *ARID5B* in ALL children for establishing their possible association with susceptibility to develop the disease. The study population included 112 children with ALL and 114 controls. Informed consent was obtained before taking saliva samples. The DNA was extracted, and genotyping analysis was performed using TaqMan probes for the SNPs rs10821936 and rs10994982 of *ARID5B*. The allele and genotype frequencies of ALL cases and controls were calculated. Odds Ratio (OR) with 95% confidence intervals [95% CI] were performed using FINETTI software. *P* values less than 0.05 was considered statistically significant. For the SNP rs10821936, 112 ALL cases and 114 controls were analyzed and for SNP rs10994982 only 52 ALL cases and 99 controls were studied. The frequency of the risk allele C of the SNP rs10821936 was significantly higher ($p<0.0001$) in cases (89.3%) than in controls (70.2%); the behavior of the frequencies for the allele A of the SNP rs10994982 was similar, the frequency in cases was 92.3% and in controls 78.8% ($p=0.0001$). The risk for developing ALL was increased in the presence of the homozygous CC ($p=2.8e-6$, OR 6.5, CI) and AA ($p=0.0009$ OR 6.6, CI) genotypes compared with the heterozygous CT ($p=0.02$, OR 2.3 CI) and AG ($p=0.3219$, OR 1.8 CI). The frequency of the risk alleles C and A in our patients was higher than reported in other populations. The frequency of these alleles in the control group is similar to that observed in healthy indigenous population from Mexico and Guatemala. These results suggest that the genetic component of our patients could be influencing the susceptibility to develop ALL and could also explain its high incidence in Mexico. This work was supported by Fondos del Presupuesto Federal para la Investigación 2012 (085/2012).

3468M

Cell cycle progression gene expression score and prostate cancer outcomes in a population-based cohort. *R. Rubicz[1], S. Zhao[1], C. April[2], J.L. Wright[1,3], S. Kolb[1], I. Coleman[4], D.W. Lin[1,3], P.S. Nelson[4,5,6], E.A. Ostrander[7], Z. Feng[8], J.-B. Fan[2], J.L. Stanford[1].* 1) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Illumina, Inc., San Diego, CA; 3) Department of Urology, University of Washington School of Medicine, Seattle, WA; 4) Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA; 5) Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA; 6) Department of Medicine, University of Washington, Seattle, WA; 7) Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 8) Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX.

Prostate cancer (PCa) is clinically heterogeneous, making it difficult to accurately predict which patients will have an indolent vs. aggressive disease course. Therefore markers that can help distinguish aggressive tumors in high-risk patients are urgently needed. In 2012 Cuzick et al. reported a cell cycle progression (CCP) score (based on expression of 31 genes in tumor tissue) as a predictor of PCa death in a conservatively managed patient cohort. To further evaluate the CCP score we tested its performance in a population-based cohort of 383 men diagnosed with clinically localized PC in 1993-1996 or 2002-2005 who underwent radical prostatectomy (RP). Recurrence information was obtained from participant surveys, medical records, and the SEER cancer registry. Over a follow-up period of 12 years, 278 men (73%) had no evidence of recurrence and 105 (27%) had recurrence events, including 27 (7%) patients who developed metastasis or died of PCa (lethal PCa). Tumor RNA was isolated from FFPE blocks of RP samples and gene expression data were obtained using the Illumina Whole-Genome DASL HT Assay. CCP score was based on average expression levels of 30 of the 31 progression genes (c18orf24 was excluded due to low quality). Based on Cox proportional hazards models, the hazard ratio (HR) associated with a change from the 25th to 75th percentile of the CCP score (range 7.97-9.96) was 1.25 (95% CI 0.96-1.62; P = 0.10) for overall recurrence, and for the subset of patients with lethal PCa it was 2.01 (95% CI 1.26-3.21; P = 0.004), adjusted for clinical variables (ie, age at diagnosis, Gleason score, PSA level, and pathologic stage). Adding the proliferation marker *Ki67* to the model did not change the results. CCP score was only weakly correlated (by Pearson correlation) with Gleason score (0.13) and PSA (0.16). Secondary analyses showed similar associations when patients were stratified by Gleason score, *TMPRSS2-ERG* fusion status, or family history of PCa. The ROC curve for CCP score (AUC = 0.740) did not perform as well as clinical variables (AUC = 0.803) for predicting lethal PCa, and the ROC curve combining CCP score with clinical variables was only slightly improved (AUC = 0.831). These results confirm a significant association between CCP score and risk of lethal PCa. However, the CCP score did not substantially improve the AUC and had less discriminatory accuracy than clinical variables alone.

3469T

Gene expression profile of human telomere and telomerase complex in intestinal-type gastric cancer. *L.C. SANTOS[1], F. Wisnieski[1], M.F. Leal[1], D.Q. Calcagno[2], C.O. Gigek[1], E.S. Chen[1], S. Demachki[2], P.P. Assumpção[2], R. Artigiani[3], L.G. Lourenço[4], R.R. Burbano[5], M.A.C. Smith[1].* 1) Genetics Division, Department of Morphology and Genetic, Federal University of São Paulo, São Paulo, Brazil; 2) Nucleu of Research in Oncology, João de Barros Barreto University Hospital, Federal University of Pará, Belém, Brazil; 3) Department of Pathology, Federal University of São Paulo, São Paulo, Brazil;; 4) Department of Surgical Gastroenterology, Federal University of São Paulo, São Paulo, Brazil; 5) Human Cytogenetics Laboratory, Institute of Biological Sciences, Federal University of Pará, Belém, Brazil.

Gastric cancer is the fourth most prevalent and the second cause of cancer death worldwide. Changes in gene expression of telomere and telomerase complex have been observed in carcinogenic processes. The investigation of these genetic mechanisms may help to determine more accurate diagnosis, prognosis and the establishment of a therapeutic approach. In the present study, we aimed to identify differentially expressed genes of telomere and telomerase complex in gastric tumors compared to matched noncancerous gastric samples. The expression of 84 key genes involved in telomere replication and maintenance was evaluated using the Human Telomeres & Telomerase RT² Profiler™ PCR Array (SABiosciences, PAHS-010). We analyzed 22 paired intestinal-type gastric cancer and adjacent noncancerous gastric tissues from individuals of Northern Brazil. We observed that ACD, BCL2, MUS81, RAD17, RAPGEF1, SIRT2 and SIRT6 expression were downregulated in neoplasic samples compared to matched noncancerous (-1.9, -2.3, -1.8, -1.6, -1.8, -1.5 and -1.6 - fold regulation; and p = 0.004, p = 0.01, p = 0.04, p = 0.008, p = 0.01, p = 0.04, p = 0.04; respectively). And HSP90AA1 and NBN were upregulated (1.8 and 1.8 - fold regulation; and p = 0.008 and p = 0.02; respectively). ACD gene encodes a protein that is involved in telomere maintenance. This protein is one of six core proteins of shelterin telomeric complex. RAPGEF1 mediate binding events that control the activity of shelterin complex. MUS81 and RAD17 are required for efficient DNA repair of telomere-associated complexes. The BCL2 gene encodes a protein that has a role in the control of apoptosis and inhibition of the telomerase action. SIRT2 and SIRT6 have been implicated in influencing a wide range of cellular processes like aging, DNA repair, telomere maintenence and energetic metabolism. HSP90AA1 has anti-apoptotic function and NBN promotes repair of damaged DNA. These differentially expressed genes highlight the importance of telomere and telomerase complex in intestinal-type gastric carcinogenesis. These results contribute to the understanding of gastric cancer biology with possible future clinical implications to be tested in a subsequent study in a larger number of gastric cancer patients, and to be associated to sex, age, tumor location, H. pylori, tumor extension, limph node metastasis and distant metastasis.

3470S

Study of expression of AKAP4, RGS22, SPAG9 and NY-ESO-1 genes as probable diagnosis and prognosis biomarkers in colorectal cancer. *A. Tavakoli Koudehi[1], F. Mahjoubi[1], B. Mahjoubi[2], R. Mirzaee[2].* 1) National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran; 2) Hazrat Rasoul Hospital, Iran University of Medical Sciences and Health Care Services, Tehran, Iran.

Background: Colo-rectal cancer (CRC) is the most common gastrointestinal cancer and the second leading cause of death in women after breast cancer and the third leading cause of death in men after lung and prostate carcinomas in the world. Because of the slow progression of CRC and the ability of treatment in primary levels (before lymph-vascular invasion), there is an urgent need to identify non-invasive tumor bio-markers (biological markers) that can prognostic the CRC in primary levels and subsequently decrease the prevalence and mortality of it. Cancer-Testis Antigens (CTAs) are a group of tumor-associated proteins which typically are expressed in normal reproductive cells of men, but their expression in normal somatic cells is off. Several reports have been identified the CTAs expression in various human tumors, such as, ovarian, endo-metrial, cervical, esophageal, and breast cancers. CTAs due to their limited expression pattern, are as promising targets for cancer diagnosis and immuno-therapy. Method: We study the expression of AKAP4, RGS22, SPAG9 and CTAG1B genes from the CTAs family in both tumor and normal tissues of 80 Iranian CRC patients by RT-PCR with the aim of comparing the genes expression and finding a biomarker for early detection and anticipated progress to CRC. According to studies, AKAP4, RGS22, and NY-ESO-1 gene expressions in colorectal cancer tissues has not been investigated so far. The SPAG9 gene expression by RT-PCR was first investigated in Iranian patients. Result: Results are pending.

3471M

Gene polymorphisms as risk factors for Cervical Cancer in a South Indian Population. *P. Upendram[1], S. Poornima[1], G. Apoorva[1], K. Jayanthi[1], V. Kiran Kumar[1], A. Shah[2], Q. Hasan[1].* 1) Department of Genetics & Molecular Medicine, Kamineni Hospitals, L.B.Nagar, Hyderabad, Andhra Pradesh, India; 2) Department of Oncology, Kamineni Hospitals, L.B.Nagar, Hyderabad, Andhra Pradesh, India.

Carcinoma of the uterine cervix is the most frequent gynecological malignancy affecting women in developing countries despite being a potentially preventable disease. Globally, it is the second most frequent cancer affecting women with a steady rise in incidence among younger women. It is evident from current literature that specific gene polymorphisms in hormone receptor genes and the cell cycle regulating genes along with ineffective host immune response may enhance the risk for gynecological malignancies. The aim of this case control study from South India was to evaluate four gene polymorphisms (i) T/C SNP of estrogen receptor (ER) alpha gene recognized by PvuII enzyme (rs2234693), (ii) 306bp Alu insertion in the progesterone receptor (PGR) gene (rs1042838), (iii) G870A SNP of CCND1 gene (rs9344) and (iv) C850T SNP of TNF alpha gene (rs909253) in cervical cancer patients. DNA was isolated from a total of 200 women comprising 100 cases with cervical cancer and 100 age-matched healthy controls using the salting out technique. Polymorphisms were evaluated by Polymerase Chain Reaction followed by restriction enzyme digestion / gel electrophoresis. Our results showed a statistically significant association of the C allele of ER gene (OR= 5.8072; 95% CI=3.7671 to 8.9521; p<0.0001) with cervical cancer. While, the T2 allele of PR gene did not show any significant association with cervical cancer (OR=1.3700; 95% CI=0.7193 to 2.6093; p= 0.33). The A allele CCND1 gene was associated with cervical cancer (OR= 1.917; 95% CI=1.286 to 2.857; p<0.0014) and the T allele of TNF alpha gene was not significantly associated with this cancer (OR= 1.413; 95% CI=0.95 to 2.102; p<0.087), however the heterozygous CT genotype was protective. In Multifactor Dimensionality Reduction (MDR) analysis, the ER and CCND1 gene polymorphisms showed an interaction with cervical cancer (p<0.0001). Our results suggest that the ER PvuII polymorphism and CCND1 G870A polymorphism can be used as biomarkers in identifying women with an increased risk of getting cervical cancer. A larger study in different ethnic groups is warranted for establishing them as biomarkers for cervical cancer.

3472T

Association of common Cancer stem cells (CSCs) genes variants with gallbladder cancer susceptibility and prognosis in North Indian population. *A. Yadav[1], B. Mittal[2], A. Gupta[3], K.L. Sharma[4], V. Kumar[5], A. Kumar[6].* 1) Medical genetics, Sanjay Gandhi Post Graduate Institute of medical Sciences, Lucknow, Uttar pradesh, India; 2) Dr. Balraj Mittal, Medical genetics, Sanjay Gandhi Post Graduate Institute of medical Sciences, Lucknow, Uttar pradesh, India; 3) Annapurna Gupta, Department of Medical Genetics, Medical genetics, Sanjay Gandhi Post Graduate Institute of medical Sciences, Lucknow, Uttar pradesh, India; 4) Kiran lata Sharma, Medical genetics, Sanjay Gandhi Post Graduate Institute of medical Sciences, Lucknow, Uttar pradesh, India; 5) Dr. Vijay Kumar, King George's Medical University, Lucknow, Uttar pradesh; 6) Dr. Ashok Kumar, Medical genetics, Sanjay Gandhi Post Graduate Institute of medical Sciences, Lucknow, Uttar pradesh, India.

Introduction: Genes important to stem cell development have been significantly implicated in the etiology and clinical outcome of various cancers. However, the associations of genetic variations in these genes with gallbladder cancer (GBC) have not yet been established. We investigated germline variants in Cancer stem cell genes (CSCs) to predict patient's susceptibility and efficacy with chemoradiotherapy treatment response. Methods: The study included 830 subjects, 580 GBC patients and 250 healthy controls. Among 580 patients, 164 patients were analyzed for chemoradiotheruaptic response. This study examined association of CD44 rs13347C/T, CD44 rs353639A/C, CD44 rs187116G/A, CD44 rs187115T/C, ALCAM rs1157G/A, ALCAM rs10511244 T/C, EpCAM rs1126497T/C, EpCAM rs1421T/C, CD133 rs3130C/T, CD133 rs2240688T/C, ALDH1A1 rs13959A/G, OCT-4 3130932T/G, LIN-28B rs4274112T/C, NANOG rs11055786C/T, SOX-2 rs11915160A/C polymorphisms. Genotyping was done by PCR-RFLP, ARMS-PCR and Taqman allelic discrimination assays. The association of clinicopathological, treatment response and toxicity with CSCs genetic variants was performed by SPSSver16. In-silco analysis was performed using Bioinformatics tools (F-SNP, FAST-SNP). Results: Statistical analysis by logistic regression showed significant association of CD44 rs353639A/C [CC, p=0.02], ALCAM rs1157G/A [AA, p=0.007], ALCAM rs10511244T/C [CC, p=0.009], OCT-4 rs3130932T/G [GG, p=0.01], LIN-28B rs4274112T/C [TC, p=0.03 CC, p=0.03], NANOG rs11055786 C/T [CT, p=0.03] polymorphisms with increased risk of GBC. Haplotype of CD44 [TAAC] showed significant association with reduced risk in GBC males and ALCAM [AC] haplotype associated with increased risk in GBC females. EpCAM rs1421T/C [TC, p=0.01] showed a statistically significant association with hematological toxicity and negative Lymph node status while SOX-2 rs11915160 A/C [CA, p=0.02] and NANOG rs11055786T/C [CT, 0.01] were associated with positive lymph node status. CD44 rs13347C/T [TT, p=0.02] and EpCAM rs1126497T/C [CC, p=0.04] were associated with good treatment outcome response. In-silico analysis of SNPs revealed variable change in transcriptional and splicing regulation. Conclusion: This is the first report of association of cancer stem cell genetic variants with GBC susceptibility and prognosis. Our results suggest that cancer stem cell genetic variants may also have a role with GBC susceptibility and clinical outcomes in North Indian population.

3473S

Association of GST polymorphism with susceptibility to Leuekmia and differential chemotherapy response. *S. Caplash[1], S. Kaur[1], R. Arora[2].* 1) Department of Human Genetics, Punjabi University, Patiala, India; 2) Oswal Cancer Hospital, Ludhiana.

Glutathione S-transferases (GST) isoenzymes play a significant role in phase II biotransformation and detoxification of many xenobiotics including environmental carcinogens, pollutants and drugs. Both the genetic polymorphisms and expression pattern of GST genes may have a major impact on cancer susceptibility, inter-individual variability in the prognosis, drug effects and toxicity. Of these, GSTT1 and GSTM1 isoenzymes are highly polymorphic with homozygous deletion of either or both genes resulting in absence of enzyme activity. We carried out a case-control study involving 150 Leukemia patients and 134 normal healthy controls from Punjab (North India). Multiplex PCR was carried out to determine GSTM1 and GSTT1 polymorphism. The frequency of individuals carrying GSTM1 and GSTT1 null genotypes was higher among Leukemia patients (42 % and 23 %) as compared to the control group (41 % and 19 %), although the difference found was not statistically significant (p>0.05). Heterogeneity in patients's response to chemotherapy is consistently observed across populations so follow up of patients is in process to analyze the association of GST genotypes with differential chemotherapy drug response.

3474M

BCR-JAK2 Translocation with Compleks Chromosomal Karyotype. *A. Ozturk Kaymak[1], C. Sonmez[2], S. Oztomurcuk[1], G. Guntas[2], S. Civriz Bozdag[3], F. Altuntas[3].* 1) Medical Genetics, Dr. A.Y. demeteveler Oncology Research and education Hospital, ANKARA, Turkey; 2) Biochemistry, Dr. A.Y. demeteveler Oncology Research and education Hospital, ANKARA, Turkey; 3) Heamatology,Dr. A.Y. demeteveler Oncology Research and education Hospital, ANKARA, Turkey.

sequencing is routinly used for diagnosis in chronic myeloproliferative neoplasms(2). Polycythemia vera, primary myelofibrosis and essential thrombocythemia have been occured because of V617F, the common gain of function mutation. Although the high ratio of JAK2 point mutations, JAK2 rearrangments are rare and have been associated with ALL and CML. Today only about 22 cases are reported with JAK2 rearrangment. Previous cases with JAK2 rearrangments are diagnosed with acute myeloid leukemia and lymphoblastic leukemia, atypical chronic leukemia, myelofibrosis, myelodysplastic syndrome (4, 5). The partners of chromosome 9 are various such as chromosome 2, 3, 4, 5, 6, 8, 12 and 22 at JAK2 reaarangments. The fusion genes are ETV6, PCM1, RPN1, NFE2, PAX5,SSBP2 and BCR. BCR-JAK2 fusion has leukemogenic role in myeloproliferative disorders. Our patient had also chromosme 5q deletion that is one of the most common structural rearrangements in MDS. The last anomaly of our patient was 8p11 deletion that is also associated with myeloproliferative neoplasms. Here we present BCR- JAK2 translocation with kompleks chromosomal karyotype. With our case, the total number of cases with BCR-JAK2 fusion in leukemia become five. But our case is the first one with kompleks chromosomal karyotype.

3475T

Evaluation of rapid whole-body magnetic resonance as screening strategy for early cancer detection in Li-Fraumeni syndrome patients. *M. Achatz[1, 2], D. Paixao[1], M.D. Guimaraes[3], A. Nogueira[1], R. Chojniak[3].* 1) Department of Oncogenetics, A.C.Camargo Cancer Center, Sao Paulo, Brazil; 2) International Research Center, A.C. Camargo Cancer Center; 3) Department of Imaging, A.C. Camargo Cancer Center.

Li-Fraumeni Syndrome (LFS) is a rare autosomal dominant syndrome that predisposes to a high-risk for developing multiple early onset cancers that includes breast, sarcomas, brain and adrenocortical carcinoma. It is related to germline mutations in TP53 gene. Screening strategies for early diagnosis in carriers constitutes a major challenge due to the wide tumor spectrum. Rapid whole body MRI (RWB-MRI) has been proposed as a screening strategy according to the Toronto protocol and its effectiveness needs to be accessed in different populations. In Brazil, there is a high prevalence of TP53 mutation carriers due to the occurrence of a founder mutation, p.R337H TP53, detected in 0.3% of South and Southeastern Brazilian population. This mutation occurs in the oligomerization domain and has a lower penetrance than regular DNA biding domain mutations in the TP53 gene. Moreover, due to genetic modifiers, tumors occur at a later age than in the regular LFS carriers. This was probably one of the reasons why the mutation spread to such proportions in Brazil. This constitutes a public health issue since its occurrence may be present in a large number of people in the most populated area in the country. The aim of this study is to evaluate the efficacy of RWB-MRI for early cancer detection in p.R337H TP53 mutation carriers and compare it to classical LFS mutation carriers.RWB-MRI was performed in 33 TP53 germline mutation carriers, including 29 cases with p.R337H. One malignant lesion was detected in an eighteen years-old p.R337H female carrier. RWB-MRI evidenced bilateral renal cortical alterations. Images were further confirmed with abdominal MRI which showed a solid lesion with enhancement in right kidney and a lesion with benign aspects in the left kidney. Lesion in the right kidney was surgically removed and anatomic-pathological findings indicated a papillary renal cell carcinoma. Incidental findings were detected in two out of the 33 patients. Further imaging techniques were done to detail diagnosis and no unnecessary biopsies were performed. No lesions were detected among the classical LFS carriers. The same cohort will further undergo WB-MRI in a 12- and 14 month interval. Preliminary results from this study demonstrated the effectiveness of RWB-MRI in detecting one case of a malignant lesion among 29 p.R337H LFS patients. It indicates that this might be an important imaging tool for early detection in LFS p.R337H TP53 mutation carriers.

3476S

Spectrum of mutations in BRCA1 and BRCA2 genes in Hereditary Breast/Ovarian Cancer families from Algeria: current knowledge and implications in genetic counseling and testing. *F. Cherbal[1], R. Bakour[1], C. Mehemmai[1], H. Gaceb[1], I. Derouiche[1], K. Akli[1], K. Boualga[2].* 1) Unit of Genetics, LMCB, Faculty of Biological Sciences, USTHB, Algiers, Algeria; 2) Anti Cancer Center, Blida, Algeria.

Background: Breast cancer is currently the leading cause of cancer morbidity and mortality among Algerian women. To date, the mean age of breast cancer cases in Algeria is younger than in North America or Europe. Between 2008 and 2014, 136 Algerian breast/ovarian cancer families were screened for BRCA1 and BRCA2 germline mutations (70 families) and for common mutations detected previously in BRCA1 gene of Algerian, Moroccan and Tunisian patients (66 families), respectively. Methods: 152 Algerian probands and relatives were screened for complete BRCA1 and BRCA2 germline mutations by using HRM- direct sequencing (86 individuals) and for 5 recurrent germline mutations detected previously in BRCA1 gene of patients from Algeria, Morocco and Tunisia by using PCR-direct sequencing (66 individuals). Screening for large genomic rearrangements (LGR) in BRCA1 and BRCA2 genes was performed in 76 patients tested negative for small point mutations by using MLPA technique. To confirm the genomic rearrangement and its genomic breakpoints, we performed long-range PCR of genomic DNA. In silico prediction for novel missense variants effects was carried out by using 5 bioinformatics programs: Align-GVGD, Polymorphism Phenotyping2, SIFT program, Mutation Taster and Mutation Assessor. Results: The analysis of DNA samples of 152 individuals revealed that 12 patients and one male relative carried pathogenic germline mutations: 11 within BRCA1 and 2 within BRCA2. In addition, 10 new missense variants have been detected in our study (3 BRCA1 and 7 BRCA2). Four distinct pathogenic mutations, c.19_47del, c.83_84delTG, c.181T>G, c.798_799delTT and two LGR were identified in BRCA1 gene. A novel deletion of BRCA1 exon 2 (c.-19-?_80?del/p.?) and a novel genomic breakpoint in deletion of BRCA1 exon 8 (c.442-?_547?del/p?) were identified in two unrelated breast cancer patients. No LGRs were detected in BRCA2 gene. We found by using long range PCR technique that our patient with the BRCA1 exon 8 deletion is heterozygous for a 2.6 kb deletion. Interestingly, the three common BRCA1 mutations, c.83_84delTG, c.181T>G and c.798-799delTT have been detected in 6 unrelated families (each mutation/2 families). We noted that three new missense variants BRCA1 R1753G, BRCA2 H2116D and BRCA2 G3086R were predicted to be likely pathogenic in silico. Conclusions: The identification of common mutations in BRCA genes in Algerian population may facilitate genetic counseling and testing.

3477M

Single-cell genetic analysis reveals insights into clonal development of prostate cancers and indicates loss of *PTEN* as a marker of poor prognosis. *K.M. Heselmeyer-Haddad[1], L.Y. Berroa Garcia[1], A. Bradley[1], L. Hernandez[1], Y. Hu[1], J. Habermann[3], C. Dumke[3], C. Thorns[4], S. Perner[5], E. Pestova[6], C. Burke[10], S.A. Chowdhury[7,8], R. Schwartz[8,9], A.A. Schäffer[2], P.L. Paris[10], T. Ried[1].* 1) Genetics Branch, CCR, NCI, NIH, Bethesda, MD; 2) Computational Biology Branch, NCBI, NIH, Bethesda, MD; 3) Section for Translational Surgical Oncology and Biobanking, Department of Surgery, University of Lübeck, Germany; 4) Institute of Pathology, University of Lübeck, Germany; 5) Department of Prostate Cancer Research, Institute of Pathology, University Hospital of Bonn, Germany; 6) Abbott Molecular, Des Plaines, IL; 7) Joint Carnegie Mellon/University of Pittsburgh Ph.D. Program in Computational Biology, Carnegie Mellon University, Pittsburgh, PA; 8) Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA; 9) Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA; 10) Department of Urology, Helen Diller Family Comprehensive Cancer Center, UCSF, San Francisco, CA.

One of the major challenges in the clinical management of patients with prostate cancer is to gauge the risk of developing progressive disease. We therefore used genetic markers to understand the dynamics of genomic alterations during disease progression. Using a novel, advanced multicolor fluorescence in situ hybridization (FISH) approach, we enumerated copy number of six genes previously identified by array CGH to be involved in aggressive prostate cancer - *TBL1XR1*, *CTTNBP2*, *MYC*, *PTEN*, *MEN1*, and *PDGFB* - in six non-recurrent and seven recurrent radical prostatectomy cases. An *ERG* break-apart probe to detect *TMPRSS2-ERG* fusions was included. Subsequent hybridization of multicolor probe panels and relocation of cells resulted in signal counts for all probes in each individual cell analyzed. We neither observed differences in the degree of chromosomal and genomic instability, i.e., intratumor heterogeneity, nor in the percentage of cells with *TMPRSS2-ERG* fusion between samples with or without progression. However, tumors from patients that progressed had more chromosomal gains and losses, and showed a higher degree of selection for a predominant clonal pattern. The loss of *PTEN* was the most frequent aberration in progressors (57%), followed by the gain of *TBL1XR1* (29%). The gain of *MYC* was observed in one progressor, which was the only lesion with an *ERG* gain, but no *TMPRSS2-ERG* fusion. According to our results, a probe set consisting of *PTEN*, *MYC* and *TBL1XR1* would detect progressors with 86% sensitivity and 100% specificity. This will be evaluated further in larger studies.

3478T

Sequence variations in known cancer susceptibility genes identified in high-risk breast cancer cases from the French GENESIS study. *F. Lesueur[1], E. Girard[1], F. Damiola[2], M.G. Dondon[1], L. Barjhoux[2], S. Eon-Marchais[1], M. Gauthier-Villars[3], B. Buecher[3], C. Lasset[4], P. Berthet[5], C. Delnatte[6], J.P. Fricker[7], M. Longy[8], O. Caron[9], P. Pujol[10], A. Chevrier[11], P. Gesta[12], E. Mouret-Fourme[13], C. Maugard[14], N. Servant[1], V. Meyer[15], A. Boland-Auge[15], E. Barillot[1], J.F. Deleuze[15], S. Mazoyer[2], N. Andrieu[1], O. Sinilnikova[2, 16], D. Stoppa-Lyonnet[3], CGEI platform, GENESIS investigators.* 1) Inserm U900, Institut Curie, Mines ParisTech, Paris, France; 2) Genetics of Breast Cancer group, Cancer Research Center of Lyon, INSERM U1052, CNRS UMR5286, Université de Lyon, Centre Léon Bérard, Lyon, France; 3) Service de Génétique, Institut Curie, Paris, France; 4) Unit of Genetic Epidemiology and Prevention, Centre Léon Bérard, Lyon, France; 5) Unit of gynecological pathology, centre François Baclesse, Caen, France; 6) Centre René Gauducheau, Nantes, France; 7) Unit of Oncology, Centre Paul Strauss, Strasbourg, France; 8) Laboratory of Molecular Genetics, Institut Bergonié, Bordeaux, France; 9) Department of Medical Oncology, Institut Gustave Roussy, Villejuif, France; 10) Unité d'oncogénétique, Hôpital Arnaud de Villeneuve, Montpellier, France; 11) Centre Henri Becquerel, Rouen, France; 12) Centre Georges Renon, Niort, France; 13) Department of Public Health, Institut Curie, France; 14) Laboratoire de diagnostic génétique, Nouvel Hôpital civil, Strasbourg, France; 15) Centre National de Génotypage, Evry, France; 16) Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon, Centre Léon Bérard, Lyon, France.

In France, a pathogenic BRCA1 or BRCA2 mutation is found in at most 16% of tested index cases of families predisposed to breast and ovarian cancers. Other breast cancer (BC) predisposing genes have been recently identified (PALB2, BRIP1, RAD51C, RAD51D, XRCC2, etc.), and with the introduction of massive-parallel sequencing (MPS) in clinical laboratories some of these genes are already being screened in patients referred to the clinics as part of multigene panel testing. However their use remains limited in clinical practice because their associated cancer risks have not been precisely estimated yet due to the too small number of families segregating the mutations identified so far. Our goal was to assess the frequencies of protein-truncating variants and rare likely deleterious missense substitutions in the newly identified BC susceptibility genes in a sample of high-risk women participating to the GENESIS study, in order to estimate the number of families that would need to be tested to get a precise estimate of cancer risks in mutation carriers. GENESIS is a unique French national existing resource for familial BC research where >1700 index cases affected with BC and negative for BRCA1/2, >800 affected sisters and >1,500 unrelated matched healthy women have been enrolled. Blood samples, clinical and epidemiological data have been collected for each study participant through the national network of cancer genetics clinics. High-throughput genotyping and MPS projects have been initiated in order to decipher the genetic landscape of BC. Here we examined rare variants (with reported frequency <0.5% in the public sequencing databases) in known BC susceptibility genes, mismatch repair genes and Fanconi Anemia genes identified through exome sequencing in a subset of 100 GENESIS high-risk cases. We found 2 protein-truncating mutations (TM) and 3 likely deleterious missense substitutions (MS) in ATM, 1 TM and 4 MS in CHEK2, and 1 TM and 5 MS in genes of the MRN complex (MRE11A, RAD50 and NBN) occurring in unrelated families. We will next conduct full mutation screening of the gene panel in GENESIS cases and controls to estimate the frequency and the distribution of the different classes of variants in the two groups. National family studies will be set up to assess penetrance and to study cancer spectrum associated with the pathogenic variants. This should help clarifying relevance of mutation screening of these genes to familial cancer clinics and counselors.

3479S

Genome-wide association study (GWAS) for transaminase elevations in pazopanib-treated patients. *X. Wang[1], Z. Xue[1], C. Carpenter[2], P. Harter[3], K. King[1], S. Stinnett[1], T. Johnson[4], J. Shen[1], C. Spraggs[4], M.R. Nelson[1], M. Chen[2], K. Deen[2], I. Mitrica[2], L. Pandite[1], N. Kaplowitz[5], C-F. Xu[4].* 1) GlaxoSmithKline, Research Triangle Park, NC, USA; 2) GlaxoSmithKline, Collegeville, PA, USA; 3) Kliniken Essen-Mitte, Essen Germany; 4) GlaxoSmithKline, Stevenage, UK; 5) Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

Background: Pazopanib (Votrient™, GlaxoSmithKline), an oral angiogenesis inhibitor, has been approved for the treatment of advanced renal cell carcinoma and soft tissue sarcoma. Transaminase elevations have been commonly observed in pazopanib-treated patients in clinical studies. We conducted a GWAS using data from 8 phase II and III pazopanib clinical studies to evaluate association between genetic variants and treatment-emergent alanine transaminase (ALT) elevation. **Methods:** 1271 pazopanib-treated (monotherapy) patients with cancer were included in the analysis. Thirty million genotyped or imputed variants were tested for association with on-treatment peak ALT and time to first ALT ≥3x upper limit of normal (ULN) or ≥5xULN. Further analysis with sequence-based 4-digit HLA alleles from 5 genes was also performed. **Results:** One genetic variant, near NNT, rs80228453 (minor allele frequency 7.5%) was associated with peak ALT ($P=2\times10^{-8}$) at a genome-wide significance level for common variants ($P\leq5\times10^{-8}$), with the variant genotype associated with increased ALT. Among patients with ALT≥3xULN, 18% carried the NNT variant genotype (vs 6% in patients with ALT≤ULN). NNT encodes nicotinamide nucleotide transhydrogenase, which is essential for mitochondrial defense against oxidative stress and redox detoxification. Additional variants associated with ALT elevation at $P\leq5\times10^{-7}$ were observed, some of which are in genes/regions of known implications for drug-induced liver injury, including the MHC region. Subsequent HLA analysis identified an association between HLA-B*57:01 and maximum ALT ($P=1\times10^{-4}$) as well as time to 5xULN ($P=9\times10^{-5}$), statistically significant after adjustment for the number of HLA alleles tested ($P\leq5\times10^{-4}$). HLA-B*57:01 carriers were seen in 5% of all patients, 10% of patients with ALT≥3xULN and 12% with ALT≥5xULN. The incidence rates of ALT≥3xULN and ALT≥5xULN were 36% and 27%, respectively, in B*57:01 carriers, and 19% and 11%, respectively, in B*57:01 non-carriers, with an odds ratio (95% confidence interval, P value) of 3.2 (1.6-6.6, 9×10^{-4}) and 4.6 (2.1-9.9, 1×10^{-4}), respectively. **Conclusions:** We identified statistically significant associations at NNT and HLA-B for ALT elevation in pazopanib-treated patients. If replicated, these findings would provide novel insight into independent mechanisms of pazopanib liver toxicity and may influence clinical management.

3480M

Correlation among MDR1, MRP and hTERT expression level and clinical response in colon cancer patients. *s. sha'bani[1], f. mahjoubi[1], s. samanian[1], b. mahjoubi[2], r. mirzaei[2], a. tavakoli[1].* 1) national institue of genetic engineering and biotechnology, tehran, Select a Country; 2) iran university of medical science.

Introduction: Colon cancer is one of the common cancers in the world. Despite current advances in the treatments of cancer, the clinical result is far away from expectation yet. Drug resistance is still a major obstacle in treatment of cancer .In this study, we attempted to investigate the possible correlation among MRP1, MRP and hTERT expression level and multidrug resistance in colon cancer patients. Materials and Methods: Tumor and adjacent normal tissues from 35 colorectal cancer patients were assessed for the mRNA expression level of MDR1, MRP and hTERT by Real Time RT-PCR. Results: A statistically significant increase in MDR1 and hTERT expression level was observed in tumoral tissues in comparison with normal tissues. However MRP expression level was not significantly increased in tumoral tissues. Furthermore, no correlation was seen among MDR1, MRP and hTERT expression level. Conclusion: MDR1 and hTERT have no direct correlation, but mRNA expression of these two genes in addition to other factors indirectly helps to tumorgenesis and cancer progression.

3481T

Association of IL-1β gene Polymorphism with HCC related to viral causes. *H.A. Abdalla[1], A.A. Badawy[1], R. Monir[1], K. Farag[2].* 1) Medical Biochemistry, Faculty of Medicine, Mansoura University, Mansoura, Egypt; 2) Oncology Center, Mansoura University, Egypt.

Hepatocellular carcinoma (HCC) is the 5th most common malignancy in the world. In Egypt, the incidence of HCC is doubled over the last decade. Recently, numerous functional gene polymorphisms among pro-inflammatory and pro-fibrogenetic factors and their inhibitors have been thought to influence carcinogenesis including Interleukine-1β (IL-1β), which is considered one of the potent cytokines that has effects on cell survival and proliferation. The aim of the current study was to investigate the association of IL-1β gene polymorphism with HCC patients on top of viral versus non viral cases. The study included 134 subjects; 89 patients suffering from HCC; 61 patients with HCC on top of viral hepatitis (37 HCV and 24 HBV) and 28 patients with HCC without viral hepatitis, and 45 healthy individuals with no liver affection. IL-1β -31 gene polymorphism was studied by PCR amplification followed by RFLP analysis. Quantitative determination of serum IL-1β level was performed by ELISA. In patients with HCC on top of chronic HCV infection, IL-1β TT genotype was found to be more frequent than other genotypes ; whereas in patients with HCC on top of chronic HBV infection, IL-1β CT genotype was more frequent than other genotypes. IL-1β is significantly elevated in HCC patient on top of both types of virus hepatitis when compared with HCC due to non-viral causes ($p<0.05$). It is shown that IL-1β was highly significant elevated in patients with TT genotype of IL-1β gene rather than other IL-1β genotypes ($p=0.004$). In conclusion, People with T homozygous of IL-1β -31 gene polymorphism in combination with serum level of IL-1β could be considered as candidate molecular biomarkers for development of HCC on top chronic HCV or HBV infection. Further studies are recommended to explain and confirm these conclusions.

3482S

Opposite expression regulation of ATP-binding cassette transporters (ABC) genes in leukemic cells during granulocyti maturation processes on imatinib therapy. *L.SRA Pedroza[1], C.A.F. Nunes[1,2], J.A.R. Lemos[1,2].* 1) Federal University of Pará State (UFPA) Belem, Brazil; 2) HEMOPA and Fundação Centro de Hemoterapia e Hematologia do Pará (HEMOPA), Brazil.

Despite the results of imatinib (TKI) therapy to control patients Chronic Myeloide Leukemia (CML) progress, prolong overall survival of 88% of new diagnosed patients, 20% will fail. Indeed, TKI are not curative options, cause it can not eliminated all cancer stem cells that differentiated to mature hematopoietic blood cells, and disease can relapse. The search for mRNA related to prognostic predicts, and membrane transports responsible for imatinib efflux from the stem cells became a way to overcome resistance. Multiresistant protein - MRP and Multidrug resistant - MDR, associated to drug resistance in a tumor/drug specific way and ABCG2/B1 responsible for the imatinib extrusion are members of ABC superfamily. To gain more insight about the ABCs expression on leukemic cells on imatinib therapy we assess the transcriptome of CD34+ cells and CD66b+ cells into a pool of 4 CML patients in sustained major molecular response - MMR in imatinib treatment, and a control (1). A total of 19 (A3, A4, A7, C5, C11, C6P1, A13, A9, A8, B5, B4, B7, B6, C9,C12,C2,D2,D3) and 16 (A4,A5, A8, B5, B6,B8,C1,C5,C13,D4,E1,A13,A9,C6,C4) genes have differential expression in CD34+ cells and CD66b+ leukemic cells when compared to normal counterpart respectively. Leukemic CD34+ cells are less different from their control counterpart than leukemic CD66b+. These cells have a lake of expression to three subfamilies as in stem cells, all subfamilies presenting in patient group have at least one representative expresses gene in normal sample and the ranking of representatives genes to each subfamilies is the same for CD34+ cells from the two groups. We observe threaten upregulated ABC genes in CML CD34+ cells and twelve downregulated ABC genes in leukemic CD66b+ cells compared to the health counterparts, highlight a possible opposite regulation of ABC genes in leukemic cells during granulocyti maturation processes. We are choice possible candidate genes to modulation through imatinib therapy and understanding the ABCs possible functions in hematopoietic stem cells and mature blood cells. We aim to validate these finds on peripheral blood and in cito. Understand the pathways involved in ABCs function and know patients ABC expression profile considering pharmacogenetic interaction are crucial to test new therapy protocol in vitro or to clinical trial and management.

3483M

Inherited NK Cell Defective Mutations in Chinese Lymphoma Patients with HHV infection. *Y. Zhang[1], H. Liu[1], F. Wang[1], W. Teng[1], X. Chen[2], P. Zhu[2], C. Tong[1], Y. Lin[1], J. Yang[1], D. Lu[1].* 1) Clinical Laboratory Division, Ludaopei Hematology & Oncology Center, Beijing, China; 2) Department of Hematology, Peking University First Hospital, Beijing.

Background: It has been reported that some patients with lymphoma may harbor mutations of PRF1, UNC13D and STXBP2, which lead to NK cell deficiency and are causative genes of familial hemophagocytic lymphohistiocytosis (FHL). Data of the association between genetic defects and lymphoma is limited for Chinese patients. **Methods and Cases:** This study included 45 unrelated Chinese patients with either Hodgkin or non-Hodgkin lymphomas (26 males, 19 females; age range 3 to 60 years) and with active HHV1-8 infection. PCR and Sanger sequencing of all coding exons and flanking sequences of UNC13D, PRF1, STXBP2 and STX11 were performed. Active HHV1-8 infections were tested by PCR method. **Results:** Mutations were observed in 12/45 patients (26.7%). 9/12 (75%) patients carried a total of 8 different UNC13D mutations, including 6 with monoallelic mutations, 1 with homozygous mutation, and 2 with compound heterozygous mutations. 2 patients had PRF1mutations, one with monoallelic mutation and the other with biallelic mutations. 1 patient was detected to carry STX11 monoallelic missense mutation. No patient carried STXBP2 mutation. All mutations in these patients were confirmed to be germline derived by pedigree analysis or sequencing of non-lymphoma tissue of the same patient. **Conclusions:** Germline UNC13D mutations are frequent in Chinese lymphoma patients with HHV infection. Involvement is less frequent for PRF1 and STX11. Inherited NK cell defective mutations might be the innate predisposing factor of active HHV infection and together lead to the lymphomagenesis.

3484T

Whole genome exome sequencing to identify novel candidates for hereditary predisposition to UM uveal melanoma. *M.H. Abdel-Rahman[1, 2, 3], D. Hedges[2], R. Pilarski[2], G. Boru[1], J.B. Massengill[1], K. Rai[2], F.H. Davidorf[1], C.M. Cebulla[1].* 1) Dept Ophthalmology, The Ohio State University, Columbus, OH; 2) Division of Human Genetics, Department of Internal Medicine, The Ohio State University, Columbus, OH; 3) Pathology Department, Menoufiya University, Shebin Elkom, Egypt.

Background: About 12% of uveal melanoma (UM) patients have features suggestive of hereditary cancer predisposition. Germline mutation in BAP1 explains only a small subset of these patients suggesting the existence of other candidate genes. Methods: Whole exome sequencing (WES) was carried out on six BAP1 mutation-negative UM patients from 5 different families. Three patients had family history of UM and two patients had personal history of UM and personal or family history of cancers associated with the BAP1 tumor predisposition syndrome (TPDS). In one family, the index case and a paternal 1st cousin once removed with UM were sequenced. Results: We focused our primary analysis on identifying unreported, deleterious or potentially deleterious (stop loss, stop gain, nonsynonymous, splice variant, frame shift) germline variants in any of the 522 COSMIC "cancer census genes". We identified deleterious or possibly deleterious mutations in MAML2, TPR, ARID1A, PBRM1, PRDM1 and MUTYH. These variants were subsequently confirmed by direct sequencing. Segregation analysis ruled out contribution of MAML2, ARID1A, PBRM1, PRDM1 and MUTYH. Also, sequencing of the tumor tissue from the patient with germline TPR (translocated promoter region) mutation showed significant loss of the mutant allele in tumor suggesting that it is likely not contributing to the pathogenesis of the disease. No pathogenic variant was detected in any of the cancer predisposition gene in more than one patient. However, we identified several potential candidate genes that have not been previously implicated in hereditary cancer predisposition. Conclusions: This pilot whole exome sequencing (WES) study suggests that hereditary cancer predisposition in these patients is not caused by coding mutation in any of the 522 COSMIC cancer census genes. It also suggests locus heterogeneity across families, causative genetic alterations in non-coding regions and/or both as mechanisms for hereditary predisposition to UM. We have identified several additional potential candidate genes that we are further exploring. Our study provides important preliminary data to further identify potential candidate genes for hereditary predisposition to UM.

3485S

Two novel germline *BAP1* mutations in two unrelated families with features of the *BAP1* Tumor Predisposition Syndrome. *C.M. Cebulla[1], R. Pilarski[2], K. Rai[2], JB. Massengill[1], G. Boru[1], FH. Davidorf[1], MH. Abdel-Rahman[1,2,3].* 1) Ophthalmology and Visual Science, The Ohio State University, Columbus, OH; 2) Division of Human Genetics, Department of Internal Medicine and Comprehensive Cancer Center, The Ohio State University, Columbus, OH; 3) Department of Pathology, Menoufiya University, Egypt.

Objective: The *BAP1* Tumor Predisposition Syndrome (TPDS) is a recently recognized hereditary cancer predisposition syndrome. Four main cancers, uveal melanoma (UM), cutaneous melanoma, mesothelioma, and renal carcinoma, are associated with this syndrome, in addition to other cancers. Here we report the clinical phenotype of two new independent families with novel germline *BAP1* mutations and features suggestive of the *BAP1* TPDS.

Methods: Germline *BAP1* mutations were evaluated by direct sequencing of all *BAP1* exons and adjacent intronic regions. Two new families were evaluated, one presenting with cutaneous melanoma and mesothelioma in the proband and a family history of UM and several other cancers. In the second family, the proband presented with peritoneal mesothelioma and a strong family history of pleural mesothelioma as well as UM.

Results: In the first family we identified a nonsynonymous mutation in the ubiquitin carboxyl-terminal hydrolase domain of *BAP1* (c. 539T>C, p. Leu180Pro). The mutation was present in a first degree relative with UM (obligate carrier) and in an unaffected second degree relative of the proband. However, the mutation was not identified in other first and second degree relatives in the family, including those with cutaneous melanoma and liposarcoma. In the second family a 3 bp intronic deletion was identified (g.2270_2272delACA, c. 256-4_256-2del). This mutation was predicted to be a splice site pathogenic variant by NetGene2 and Mutation Taster software. Validation studies and testing of additional family members are ongoing.

Conclusion: We report 2 novel germline mutations in *BAP1* in two unrelated families associated with the typical *BAP1* TPDS phenotype.

3486M

Predisposition to Burkitt Lymphoma in Williams-Beuren syndrome. *D. Guenat[1,2], J. Soulier[3], C. Rizzari[4], H. Fryssira[5], C. Lundin[6], C. Borg[2,7], PS. Rohrlich[2,8].* 1) Laboratory of Cellular and Molecular Biology, University Hospital of Besançon, Besançon, France; 2) UMR1098 Inserm/EFS-BFC/UFC, LabEx LipSTIC, Besançon, France; 3) Saint-Louis Hospital APHP and Hematology University Institute (IUH), University Paris-Diderot, Paris, France; 4) Department of Pediatrics, San Gerardo Hospital, University of Milano-Bicocca, Monza, Italy; 5) Department of Medical Genetics, Medical School, University of Athens, Greece; 6) Department of Clinical Genetics, University and Regional Laboratories, Skåne University Hospital, Lund University, Lund, Sweden; 7) Department of Medical Oncology, University Hospital of Besançon, Besançon, France; 8) Department of Hematology, University Hospital of Nice, Nice, France.

Chromosomal disorders are frequently associated with predisposition to cancer as shown by the increased incidence of leukemia in Down syndrome. Williams-Beuren Syndrome (WBS [MIM 194050]) is a multisystem disorder caused by a hemizygous microdeletion on chromosome 7q11.23 spanning 1.5 Mb and encompassing 28 genes and 2 miRNA loci. WBS is not currently considered as a risk factor for cancer. However, a thorough review of the literature shows that 9 cases of pediatric cancers have been reported in WBS patients and, strikingly, 6 (66%) of them were Burkitt Lymphomas (BL). We report here 2 novel cases of BL in children with WBS. DNA isolated from lymphoma cells and normal tissue of the 2 novel cases (one from France and one from Italy) was analyzed. In addition, tumor and normal DNA of 2 other children were investigated: a Greek child with WBS who presented a BL whose case was reported in 2004 and a Swedish child without WBS but with a sporadic lymphoma that has shown a 7q11.23 deletion in a previous study. CGH-array and Next Generation Sequencing were used to characterize the size of the WBS deletion and to seek for the presence of a loss of heterozygosity (LOH) in the 7q11.23 locus. The constitutional microdeletion observed in the normal DNA of the 3 WBS patients corresponded to the typical WBS critical region with a length of 1.5 Mb and there was no evidence of LOH in the tumor DNA. Interestingly, the somatic deletion observed in the sporadic BL was mono-allelic and similar to the one observed in WBS patients. No other rearrangement was observed in this tumor. Thus, BL may arise in the presence of either constitutional or somatic microdeletion of 7q11.23 and in the absence of LOH in the WBS critical region. A number of genes mapping to this region are involved in DNA repair pathways (*BAZ1B*, *RFC2*, *GTF2I* family genes) or B-cell proliferation and differentiation (*LAT2*, *FZD9*). The haploinsufficiency of one or of a combination of these genes might be responsible for genetic instability and lymphoma initiation. Given these observations and the difficult management of epidemiological data concerning association of rare diseases such as WBS and BL, we suggest that the deletion of the WBS critical region is involved in the initiation of BL and that WBS patients have an enhanced risk of Burkitt lymphoma occurring at an early age. Functional studies are needed to identify the precise molecular pathways that drive the B-cell lymphomagenesis in WBS patients.

3487T

Familial Inflammatory Fibroid Polyps Syndrome (FIFPS): Phenotype much broader than having polyps. *J. Walia[1,2], S. Abu-Abed[2], A. Hawrysh[1], N. Perrier[1], L. Hookey[3], R. Giddell[4], J. Chien-Hung[2], R. Kirsch[4], H. Feilotter[2], P. Manley[2].* 1) Division of Medical Genetics, Department of Pediatrics, Queen's University, Kingston, ON, Canada; 2) Department of Pathology and Molecular Medicine, Queens University, Kingston, ON Canada; 3) Division of Gastroenterology, Department of Internal Medicine, Queen's University, Kingston, ON, Canada; 4) Pathology & Lab Medicine Mount Sinai Hospital, Toronto, ON, Canada.

Inflammatory fibroid polyps (IFPs) are mesenchymal benign tumors of the gastrointestinal tract that have been reported to harbor somatic mutations in Platelet Derived Growth Factor Receptor Alpha (PDGFRA) especially in exon 12, 14 and 18. Three case reports have been documented with germline mutations as well. We report a family with germline mutation is exon 18 of PGFRA gene with a further description of physical features segregating with this mutation. The proband, a 59 year old female, presented with severe abdominal pain. Surgical exploration showed multiple intussusceptions of the small intestine. Direct and indirect imaging techniques revealed hundreds of small bowel sub-mucosal nodules and a few gastric and colonic ones. These were described as 'mutiple submucosal, intramural and serosal firm circumscribed lesions'. Pathological examination confirmed them to be IFPs. Molecular testing showed two different mutations- D846V (c.2537A>T) mutation in exon 18 was found in two polyps as well as normal small bowel tissue and an additional D576Y (c.1726G>T) mutation in exon 12 in one of the two polyps. D846V mutation was also indentified in DNA derived from blood, thus confirming it to be a germline mutation. Investigations are ongoing to see any susceptibility/presence of Gastrointestinal Stromal Tumors (GISTs). Multiple other members were also tested positive for the same mutation. Physical examination of mutation positive individuals showed coarser skin, and broader hands and feet as compared to mutation negative family members. In addition, the mutation positive members lost most of their teeth in late 30s or early 40s. PDGFRA gene has been reported to have an important role in skeletal development in addition to other developmental pathways and we hypothesize that this gain of function mutation is the cause of abovementioned phenotypic features. To our knowledge, this is a first report of a family with a comprehensive genetic, morphological, pathological and clinical characterization of Familial IFPs Syndrome. Such information will not only help in the clinical recognition of affected individuals, but also in follow-up investigations for IFPs and possibly related GISTs.

3488S

High Depth HPV16 Whole Genome Sequencing of 830 PaP Cohort Specimens using Crude Exfoliated Cervical Extracts. *J. Boland[1], M. Cullen[1], L. Mirabello[2], J. Mitchell[1], S. Bass[1], X. Zhang[1], M. Yeager[1], M. Andersen[7], M. Osentoski[7], M. Laptewicz[7], J. Smith[7], N. Wentzensen[2], K. Yu[2], Z. Chen[3], T.R. Raine-Bennett[5], P.E. Castle[6], M. Schiffman[2], R. Burk[3,4].* 1) Cancer Genomics Research Lab, NCI_Leidos, Gaithersburg, MD; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD; 3) Division of Gynecologic Oncology, Department of Obstetrics & Gynecology and Women's Health, at Albert Einstein College of Medicine, Bronx, NY; 4) Department of Epidemiology and Population Health, at Albert Einstein College of Medicine, Bronx, NY; 5) Women's Health Research Institute, Division of Research, Kaiser Permanente Northern California, Oakland CA; 6) Global Cancer Initiative, Chestertown, MD; 7) Thermo Fisher Scientific, 5781 Van Allen Way, Carlsbad, CA.

In this report, complete 7.9 kb genomes of 830 PaP (Persistence and Progression) cohort specimens were sequenced on the Thermo Fisher Scientific Ion Torrent Personal Genome Machine using a custom HPV16 Ion AmpliSeq panel. Pre-pooled multiplexed primers targeted the 7.9kb circular viral genome amplifying 47 overlapping amplicons at a coverage of 99%. As little as four microliters of low quality exfoliated cervical tissue extract was amplified, regardless of viral load, in a robust multiplex reaction. Total time from sample to sequence data was three days. Sequence read alignment to the HPV16 reference genome and variant calls were done utilizing the on board Ion Torrent Suite of sequence analysis tools. Total aligned reads across all 47 overlapping amplicons translated into a depth of coverage per sample ranging from 400x to over 5000X. This depth of coverage has enabled the discovery of rare variants present at frequencies down to 2%. Over 750 variants were identified of which 480 were novel and not seen in the eight HPV16 reference genomes used to define HPV sub lineages. Concordance of variant calls across 32 sample duplicates within our study was 97.76 percent. As a fine measure of this novel method of sequencing HPV16 genomes we looked at concordance of variant calls between Sanger methods and our AmpliSeq/Ion Torrent data in a subset of 96 PAP samples. Lastly, with barcodes, your sequencing throughput can range from one sample to an entire 96 sample plate per sequencing run. Sequence output of the three different Ion Torrent sequencing chips (314, 316 and 318) enable scaling of sequencing needs. With a cost of ~$50/sample the AmpliSeq HPV16 panel is a low cost, high quality, throughput and depth (>400X) of read alternative to the Sanger method.

3489M

Genome-wide DNA methylation patterns and genetic ancestry in sporadic breast cancer patients from a Latino population. *M. Cappetta[1], L. Brignoni[1], N. Artagaveytia[2], O. Stefansson[3], M. Esteller[3,4,5], B. Bertoni[1], M. Berdasco[3].* 1) Departamento de Genética, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay; 2) Departamento Básico de Medicina, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay; 3) Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute, L'Hospitalet de Llobregar, Barcelona, Catalonia, Spain; 4) Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain; 5) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.

Genetic variants alone are not enough to explain a complex disease like cancer. Alterations in DNA methylation patterns have been associated with different types of tumors. Although methylation profiles are often tissue-specific, recent data indicate that epigenetic changes in peripheral blood leukocytes are promising risk markers for solid tumors.

In order to detect risk markers for sporadic breast cancer in the Uruguayan population, we integrated genetic and epigenetic information of individuals. We determined the level of global leukocyte DNA methylation (glDNAmet) in cancer patients by relative quantification of 5mC by HPLC-MS; and then, we focused on site-specific studies using HumanMethylation450 BeadChip.

A global DNA hypomethylation in breast cancer patients was detected when it was compared with healthy controls. This result suggests its potential use as a risk marker. Since the Uruguayan population is admixed, we studied the correlation between glDNAmet and individual genetic ancestry (determined by analyzing ancestry informative markers). We found a negative correlation between African ancestry and glDNAmet in cancer patients. These results suggest that the ancestral admixture genome structure could be modeling DNA methylation patterns and its effect and underscore the importance of considering genetic ancestry as a modifying variable in epigenetic association studies in admixed populations like Latino ones.

We detected 77 differentially methylated CpG sites in blood of breast cancer patients including known cancer genes involved in disease-specific pathways and also novel candidates with possible implication in breast tumorigenesis. This set was characterized and validated in an independent sample of breast tissue belonging to European breast cancer patients. These CpG sites are able to differentiate leukocytes of sporadic breast cancer patients from controls. It also distinguishes healthy tissue from breast tumor, even in cases of hereditary breast cancer.

We identified differential DNA methylation at global and site-specific level in leukocytes of patients with sporadic breast cancer, which suggests the existence of a systematic variation in DNA methylation associated with susceptibility to tumor development.

3490T

Routine use of massively parallel sequencing for BRCA1 and BRCA2 diagnosis: a comprehensive workflow combining PCR Multiplex and sequencing chemistry on Miseq. *F. Coulet, F. Pires, E. Guillerm, M. Eyries, C. Colas, F. Soubrier.* Oncogenetics, APHP- Pitie Salpetriere Hospital, Paris, France.

The aim of this study was to develop a massively parallel sequencing workflow for routine diagnosis for hereditary breast and ovarian cancer syndrome. The aim was to reduce the delay of results with a highly specific and sensitive method. A workflow was designed using a multiplex amplification approach (BRCA MASTR Dx) with 5 multiplex PCR reactions covering the complete coding regions of BRCA1 and BRCA2, followed by synthesis sequencing on Miseq. Bioinformatic analysis of the results was made through the use of adaptated settings in SeqNext software. A training set of 56 DNA samples containing polymorphisms, unclassified variants and pathogenic mutations previously identified by others technologies, was used. This group consisted of 448 variants controls polymorphisms, 23 deleterious mutations (2 missense, 2 nonsense, one splice mutation and 18 frameshift mutations) including 4 located in homopolymers. The worflow developped on the Miseq permitted the identification of all variants (n = 491), including those located in or close to homopolymers. The routine diagnosis was established for series of 48 patients per run to achieve a high depth reading (averaged per amplicon 4000X). Eighteen series of 48 patients (n=864) were analyzed by this approach. A series of 48 patients from DNA extraction to results is performed by a technician in 10 days, included medical validation. The workflow meets the sensitivity and specificity requirements for genetic diagnosis of breast and ovarian cancer and improves on the cost-effectiveness of current approaches. It could be respond to situations in which the expectation of the mutational status is an emergency such as the practice of contralateral prophylactic surgery or the development of targeted therapies (Anti PARP).

3491S

New genome-wide technologies and low volume, archival, formalin-fixed paraffin embedded material: are the two compatible? *L.M. FitzGerald[1], E.M. Wong[2], J.E. Joo[2], J. Pedersen[3], J. Mills[3], CH. Jung[4], J. Chung[4], G. Severi[5], M.C. Southey[2], G.G. Giles[1].* 1) Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, 3004, Australia; 2) Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Victoria 3010, Australia; 3) TissuPath Specialist Pathology, Victoria 3149, Australia; 4) Victorian Life Sciences Computation Initiative, Carlton, Melbourne, Victoria 3010, Australia; 5) Human Genetics Foundation, Torino, 10126, Italy.

Purpose: Archival formalin-fixed, paraffin-embedded (FFPE) tumour material is a widely available, valuable resource for genomic research but one that has been under-utilised due to low yields of highly degraded nucleic acids. This limitation has prevented assaying such samples on early generation, whole-genome arrays, which required high quality and quantities of starting material. However, with improved whole transcriptome amplification systems and the evolution of next-generation array technologies, it is becoming more feasible to use low quantities of FFPE material. In a pilot study, we investigated whether low volume, FFPE diagnostic samples of prostate cancers (PCa) could be successfully applied to the Infinium HumanMethylation450 Beadchip (HM450K) array and the Agilent SurePrint G3 Human Gene Expression v2 Microarray. Here we present the results of both the individual and combined analyses of these assays. Methods: Fourteen FFPE diagnostic PCa samples, consisting of six transrectal ultrasonography biopsies and six transurethral resection samples were selected from the Cancer Council Victoria Prostate Cancer Program. DNA and RNA were extracted from macro-dissected prostate tumour tissue. DNA samples were whole genome amplified and assayed on the Infinium HM450K array. RNA samples were amplified in duplicate using two different whole-transcriptome amplification kits and assayed on the Agilent SurePrint array. Quality control measures and analyses were performed using publicly available and in-house analysis tools. Results: DNA and RNA were successfully amplified for 11 samples. These samples were applied to the Infinium HM450K and the Agilent SurePrint arrays, and quality control assessment and analyses are currently underway. Conclusion: One of the greatest challenges in PCa management is distinguishing at diagnosis those men who are at risk of progressing to life-threatening disease from those whose disease will remain indolent. Diagnostic tumour samples are a vital resource for biomarker discovery but have been under-utilised due to the nature of the material. This pilot study suggests that low volume, FFPE tumour material can be successfully interrogated using next-generation, genome-wide arrays, thereby opening up a previously under-utilised resource for epidemiological studies.

3492M

Genomic sequencing reveals significantly higher mutation load in genomes of fathers of children with *de novo* germline mutation in RB1 gene. *A. Ganguly[1], E. Toorens[2], J. Richards-Yutz[1], T. Ganguly[2], KG. Ewens[1].* 1) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 2) Penn Genomic Analysis Core, University of pennsylvania, Philadelphia, PA.

Background: Retinoblastoma (RB) is a neoplasm of the retina that occurs in infants and young children. Sporadic bilateral retinoblastoma (RB), results from a *de novo* germline mutation in the RB1 gene. This *de novo* mutation occurs in the father's gamete before the child's conception in about 94% of cases. However, paternal age is not significantly associated with risk of having a child with *de novo* germline mutation in RB. We have previously shown that paternal diet, occupation, and medical exposures to mutagens before the child's conception affect this risk. **Methods:** We investigated the mutation load of the paternal genome and compared it to that of the mother and unrelated control adults using whole genome or exome sequencing from 12 parent-child trios with sporadic bilateral RB and four controls . Whole genome sequencing library was prepared using Illumina TruSeq DNA library protocol. Exome capture was performed using NimbleGen V3.0 oligonucleotide libraries. Following sequencing on Illumina HiSeq, variant calling was done with the GATK Unified Genotyper and filtered using the GATK VariantFiltration and SelectVariants tools. Exome sequencing reads were filtered against the target regions to remove off target reads. Remaining variants were annotated with Annovar before further filtering via Python script and in Excel. **Results and Discussion:** The results indicate that the distribution of mutations including non-synonymous, synonymous, frameshift and non-frameshift mutations was significantly different between the paternal and maternal genomes ($p < 0.001$). The paternal genomes included increased number of non-synonymous damaging changes. Analysis of the mutated genes indicated novel mutations in genes in nucleotide excision repair and mismatch repair pathways, as well as in notch signaling, purine and caffeine metabolism pathways. These results provide a strong association with our epidemiologic data on risk enhancing paternal exposures that includes medical radiation, diet and occupation. In conclusion, this data validates our hypothesis that paternal exposure and mutation load increase the risk of having a child with sporadic RB. We acknowledge that the data set is very small; however considering that RB is a rare disease, it is important to emphasize that we have the largest number of RB parent-child trios currently available in any institution, and we are in the process of enlarging the data set.

3493T

Differential Gene Expression In Key Oncolytic Pathways Observed Between Caucasian-American and African-American Women with Triple-Negative Breast Cancer. *J.E. Getz[1], L.L Baumbach-Reardon[1], C. Gomez[2], M.E. Ahearn[1], M. Jorda[2], C.R. Legendre[1], W. Tembe[1], S. Nasser[1], V. Yellapantula[1], M.D. Pegram[3], J.D. Carpten[1].* 1) Integrated Cancer Genomics Division, Translational Genomics Research Institute., Phoenix, AZ; 2) Miller School of Medicine, University of Miami Medical School, Miami, FL; 3) Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA.

In the U.S., breast cancer (BC) incidences among African American women (AA) and Caucasian women (CA) are similar, however, AA have a significantly higher mortality rate (20%). AA often presents with tumors of higher grade, later stage, and are more likely to be triple-negative breast cancer (TNBC). Although multiple factors contribute to these cancer disparities, it is essential that we identify the molecular characteristics and underlying biological differences between CA and AA TNBC. In this study, we performed gene expression array analysis on archived formalin fixed paraffin embedded (FFPE) blocks from a multi-ethnic U.S. cohort of node-negative (N0) TNBC samples. Total RNA was isolated from 10 μm scrolls from each FFPE block. Following cDNA synthesis, each sample was hybridized to a breast-enriched gene expression array (Affymetrix, BC DSA Research Tool). Expression analysis was conducted using GeneSpring 12.1 analytical software and after contemporary data analyses, the final study cohort consisted of 10-AA, 13-CA. PCA analysis revealed that the samples clustered well with respect to ethnicity and unsupervised cluster analysis, based on ethnicity and genes, was performed. The resulting dendogram segregated into distinct subgroups based on ethnicity, revealing a pattern of differential gene expression between the cohorts. A list of differentially expressed genes from each cohort (DEG) were selected using ANOVA analysis (fold change > 3.0, p value <.05) followed by the Benjamin/Hochberg method for multiple-testing correction. Finally, the lists of DEG were uploaded into GeneGo MetaCore to identify functionally enriched pathways. These analyses revealed differentially expressed genes pathways enriched for cytoskeletal remodeling, cell adhesion and EMT pathways. In particular, significantly deregulated genes associated with the Wnt/β-catenin pathway were observed in the AA cohort as compared to the CA, suggesting that this pathway may contribute to the more aggressive phenotype in AA women diagnosed with TNBC. In summary, our results indicate gene expression differences within several key oncolytic pathways across these ethnic groups and these results are currently being validated through several parallel approaches. These studies have important implications for further understanding BC and TNBC health disparities, as well as future tailored approaches to prediction, prevention and therapeutic advances.

3494S

Shared genetic background between chronic gastroesophageal reflux and Barrett's esophagus and esophageal adenocarcinoma, consistent with a causal relationship. *P. Gharahkhani[1], J. Tung[2], T.L. Vaughan[3], D.C. Whiteman[1], S. MacGregor[1], Barrett's and Esophageal Adenocarcinoma Consortium.* 1) QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 2) 23andMe, Mountain View, CA, USA; 3) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

Esophageal adenocarcinoma (EA) is a cancer with rising incidence and high mortality rate. Every year 0.12-0.5% of patients with Barrett's esophagus (BE), a precancerous metaplastic change, develop EA. Chronic gastroesophageal reflux (GERD) is a risk factor for both BE and EA. However, it is not clear whether GERD actually causes those diseases. In a previous study we found genetic overlap between BE and EA but did not detect significant evidence of common genetic background between GERD and BE/EA. The previous study was limited by small sample size, which may have resulted in false negative findings. Here we aimed to further investigate the polygenic overlap between GERD and BE/EA using larger datasets, and to use the Mendelian randomization approach to investigate the causality of GERD for those diseases. The top genome-wide GERD-associated SNPs from a study of 23andMe in 8,743 GERD cases and 43,932 controls were used to calculate polygenic risk scores for GERD in our study population (2,051 BE cases, 1,239 EA cases, and 2,088 controls of European descent from the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON)). We showed the 23andMe risk scores significantly ($P<5e-4$) predicted BEACON reflux status, providing the strongest evidence to date for a role of genetic variation in reflux predisposition. Further, the scores were used as an instrumental variable in a Mendelian randomization approach using logistic structural mean model estimator to investigate the causality of GERD for BE and EA. Unlike our earlier smaller study of GERD, we found a significant polygenic overlap between GERD and BE/EA. Our preliminary Mendelian randomization analysis provides estimates for the role of GERD on BE and EA risk which are consistent with causality. Our findings on causality are potentially clinically important as they support the use of GERD interventions to control cancer risk.

3495M

The Kaiser Permanente Biobank: A multi-region, multi-ethnic resource linking specimens and electronic medical records for broad research in an integrated health care delivery system. *KAB. Goddard[1], N. Cardoza[2], LA. Croen[3], HS. Feigelson[4], K. Harris[5], S. Honda[6], M. Horberg[7], C. Jonas[7], C. Koebnick[8], MK. Nguyen[8], AA. Owen-Smith[5], S. Rowell[3], C. Schaefer[3], KS. Smith[1], C. Somkin[3], J. Vu[2], MG. Wrenn[4], C. Yonehara[6], K. Emmons[2].* 1) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 2) Kaiser Permanente Program Office and Kaiser Foundation Research Institute, Oakland, CA; 3) Kaiser Permanente Division of Research, Oakland, CA; 4) Institute for Health Research, Kaiser Permanente, Denver, CO; 5) Center for Health Research Southeast, Kaiser Permanente Georgia, Atlanta, GA; 6) Center for Health Research, Hawai'i, Kaiser Permanente Hawaii, Honolulu, HI; 7) Mid-Atlantic Permanente Research Institute, Mid-Atlantic Permanente Medical Group, Rockville, MD; 8) Kaiser Permanente Southern California, Department of Research and Evaluation, Pasadena, CA.

Kaiser Permanente (KP) is an integrated healthcare delivery system spanning seven geographic regions in the US and including 9.3 million members. The diverse membership reflects the US population, where about half of the members are non-Hispanic white. Membership includes between 4% and 29% of the total population in the coverage areas, and broadly represents regional populations in overall demographic and socioeconomic status measures. In the last decade, KP established several regional biobanks that now include specimens and phenotypic data for a total of 210,000 members. We describe the design and features of a new initiative to consolidate these efforts into a single entity, called the KP Biobank, and to expand the cohort to a total of 500,000 participants. The KP Biobank will include a general cohort (410,000 participants), designed to represent the diversity of KP members, and two specialized cohorts: a cancer cohort with 30,000 participants, and a pregnancy cohort with 60,000 participants. Specimens will be stored centrally and linked to phenotypic information from the medical record. Our data systems capture comprehensive and longitudinal records on every medical encounter, including diagnoses, procedures, pharmacy data, tumor registry, pathology records, imaging, and laboratory results. Participants have about 23 years of membership on average (range 0 - 71 years), with comprehensive electronic records for about the last 15 years. We also collect patient reported information on demographics and behavioral and environmental exposures via survey. The KP Biobank is designed for both discovery and translational research. Our setting of research units embedded within a healthcare delivery system supports clinical integration of new discoveries and research across the translational continuum. The KP Biobank Translational Research Center will specifically support this function. The consolidated governance of the KP Biobank will streamline and facilitate collaboration and partnerships through a single Access Review Committee and application procedure. Any researcher, including external scientists, may apply for use of specimens. Participants consent to broad future uses of specimens and data, and allow for broad data sharing, including through dbGaP. This novel resource will facilitate future genomic research, particularly on research questions addressing diverse populations, longitudinal measures to assess changes over time, and long-term outcomes.

3496T

Diverse types of lymphoid cancers cluster in families. *S.J. Jones[1,2], J. Voong[3], R. Thomas[2], G.W. Slack[4], J. Connors[4], J. Graham[3], A.R. Brooks-Wilson[2,3].* 1) University of British Columbia, Vancouver, British Columbia, Canada; 2) Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada; 3) Simon Fraser University, Vancouver, British Columbia, Canada; 4) Centre for Lymphoid Cancers, British Columbia Cancer Agency, Vancouver, British Columbia, Canada.

Lymphoid cancers are collectively the sixth most common cancer in the USA and impose a large health and financial burden on individuals and health care systems. Lymphoid cancers are a biologically diverse group of neoplasms, yet familial aggregation of various lymphoproliferative disorders has been observed, suggesting shared genetic risk factors. Specifically, non-Hodgkin lymphoma, Hodgkin lymphoma, multiple myeloma and chronic lymphocytic leukemia aggregation may be specific to gender, age, ethnicity, infectious agents and familial relationship. Our collection of over 120 families with more than one lymphoid cancer diagnosis will aid in understanding the inter-relationship of lymphoid cancers between and within lymphoid cancer families. The relationships between different lymphoid cancers observed in these families was investigated by categorizing affected individuals by type of lymphoid cancer, age of onset, affected generation and rarity of each type of lymphoid cancer. We have observed that within these families, some types of lymphoid cancers co-occur more frequently than expected based on their incidence in the general population. This co-aggregation of specific types of lymphoid cancers may be due to genetic and/or other risk factors that are shared within families. In addition, we observe anticipation in the set of families; the mean age of diagnosis among generations was significantly different for non-Hodgkin lymphoma, Hodgkin lymphoma, chronic lymphocytic leukemia and all lymphoid cancers collectively.

3497S

Cancer Predisposition Genes in Whole Exome Sequencing: How Do Findings Correlate with Cancer Histories? *N.M Lindor[1], S. Middha[2], S.K. McDonnell[2], J.E. Olson[2], J.R. Cerhan[2], E.D. Wieben[3], K.J. Johnson[4], G. Farrugia[5], S.N. Thibodeau[6] For the Mayo Clinic Community Biobank, Center for Individualized Medicine.* 1) Dept Health Science Research , Mayo Clinic, Scottsdale, AZ; 2) Health Sciences Research, Mayo Clinic, Rochester, MN; 3) Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN; 4) Ariosa Diagnostics, Inc, San Jose CA; 5) Internal Medicine, Div of Gastroenterology, Ctr for Individualized Medicine, Mayo Clinic, Rochester, MN; 6) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

Background: Prospective demonstration of clinical utility of Whole Exome Sequencing (WES) will require many years of data collection, so retrospective analyses are needed to inform future uses. In this study we examined how DNA variants in known cancer predisposition genes correlated with cancer histories of individuals who had completed their lifespan. **Methods**: WES was performed on 89 deceased individuals from the Mayo Clinic Biobank. Mean age at death was 74 yrs (range 28-93). The Mayo medical record was abstracted for each individual. WES findings including SNVs and indels were filtered and sorted to identify high quality Tier 1 (likely damaging) changes and Tier 2 (missense variants) for which either SIFT or Polyphen predicted pathogenicity in 58 cancer predisposition genes. No copy number variants were evaluated. Variants were scored for likely pathogenicity using IARC scale (1=neutral; 5=pathogenic). **Results:** Of the 89 individuals studied, 39 (26 M, 13 F) had invasive cancers and 50 had no cancers (25M, 25F). A total of 115 Tier 1 or Tier 2 variants were found. Among those with cancers, a mean of 1.5 variants (range 0-3) involved 24 different genes; in those without cancer, a mean of 1.1 (range 0-4) were found involving 20 different genes. Manual curation of pathogenicity scores, blinded to cancer status, resulted in most variants remaining Class 3(VUS:variant of unknown significance): 46/58 variants in those with cancer and 48/57 in those without cancer were VUS. However, there were 3 Class 4 or 5 variants in the subjects with cancers (likely pathogenic) compared to zero in the subjects without cancer. **Conclusions:** Among 89 individuals, 39 who had developed a cancer were apparently more likely to have likely pathogenic mutations (3/39) within cancer predisposition genes than those individuals without cancer (0/50). However the majority of DNA variants found on WES of these genes currently will not be able to be interpreted beyond calling them a VUS. Copy number variants, inadequately covered regions in some genes, and untested genes could still account for some of the cancers. Nevertheless, because VUS with *in silico* assignments of pathogenic/damaging in cancer predisposition genes occurred at about the same frequency in those who had cancer as in those without cancer, great caution is warranted in assigning significance to VUS in these genes in an unselected population.

3498M

The prostate cancer risk mutation G84E in *HOXB13* is associated with the subtype of ETS fusion negative adenocarcinoma with early age of diagnosis. *M. Luedeke[1], A. Rinckleb[1], J. Stanford[2,3], L. FitzGerald[2], J. Schleutker[4,5], T. Wahlfors[4], R. Eeles[6,7], Z. Kote-Jarai[6], S. Weikert[8], H. Krause[8], K. Herkommer[9], J. Hoegel[10], C. Maier[1], and. the ICPCG[11].* 1) Department of Urology, University Hospital of Ulm, Ulm, Germany; 2) Fred Hutchinson Cancer Research Center, Division of Public Health Science, Seattle, Washington, USA; 3) Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA; 4) Institute of Biomedical Technology, University of Tampere and FimLab Laboratories, Tampere, Finland; 5) Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland; 6) The Institute of Cancer Research, Sutton, UK; 7) Royal Marsden National Health Service Foundation Trust, London and Sutton, UK; 8) Department of Urology, University Hospital Charité, Berlin, Germany; 9) Department of Urology, Klinikum rechts der Isar der Technischen Universität München, Munich, Germany; 10) Institute of Human Genetics, University Hospital of Ulm, Ulm, Germany; 11) International Consortium for Prostate Cancer Genetics.

HOXB13 was discovered as the first prostate cancer (PrCa) specific high-risk susceptibility gene. The most prevalent *HOXB13* germline mutation in PrCa patients of European descent is *HOXB13* G84E, which likely originated in Northern Europe. Previous molecular examination of a set of G84E driven tumors suggested a distinct somatic phenotype, where oncogenic ETS gene fusions appear at unusually low frequencies as compared to the general prevalence of ETS fusions in PrCa (22 % vs approx. 50 %). Hypothesizing that *HOXB13* could predispose to ETS fusion negative PrCa we have analyzed 942 cases from three European ancestry populations for the coincidence of *HOXB13* G84E and the most common ETS fusion, *TMPRSS2-ERG* (T2E), in corresponding tumor samples.

While the prevalence of T2E fusions was similar among study sites (range: 57.2% - 60.1%), the frequency of G84E genotypes differed markedly between US (1.5%), German (3.6%) and Finnish samples (8.3%). Despite the expected frequency gradient among study populations, all subsamples showed a strong enrichment of G84E mutation carriers among T2E fusion negative cases as compared to fusion positive cases (center adjusted OR = 4.96; 95% CI = 2.30 - 11.9; p = 0.0001). Consistent with the previous study, the crude frequency of the T2E fusion in *HOXB13* G84E carriers was 23.5 % (range 16.7 % - 28.5 %). Examination of disease characteristics highlighted age at diagnosis, with fusion positive cases being diagnosed 1.85 (0.95 - 2.74) years earlier than negative cases ($p < 0.0001$). Age at diagnosis in G84E carriers did not differ significantly from non-carriers (p = 0.13). However, within the subtype of fusion negative carcinoma, which is usually associated with later ages at diagnosis, carriers of G84E were diagnosed on average 3.45 (0.61 - 6.3) years earlier (p = 0.018). No associations were seen for tumor stage, tumor grade or diagnostic PSA levels.

In conclusion, this study has demonstrated a tumor type specific association for *HOXB13* G84E mutation carriers having a higher frequency of T2E fusion negative PrCa. Although the T2E fusion negative subtype is known to be associated with later ages of diagnosis, *HOXB13* driven tumors within this subtype may represent an early onset subgroup.

3499T

Complexities in genetic testing for allogeneic bone marrow transplant recipients and patients with hematologic malignancies. *D. Mancini-DiNardo, M. Landon, B. Abbott, M. Elias, J. Rinsky, B. Roa.* Myriad Genetic Laboratories, Inc., Salt Lake City, UT.

Allogeneic bone marrow transplant (allo-BMT) recipients and patients with a hematologic malignancy present unique challenges when their blood/buccal samples are submitted for genetic testing. White blood cells from allo-BMT recipients are usually entirely derived from the transplant donor. Buccal samples from such patients can be chimeric, containing cells derived from both the recipient and the bone marrow donor. Therefore, molecular genetic testing performed on DNA derived from blood/buccal samples of allo-BMT patients may reflect the genetic status of the transplant donor and not the recipient. Blood/buccal samples from individuals with a hematologic malignancy (HM) are also not recommended for genetic testing. This is because genetic analysis may detect somatic, non-heritable mutations caused by the hematologic malignancy, which are present only in the patient's blood. Genetic testing on blood/buccal samples submitted by patients with hematologic malignancies may therefore reflect the genetic status of the tumor rather than the germline genetic status of the patient. Here we present data from three representative cases that illustrate the complexity of genetic testing for allo-BMT recipients and patients with a hematologic malignancy. In one case, we tested a buccal sample from an allo-BMT patient that was sufficiently chimeric such that the donor DNA was detected on the genetic test results. In the second case, we illustrate the potential to report out a false negative result on a blood sample from an allo-BMT patient. In the final case, we demonstrate the detection of a mosaic large rearrangement in a patient with a hematologic malignancy. This large rearrangement was subsequently determined to be a somatic mutation present only in the patient's blood cells. We have therefore determined that cultured cells derived from skin fibroblasts are the most suitable sample type for allo-BMT patients, which is in accordance with the published 2012 NCCN guidelines. Fibroblasts are also recommended for HM patients where there is a high likelihood of leukemic cells circulating at a quantity that is detectable by our genetic analyses. We have collaborated with ARUP to assist in the procurement of fibroblast cells. It is imperative that germline genetic testing results reflect the heritable mutation status of the patient being tested, as they may impact the interpretation of test results and subsequent clinical management of the patient and their family members.

3500S

Familial lung cancer: A genetic epidemiology study. *D. Mandal[1], A. Bencaz[1], J. Chambliss[1], J.E. Bailey-Wilson[2].* 1) Department of Genetics, LSU Health Sciences Center, New Orleans, LA; 2) National Human Genome Research Institute, National Institutes of Health, Baltimore, MD.

The association between lung cancer (LC) and smoking is well known. However, only 15% of smokers are diagnosed with LC. Inherited genetic factors have a major role in lung cancer etiology. About 10% of LC cases (22,000 cases per year in the U.S.) have at least one first-degree relative affected with LC, and 25% of cases at least one first- or second-degree affected relative, indicating that family history is a significant risk factor. In the last few years histology specific genetic alterations in lung cancer have been reported in epidemiologic studies. In our previous studies, we have reported adenocarcinoma (38%) being the most common histologic subtype, followed by squamous cell (25%) and small cell (15%) in familial cancer cases. The objective of the present study is to describe the rapid ascertainment of familial lung cancer cases and to determine the co-segregation of histologic subtypes transmitted in familial lung cancer families. Eligible subjects (N=203) with two or more relatives affected with primary lung cancer were recruited from 24 parishes across southern Louisiana. Diagnosis of primary LC was confirmed by medical records, and histologic subtype (N= 173) was abstracted from pathology reports. About 20 of these families were developed as multigenerational families and biological specimens were collected on affected lung cancer cases and unaffected family members. Annual follow up on these families confirm a change of health status and new diagnosis in the families. Histological data on these families will be analyzed and co-segregation of specific histologic subtypes and lung cancer affection status will be determined in this unique collection of families. Methods for rapid ascertainment of lung cancer cases maintaining HIPAA compliance and communication about dbGaP policies with general public in the development of bio-repository will be discussed in detail.

3501M

Identification of a novel founder *MSH2c.705delA mutation causing colon cancer in a Druze population.** *M. Melas[1], C.C. Studenmund[1], K. McDonnell[1], L. Raskin[2], P.S. Boonstra[3], M. Zawistowski[3], B. Mukherjee[3], F. Lejbkowicz[4], H.S. Rennert[4], G. Rennert[4], S.B. Gruber[1].* 1) USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; 2) USC Division of Epidemiology, Department of Medicine, Vanderbilt University, Nashville, Tennessee; 3) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 4) CHS National Israeli Cancer Control Center, Haifa, Israel.

Lynch Syndrome is an autosomal dominant cancer genetic syndrome resulting from mutations in the DNA mismatch repair genes *MSH1, MSH2, MSH6, PMS2* and *EPCAM*. Individuals carrying mutations in these genes characteristically develop cancer of the colon, rectum, endometrium and other cancers at an early age. A number of founder mutations in Lynch-causing genes have previously been described as important in identifying individuals at risk within specific populations. The Druze are a unique religious and ethnic population that originated approximately 11,000 years ago in the Arabian Peninsula. The population subsequently migrated throughout the Middle East, primarily to modern Syria, Lebanon, Israel and Jordan, between 1,000 and 500 years ago. The Druze are culturally discouraged from marrying outside their communities and endogamy rates are relatively high. In the present study we report a novel *MSH2* mutation, c.705delA, in a group of Druze individuals who developed Lynch Syndrome. To establish whether this variant is a founder mutation we performed haplotype analysis using SNP genotyping data from 12 Druze carrying *MSH2**c.705delA mutation (Infinium HumanCore Beadchip) and 47 Druze individuals from the Human Genome Diversity Project (HGDP) with chromosome 2 phasing. Concomitantly, recombination length was determined by carrying out a short tandem repeat genetic analysis of the same Druze carriers together with 9 Druze non-carriers and 8 Christian Arab non-carriers employing six highly polymorphic microsatellite markers near the *MSH2* gene: D2S367, D2S370, D2S378, D2S391, D2S1248 and D2S2238. Haplotype phasing using ShapeIt2 software showed a 13 Mb haplotype common to all *MSH2**c.705delA mutation carriers. This haplotype was not found in non-carriers and also not in the 47 Druze individuals from the HGDP. Attempts to determine the age of the mutation were confounded by significant consanguinity and complicated population structure of the Druze, which highlights the difficulty of estimating the age of a mutation that depends on the assumption of independent ascertainment. In summary, the present study describes a novel *MSH2**c.705delA mutation among the Druze population, establishes this deletion as a founder mutation and illustrates the challenges of statistical genetic analyses in highly inbred populations. Additionally, the results of this study may contribute to improved cancer genetic screening and cancer care in the Druze.

3502T

Investigating the Genetic Basis of Breast Cancer Disparities Using Whole Genome Sequencing and Parallel Computing. *J.J. Pitt[1,2], TY. Yoshimatsu[3], Y. Zheng[3], AJ. Grundstad[2], J. Tuteja[2], A. Odetunde[4], G. Khramtsova[3], TO. Ogundiran[5], CP. Babalola[4], OA. Ojengbede[6], CO. Olopade[7], D. Huo[8], KP. White[1,9], OI. Olopade[3,9].* 1) Committee on Genetics, Genomics, and Systems Biology, The University of Chicago, Chicago, IL, USA; 2) Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL, USA; 3) Center for Clinical Cancer Genetics and Global Health, Department of Medicine, The University of Chicago, Chicago, IL, USA; 4) Institute for Advanced Medical Research and Training, College of Medicine & Faculty of Pharmacy, University of Ibadan, Ibadan, Oyo, Nigeria; 5) Department of Surgery, University of Ibadan, Ibadan, Oyo, Nigeria; 6) Centre for Population & Reproductive Health, College of Medicine, University of Ibadan, Ibadan, Oyo, Nigeria; 7) Section of Pulmonary and Critical Care, Department of Medicine, The University of Chicago, Chicago, IL, USA; 8) Department of Health Studies, The University of Chicago, Chicago, IL, USA; 9) Department of Human Genetics, The University of Chicago, Chicago, IL, USA.

Breast cancer incidence and mortality rates vary widely based on geography and ethnicity. Compared to Caucasians, individuals with African ancestry are more likely to die from breast cancer, which is in part attributable to higher prevalence of the triple negative subtype. It is hypothesized that these disparities manifest through population-specific inherited and somatic genetic variation. Breast cancer specimens and germline DNA have been collected from the University College Hospital Ibadan in southwest Nigeria. To date, 17 tumor-normal pairs and 50 germline breast cancer samples have undergone whole genome sequencing on the Illumina platform, with neoplastic and non-neoplastic sequenced to average depths of 90x and 30x, respectively. In addition, 25 asthmatics and 75 healthy Nigerians from the surrounding community have been sequenced as controls. To handle the computational burden inherent to large-scale sequencing analyses, we have developed SwiftSeq, a modular, highly-parallel workflow for fast, efficient, and robust processing of DNA sequencing data. Using Genome Analysis Toolkit's best practices, SwiftSeq is able to completely align, process, genotype, and annotate a 30x genome in ~36-40 hours. By scaling with compute resources, our framework can analyze hundreds of genomes in days, rather than weeks. Currently, 19 germline case and 21 control genomes (mean depth ~27x) have been pushed through SwiftSeq. Each harbor, on average, 3,921,264 SNPs; 621,141 indels; and 306,882 variants not reported in dbSNP 137. To investigate the effects of below average sequencing depth, we down-sampled to 30x, 25x, and 20x in triplicate using two well-covered samples (~35x and ~37x). Genotypes from all replicates showed only marginal differences relative to the original data, indicating our germline calls are robust to dispersions from the targeted sequencing depth. While early onset and/or familial cases lacking deleterious BRCA1/2 mutations were explicitly selected for sequencing, damaging mutations were found in other known cancer susceptibility genes such as BRIP1, FANCC, and SMARCE1. This underlies a potentially expanded role for these genes in African breast cancer pathogenicity. With continued sequencing, analysis, and comparison to Caucasian tumor-normal genomes from The Cancer Genome Atlas, we will elucidate the unique characteristics of African breast cancer genomes.

3503S

The Genetic Testing in Epithelial Ovarian Cancer (GTEOC) Study: Direct access to BRCA1/2 genetic testing in oncology. *M. Tischkowitz[1,2], J. Drummond[2], E. Thompson[2], G. Sagoo[3], B. Newcombe[4], E. Barter[5], P. Ridley[6], S. Miller[7], F. Thompson[8], H. Webb[9], C. Hodgkin[4], L.T. Tan[7], M. Daly[9], S. Ayers[5], B. Rufford[6], C. Parkinson[10], H. Earl[10], T. Duncan[8], P. Pharoah[11], S. Abbs[2], N. Hulbert-Williams[12], R. Crawford[10], J. Brenton[10,13], H. Shipman[1,2].* 1) University of Cambridge, Cambridge, United Kingdom; 2) East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK; 3) PHG Foundation, Cambridge, UK; 4) Cambridge Cancer Trials Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK; 5) Clinical Oncology, Peterborough and Stamford Hospitals NHS Foundation Trust, Peterborough, UK; 6) Cancer Services, The Ipswich Hospital NHS Trust, Ipswich, UK; 7) Clinical Cancer Services, Hinchingbrooke Health Care NHS Trust, Huntingdon, UK; 8) Department of Obstetrics and Gynaecology, Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK; 9) Department of Oncology, The Queen Elizabeth Hospital King's Lynn NHS Foundation Trust, King's Lynn, UK; 10) Cancer Services, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK; 11) Department of Primary Care and Public Health, University of Cambridge, Cambridge, UK; 12) Department of Psychology, University of Chester, Chester, UK; 13) Cancer Research UK Cambridge Institute, Cambridge, UK.

Background: Genetic testing for germline mutations in BRCA1/2 has been available for many years, but until recently technical complexity and cost has limited testing access to high-risk cases. Further, many epithelial ovarian cancer (EOC) patients who would be eligible for testing are not referred for a genetic assessment. The advent of targeted therapies for BRCA1/2 mutation carriers will necessitate new service delivery models for genetic testing. The Genetic Testing in Epithelial Ovarian Cancer (GTEOC) Study started in July 2013 and explores the feasibility and acceptability of offering direct genetic testing to all women recently diagnosed with EOC. Methods: Recruitment is on-going through six centres in East Anglia, UK, which has an outbred population of 2.5 million with no BRCA1/2 founder mutations. All women with serous or endometriod ovarian cancer diagnosed with in the last 12 months are eligible, irrespective of age and family history of cancer. Consent for genetic testing is obtained by the research team after participants have reviewed the study information sheet. Descriptive statistics are used to present data collected to date. Economic analysis is being undertaken using a cost effectiveness evaluation. The psychosocial arm of the study is utilising quantitative questionnaires and qualitative interviews subjected to Interpretive Phenomenological Analysis. Results: To date we have recruited 76 participants out of a total of 139 eligible women. The mean age at diagnosis was 65.5 years (range 29-85), mean time since diagnosis to recruitment = 142 days (range 15-355 days). BRCA1/2 gene testing has been completed on 65 samples, resulting in detection of pathogenic mutations in 6 women and variants of uncertain clinical significance (VUS) in 6 women. Of the women found to have a pathogenic mutation, 3 had no significant previous family history of cancer to their knowledge. We will present our preliminary psychosocial findings and economic analysis, and highlight the challenges and success of our model and implications for genetic testing in oncology settings. Conclusion: The initial GTEOC study recruitment rate is 55%. Pathogenic BRCA1/2 mutations have been identified in 9.2% of women and VUS in 9.2%. This method of population-based genetic testing appears to be acceptable to patients and is less resource-intensive than standard practice where all patients have a full assessment by the genetics team prior to testing.

3504M

Insecticide Exposure Induces Leukemia-Associated Gene Aberrations. *M.P. Navarrete Meneses[1], M. Betancourt[2], E. Bonilla[2], M. Altamirano[3], C. Salas[1], A. Reyes[1], M. Sanabrais[2], P. Pérez Vera[1].* 1) Instituto Nacional de Pediatría, Mexico City; 2) Universidad Autónoma Metropolitana-Iztapalapa, Mexico City; 3) FES-Zaragoza, UNAM. Mexico City.

Background: Leukemia is the most common childhood cancer. It is characterized by the presence of cells with chromosomal abnormalities, which are proposed to arise in utero. It is associated with prenatal exposure to pesticides. Epidemiological studies have demonstrated that environmental exposures to these chemicals could be implicated in leukemogenesis, however there is limited biological evidence showing that pesticides can induce leukemia-associated gene aberrations. Malathion and permethrin are two widely used insecticides; nevertheless their genotoxic and carcinogenic potentials are controversial. The aim of this study was to detect numerical and structural rearrangements in leukemia-related genes on human peripheral blood lymphocytes exposed in vitro to insecticides. Methods: Mononuclear cells from two healthy volunteers were cultured in triplicates for 72h and exposed to 200μM of permethrin or malathion for the last 24h. MLL, ETV6 and RUNX1 genes were analyzed with fluorescence in situ hybridization. Besides, numerical aberrations were assessed using centromeric probes for chromosomes 12, 18 and 14/21. Structural and numerical analysis were performed separately, 1000 nuclei were scored. Groups were compared with U-Mann Whitney test. Results: Permethrin exposure increased the frequency of cells with MLL structural aberrations and copy number deviations significantly, as well as the diversity and complexity of damage. It also induced aneuploidy significantly. An increased frequency of cells with abnormalities in ETV6 and RUNX1 was detected, however it was not significant. On the other hand, a significant increase in the frequency of cells with structural damage in MLL gene was observed with malathion exposure. This effect was not observed on ETV6 and RUNX1 genes. Neither chromosome gains nor losses were significantly induced by malathion exposure. Discussion: Permethrin exposure can affect leukemia-related genes in structure and number. The results show that this pyrethroid insecticide has clastogenic and aneuplogenic potential. Similar patterns of alterations have been reported in cells treated with leukemogenic agents such as etoposide. Malathion also increased the frequency of altered cells, however a greater effect was observed with permethrin. Conclusion: This study provides biological evidence showing that insecticide exposure is implicated in the generation of leukemia-associated gene aberrations. CONACYTCB-2012-01/183467.

3505T

A systematic approach to clinical classification of DNA sequence variants in mismatch repair genes: the InSiGHT initiative. *B.A. Thompson[1,2], A.B Spurdle[1], J. Plazzer[3], M.S. Greenblatt[4], P. Moller[5], F. Macrae[3], M. Genuardi[6,7], InSiGHT Variant Interpretation Committee.* 1) Genetics & Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 2) School of Medicine, University of Queensland, Brisbane, Australia; 3) Department of Colorectal Medicine and Genetics, Royal Melbourne Hospital, Australia; 4) Vermont Cancer Center, University of Vermont College of Medicine, Burlington, VT, USA; 5) Research Group on Inherited Cancer, Department of Medical Genetics, Oslo University Hospital, The Norwegian Radium Hospital, Oslo, Norway; 6) Department of Biomedical, Experimental and Clinical Sciences, University of Florence, Italy; 7) Fiorgen Foundation for Pharmacogenomics, Sesto Fiorentino, Italy.

Clinical classification of sequence variants identified in hereditary disease genes directly affects management of patients and their relatives. The International Society for Gastrointestinal Hereditary Tumours Variant Interpretation Committee (InSiGHT VIC) undertook a collaborative effort to develop, test and apply a standardised classification scheme to germline variants in the Lynch syndrome-associated genes MLH1, MSH2, MSH6 and PMS2 (accounting for ~2% of all colorectal and endometrial cancers). Unpublished data submission was encouraged to assist in variant classification and was recognised through microattribution. The scheme was refined by multidisciplinary expert committee review of clinical and functional data available for variants, applied to 2,360 sequence alterations, and disseminated online. Assessment using validated criteria altered classifications for 66% of 12,006 database entries. Clinical recommendations based on transparent evaluation are now possible for 1,370 variants that were not obviously protein truncating from nomenclature. All data were merged into one publicly available database, which will continue to be curated according to the classification scheme. This large-scale endeavour has important implications for the clinical management of suspected Lynch syndrome families; at the same time, it provides an important model of international multidisciplinary collaboration for DNA variant interpretation.

3506S

PALB2 variant database - a joint collaboration between LOVD and ClinVar. *M.J. Landrum[1], D. Subramanian[2], G. Riley[1], D. Maglott[1], R. Villamarin-Salomon[1], S. Chitipiralla[1], J. den Dunnen[3], M. Tischkowitz[4].* 1) NCBI/NLM/NIH, Bethesda, MD; 2) Cambridge University Hospitals Trust, Cambridge, UK; 3) Human and Clinical Genetics, LUMC, Leiden, The Netherlands; 4) Department of Medical Genetics, University of Cambridge, UK.

Background Biallelic germline loss-of-function mutations in PALB2 (also known as FANCN) cause Fanconi Anemia, whereas monoallelic loss-of-function mutations are associated with an increased risk of breast cancer and pancreatic cancer. The LOVD platform, providing a flexible, freely available tool for the collection and display of DNA variants, was used to establish a PALB2 gene variant database. ClinVar was released publicly in April 2013 and provides a freely accessible, public archive of reports of the relationships among human variations and phenotypes. Here we describe a joint collaboration between ClinVar and LOVD to minimize inconsistencies in listing PALB2 variants and maximize accessibility for potential users. Methods PALB2 variants were identified through literature searches dating back to 2006 (the year PALB2 was identified), collated, classified and data uploaded to LOVD. The database was linked to Fanconi Anemia Mutation Database (Rockefeller University). Variant data was exported from LOVD and shared with ClinVar. In the ClinVar submission process, variant data were validated; the pathogenicity values were mapped to the American College of Medical Genetics (ACMG) standard terms; and variants were aggregated based on the combination of variant and disease. Much of the processing work involved synchronizing the LOVD and ClinVar data models e.g. determining whether a list of multiple phenotypes indicated individuals with all listed disorders or independent variant/disorder relationships. Results The LOVD database currently contains 242 listed variants in 671 individuals. The LOVD submission makes up 77% of the 342 PALB2 submissions to ClinVar. The other PALB2 ClinVar submissions are unpublished variants from clinical laboratories, structural variants that include PALB2 with other genes, and a handful of variants described in OMIM. 27% of the submissions in PALB2 are asserted as Pathogenic, 3% as Likely pathogenic, 32% as Uncertain significance, 41% as Likely benign, and 10% as Benign. Conclusions This provides an example of facilitated data sharing between different databases, which results in a higher-quality level of data over merely providing weblinks, because the variants have been assessed independently by different teams and data standardization is increased. Collaborations between resources such as the LOVD and ClinVar are essential for genes with important clinical consequences such as PALB2.

3507M

Combined contribution of intermediate-risk gene rare variants and modest-risk SNP genotypes to early-onset breast cancer. *E.L. Young[1], B.J. Feng[2], A.W. Stark[1], T.C. Francy[1], A.M. Paquette[1], J.S. Rosenthal[1], N. Forey[3], G. Durand[3], S. McKay-Chopin[3], M. Hashibe[4], J. Gertz[1], F. Le Calvez-Kelm[3], F. Lesueur[5], D.E. Goldgar[2], S.V. Tavtigian[1], Breast Cancer Family Registries.* 1) Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA; 2) Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA; 3) Genetic Cancer Susceptibility group, International Agency for Research on Cancer, Lyon, France; 4) Department of Family and Preventive Medicine, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA; 5) Genetic Epidemiology of Cancer team, Inserm, U900, Institut Curie, Paris, France.

Since the introduction of genetic testing of BRCA1 and BRCA2 in women with family histories of breast cancer and/or ovarian cancer, many additional genes have been identified as breast cancer susceptibility genes, and many of these genes are included in multigene panel testing. Building from our published case-control mutation screening studies of ATM, CHEK2, MRE11A, RAD50, NBN, RAD51, RINT1, and XRCC2, we evaluated the contribution of these eight genes, plus mutation screening of BARD1 and genotyping of 18 Breast Cancer Association Consortium (BCAC) confirmed modest-risk SNPs, to early onset breast cancer among a series of 1300 cases and 1120 ethnically matched controls that did not carry pathogenic variants in BRCA1, BRCA2, or PALB2. Rare missense substitutions were graded on severity using Align-GVGD, Polyphen2, CADD, and MAPP. To avoid gene-by-gene over-fitting, we applied uniform missense substitution analysis models across all nine genes. Using attributable fraction, familial relative risk, and receiver operator characteristic (ROC) analyses, we: (1) compared the contribution of known and likely pathogenic missense substitutions to the contribution of protein truncating variants across the nine gene set, and (2) compared the contribution of the nine gene mutation screening data to the 18 BCAC SNP genotyping data. Finally, setting an odds ratio (OR) of 2.5 as a threshold for the point at which a genotype-defined risk estimate is high enough to justify early screening (early mammography or breast MRI), we estimated the fraction of cases and controls that reach the $OR \geq 2.5$ threshold on the basis of mutation screening data or SNP genotyping data. From preliminary results across several analysis models, mutation screening of the nine genes suggests that approximately 8.8% of cases and 3.7% of controls carried one or more rare variants associated with an $OR \geq 2.5$, with over half of these variants being missense substitutions. Similarly, treating 15 (of 18 now genotyped) common modest-risk SNPs as a polygene and applying a simple multiplicative model, we found that approximately 3.6% of cases and 1.2% of controls had a modest-risk SNP polygene genotype associated with an $OR \geq 2.5$. We plan on combining the intermediate-risk gene mutation screening data with the modest-risk SNP genotyping data to paint a clearer picture of clinically actionable breast cancer susceptibility explained by genes other than the established high-risk susceptibility genes.

3508T

Germline Next Generation Full Gene Sequencing of *MLH1*, *MSH2* and *MSH6* Detects Pathogenic Mutations in Cases Previously Tested Negative for a Germline Mutation. *R.P. Graham[1], M.S. DeRycke[1], A. French[1], S. Gallinger[2], M. Cotterchio[3], R. Haile[4], G. Casey[5], M.A. Jenkins[6], J.L. Hopper[7], M. Woods[8], L. Le Marchand[9], J. Potter[10], P.A. Newcomb[10], D. Duggan[11], E.L. Goode[1], N. Lindor[12], S.N. Thibodeau[1] On behalf of the Colon Cancer Family Registry.* 1) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Department of Surgery, Mount Sinai Hospital, Toronto, Ontario, Canada; 3) Population Studies and Surveillance, Cancer Care Ontario, Ontario, Canada; 4) Department of Medicine, Division of Oncology, Stanford Cancer Institute, Palo Alto, CA; 5) Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA; 6) Centre for Epidemiology & Biostatistics, Melbourne School of Population & Global Health, University of Melbourne, Melbourne, Victoria, Australia; 7) Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Melbourne, Victoria, Australia; 8) Memorial University of Newfoundland, St. John's, Newfoundland, Canada; 9) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI; 10) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 11) Translational Genomics Research Institute (TGen), Phoenix, AZ; 12) Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ.

Lynch syndrome (LS) is due to germline mutations in mismatch repair (MMR) genes and tumors of affected patients show microsatellite instability. A subset of likely hereditary cases with mismatch repair deficient (dMMR) tumors has no identified germline mutations in clinical screening of coding regions of MMR genes (*MLH1*, *MSH2*, and *MSH6*). In 98 such cases, we examined both coding and non-coding regions of key MMR genes by next generation sequencing. Germline DNA from 98 suspected LS patients (IHC loss of MLH1 [n=71], MSH2 [n=18] and isolated MSH6 [n=9]) identified in the Colon Cancer Family Registry and 90 unaffected spousal controls was sequenced using Agilent's custom capture and an Illumina HiSeq2000. We targeted the coding and non-coding regions (intronic, upstream and downstream) for *MLH1*, *MSH2* and *MSH6*. Pseudogene homology impaired adequate *PMS2* capture. Quality control filtering excluded variants based on read depth (<20), genotype quality (<30), and minor allele frequency (>5% in 1000 Genomes). Remaining variants were analyzed for potential abnormal splicing, published reports and their frequencies in suspected LS patients relative to controls. After filtering, there were 10 coding and 262 non-coding variants in suspected LS patients. Three of the coding variants (*MLH1* p.L653R; *MSH2* p.N596del; *MSH6* p.F958X) were pathogenic and the remainder predicted to be benign. There were 69 non-coding variants in *MLH1*, 144 in *MSH2* and 50 in *MSH6*. All *MLH1* variants were detected once, 4of which were novel and predicted cryptic splice sites. There was one known pathogenic *MLH1* intronic variant, c.454-13A>G. Two novel *MSH2* variants were predicted to form cryptic donor sites and 3 variants were present in >10% of cases compared to 1-3% of controls. A single private *MSH6* variant was predicted to alter splicing. IHC data correlated with variant interpretation for deleterious and suspected deleterious variants. The remaining variants in *MLH1*, *MSH2* and *MSH6* were novel and predicted to be neutral. Among 98 suspected LS cases analyzed, for which standard testing was negative, we identified 4 deleterious and 6 suspected deleterious alterations (10% rate of diagnosis). The unresolved cases may still be due to the non-coding variants discovered. Alternatively, mutations in other MMR genes, structural or epigenetic changes, or somatic alterations in these genes may explain these cases. This work was supported by grant UM1 CA167551 from the National Cancer Institute.

3509S

Identification of men with a genetic predisposition to prostate cancer: targeted screening of BRCA1/2 mutation carriers and controls. The IMPACT study Quality of Life Study. *E. Bancroft[1, 2], N. Aaronson[3], C. Mikropoulos[2,1], S. Saya[2], E. Page[2], J. Pope[2], E. Farrow[2], E. Castro[4], N. Gadea[5], C. Selkirk[6], S. Buys[7], J. Cook[8], K. Ong[9], R. Davidson[10], D. Eccles[11], M. Tishchkowitz[12, 13], L. Greenhalgh[14], J. Barwell[15, 16], C. Brewer[17], A. Henderson[18], R. Eeles[2,1], IMPACT Study Steering Committee and Collaborators.* 1) Royal Marsden NHS Foundation Trust, London, UK; 2) Institute of Cancer Research, London, UK; 3) Netherlands Cancer Institute, Amsterdam, The Netherlands; 4) Spanish National Cancer Research Centre, Madrid, Spain; 5) Hospital Vall d'Hebron, Barcelona, Spain; 6) Center for Medical Genetics, NorthShore University HealthSystem, Evanston, IL, USA; 7) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA; 8) Sheffield Clinical Genetics Service, Sheffield Children's Hospital, Sheffield, UK; 9) Clinical Genetics Unit, Birmingham Women's Hospital, Birmingham, UK; 10) Duncan Guthrie Institute of Medical Genetics, Yorkhill NHS Trust, Glasgow, UK; 11) Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton, UK; 12) Addenbrooke's Hospital, Cambridge, UK; 13) The University of Cambridge, Cambridge, UK; 14) Clinical Genetics, Royal Liverpool Children's Hospital, Liverpool, UK; 15) University of Leicester, Leicester, UK; 16) University Hospitals Leicester, Leicester, UK; 17) Royal Devon and Exeter Hospital, Exeter, UK; 18) Northern Genetics Service, Newcastle upon Tyne Hospitals, UK.

Background: IMPACT is a multi-national targeted prostate cancer screening study of men with a known germline mutation in *BRCA1* or *BRCA2* which is thought to predispose to the disease. The study is recruiting male *BRCA1/2* mutation carriers and a control group, who have tested negative for a mutation present in their family, to undergo annual PSA screening +/- prostate biopsy. In addition to the screening protocol 15 centres are enrolling men in a quality of life study (QoL). The aims of the QoL study are: (1) To evaluate the psychosocial impact of screening for prostate cancer in men with *BRCA1/2* mutations in terms of (a) cancer worry, (b) distress of screening and (c) health related quality of life; (2) To identify factors associated with negative psychosocial outcomes; (3) To determine levels of compliance with PSA screening over time and identify significant predictors of compliance / non-compliance. Methods: Men enrolled in the IMPACT study at collaborating sites were invited to complete a questionnaire prior to each annual screening visit. The questionnaires included sociodemographic data and the following measures: HADS, IES, SF36, PC-MAX, Cancer Worry Scale, self reported risk perception and a knowledge questionnaire (understanding of *BRCA1/2* and prostate cancer). The results of the baseline questionnaire are presented. Results: A total of 484 men enrolled in the QoL study prior to their first screening visit (prospective cohort). A further 217 men have enrolled at subsequent visits (truncated prospective cohort). Uptake of the study was 82-100% at participating sites. Mean scores for HADS and SF36 were within reported general population norms and mean IES scores were within normal range. No statistically significant differences were observed between groups (*BRCA1* mutation carrier, *BRCA2* mutation carrier, and controls). Average knowledge score was 77%, demonstrating a good understanding. Mean MAX-PC scores were low suggesting a low level of prostate-cancer specific anxiety or PSA related anxiety in the cohort. Cancer worry levels were low and *BRCA2* carriers in particular perceived their risk of prostate cancer to be higher than average. Conclusions: Uptake of the QoL study is high in participating centres. Baseline psychosocial measures for men entering the study suggest a low prevalence of distress and participants demonstrate a good level of knowledge about prostate cancer and the *BRCA1/2* genes.